

Universität Münster

Arene Hydrogenation, Photocatalysis, Smart Screening and Data Science – Game Changing Technologies

wissen.leben

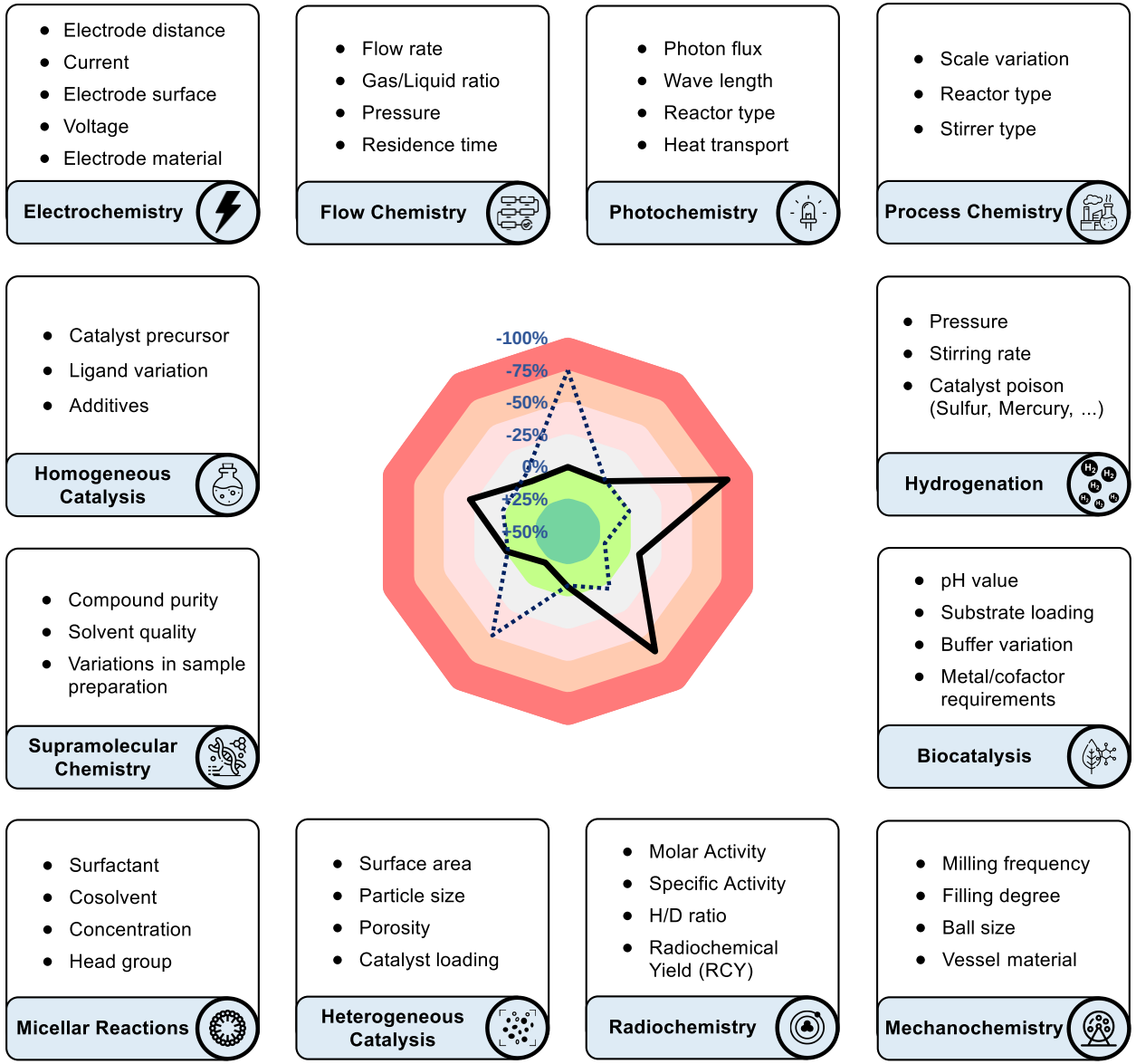
Prof. Dr. Frank Glorius


1

Chemistry is DATA science!!

Featuring “Rational design” & “Surprise findings” and “Success” & “Failure”

Question: Which data do we really care for?



Schäfer, Lückemeier, Glorius, *Chem. Sci.* **2024**, *15*, 14548-14555 .

Chemical Science

ROYAL SOCIETY OF CHEMISTRY

PERSPECTIVE

Published on 30 August 2024
Licensed under CC-BY 4.0

Check for updates

Cite this: *Chem. Sci.*, 2024, **15**, 14548

All publication charges for this article have been paid for by the Royal Society of Chemistry

Received 8th May 2024
Accepted 29th August 2024

DOI: 10.1039/d4sc03017f

rsc.li/chemical-science

Improving reproducibility through condition-based sensitivity assessments: application, advancement and prospect†

Felix Schäfer, Lukas Lückemeier and Frank Glorius *

The fluctuating reproducibility of scientific reports presents a well-recognised issue, frequently stemming from insufficient standardisation, transparency and a lack of information in scientific publications. Consequently, the incorporation of newly developed synthetic methods into practical applications often occurs at a slow rate. In recent years, various efforts have been made to analyse the sensitivity of chemical methodologies and the variation in quantitative outcome observed across different laboratory environments. For today's chemists, determining the key factors that really matter for a reaction's outcome from all the different aspects of chemical methodology can be a challenging task. In response, we provide a detailed examination and customised recommendations surrounding the sensitivity screen, offering a comprehensive assessment of various strategies and exploring their diverse applications by research groups to improve the practicality of their methodologies.

Reproducibility and practicality of synthetic protocols form fundamental pillars within the realm of experimental science.[1] From basic determinations, like selecting a synthetic procedure for starting materials, to strategic judgments such as project viability and the evaluation of scientific reports for publication or funding decisions, chemists depend on the insights provided within scientific literature. Understanding a reaction's sensitivity towards a diverse set of parameters is a key factor that ultimately leads to more sustainable and affordable chemistry,[2] enabling more robust[3] and greener reactions.[4]

The range of reactions utilised in synthetic disciplines like medicinal or process chemistry has seen little expansion, despite organic chemists consistently devising more efficient, selective, and innovative methods.[5] Particularly, technologies often found in academic research, like photo- and electrochemistry, are not frequently employed in complex syntheses or medicinal chemistry projects. Scaling up reactions is a common task in the chemical industry, while miniaturisation of chemical reactions smaller than milligram scale represents an entirely different challenge.[6] As high-throughput experimentation (HTE) becomes increasingly important, alongside advancements in data science and molecular machine learning, ensuring the quantitative accuracy of chemical reactions becomes increasingly vital for building reliable datasets.[7–9]

While the general outcome of chemical reactions tends to align with existing literature, there is often variability in quantitative results such as yield, selectivity, and purity. Systematically capturing these variations helps in integrating chemical literature into practical industrial processes and application across different research domains and academic labs. Even with detailed protocols provided in ESI,† reaction parameters may differ between labs with different chemists conducting the experiments and assessing this impact on the target value (*e.g.* yield) is non-trivial.

Depending on the resources, data and time available, a chemist may choose between different approaches in determining the sensitivity of a process or reaction. Recently, data science methods like molecular machine learning have emerged as rapid and cost-effective tools for predicting yields, but they rely on dense and unbiased datasets, which are often unavailable.[8,10] Conducting Design of Experiment (DoE) studies to explore every aspect of a reaction reveals its sensitivity, but this approach is likely the most time-consuming and costly.

Convenient experimental sensitivity assessment tools offer an advantageous alternative, providing researchers with efficient means to evaluate the response of chemical processes to various factors. These tools not only streamline the analysis process but also empower scientists to swiftly identify critical parameters influencing reactions, thereby facilitating more informed decision-making in experimental design and optimisation (Fig. 1).

In response to this challenge, the sensitivity screen was developed as a general tool to tackle the aforementioned reproducibility problems.[11] In this intuitive graphical evaluation, the effect of the parameters is plotted on a radar diagram, showing their deviation from the target value (usually yield)

Universität Münster, Organisch-Chemisches Institut, Corrensstraße 36, 48149 Münster, Germany. E-mail: glorius@uni-muenster.de

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4sc03017f

© 2024 The Author(s). Published by the Royal Society of Chemistry

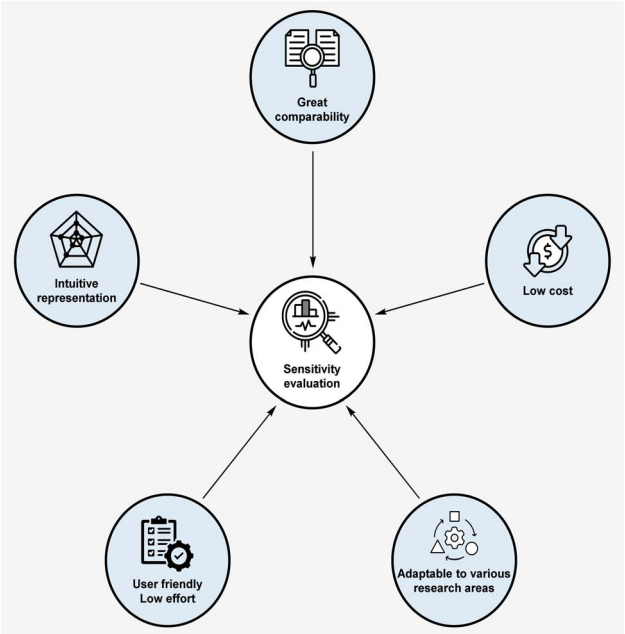


Fig. 1 Guidelines for conducting a universally applicable sensitivity assessment of reaction conditions.

compared to standard conditions. An updated version of the sensitivity screen with improved design and the option to use an additional target value (*e.g.* selectivity), together with a detailed user guide is provided in the ESI.† Our own experience showed that, if a synthetic protocol cannot be reproduced immediately, the vast number of possible sources of error is often overwhelming and thus troubleshooting can be incredibly time consuming. This especially applied to photochemical methods, as this rapidly developing field suffers from many different experimental setups, causing reproducibility issues. The basic idea to solve this issue was to use a screening approach to identify the crucial reaction parameters, which most significantly influence the reaction's behaviour. To do so, single reaction parameters are varied in both positive and negative directions, while all other parameters are kept constant. The resulting impact on the yield provides a valuable clue regarding which parameters to prioritize when troubleshooting reproducibility issues. To ensure user-friendliness, the method was designed to work with standard equipment and a stock solution approach, to help minimise experimental effort. As an intuitive graphical representation of the respective results, we introduced a colour-coded radar/spider diagram, visualising the results and highlighting the crucial parameters.

Our innovation was embraced by the scientific community and has since been widely adopted for applications beyond the originally suggested ones. Originating from a background in photochemistry, we initially considered only yield as the target value and parameters such as: temperature, concentration, oxygen and moisture levels, light intensity, and scale. The use of the sensitivity screen within the scientific community, including our own group, has led to its adoption modified with additional parameters and target values. While the target value was formerly only yield, researchers have suggested and applied conversion, product ratio, selectivity, ee, throughput, TON, radiochemical yield (RCY), molar/specific activity (A_m/A_s), *H*/*D*-ratio, purity, space-time yield and enzyme activity,[12] and the standardisation of these terms has been a central achievement of the chemical community (Fig. 2).[13–16] Especially for industrial processes with more than one desired product, the reaction can be steered towards a customised product distribution, which might alter given the demand of the chemical product. A radar

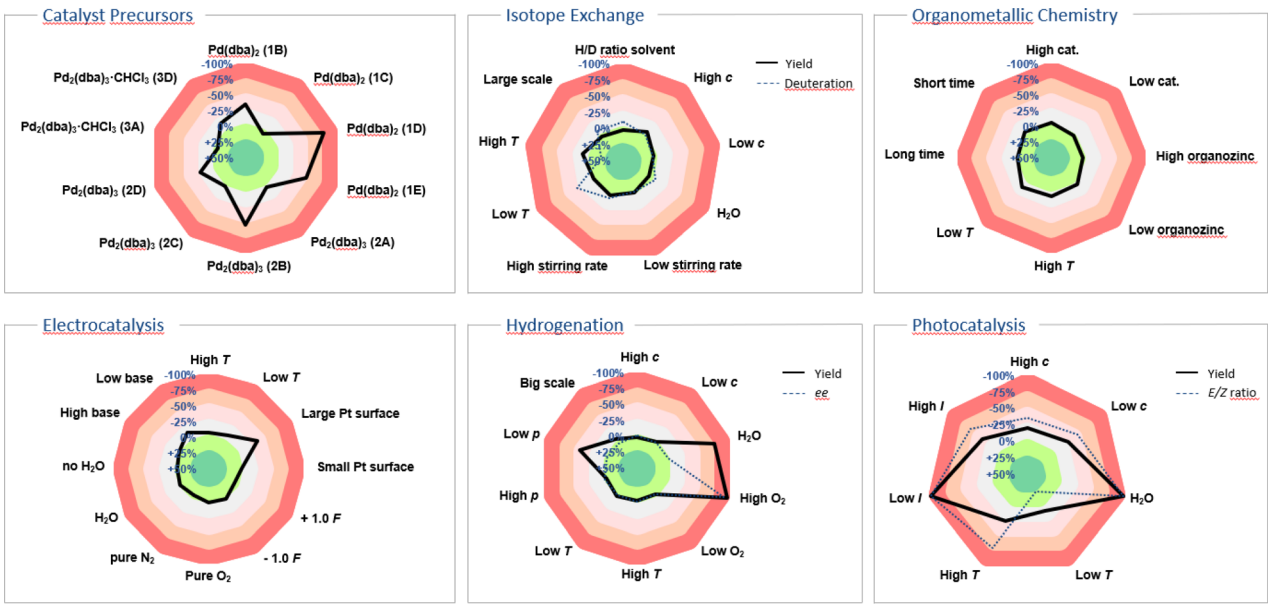


Fig. 2 Examples showcasing sensitivity screens tailored to diverse research fields.[13–16] In instances like hydrogenation, isotope exchange and photocatalysis the sensitivity screen was utilised to illustrate two target values, such as yield/deuterium incorporation, yield/enantiomeric excess and yield/*E*:*Z* ratio, along with the parameters' impact on them.

© 2024 The Author(s). Published by the Royal Society of Chemistry

diagram might also be beneficial for straightforward product ratio adjustment.

We have been pleasantly surprised by the diverse applications of the sensitivity screen, which have both refined the original parameter choices made by other scientists and extended its utility into various fields such as electrochemistry, hydrogenation, isotope chemistry, and asymmetric synthesis, among others, over the past five years.

In some cases, the critical adjustment of parameter values served as valuable feedback on the sensitivity screen: we acknowledge that our proposed range of ±10% concentration was frequently insufficient for detecting substantial yield changes beyond the detection limit of the analytical method employed. The widespread adoption of at least a ±50% concentration variation by many groups to demonstrate the impact on the target value reflects the valuable feedback from scientists in using our screening method.

For the other variables, we want to emphasise that our initial sensitivity screen was not intended to function as a comprehensive review of control reactions, where parameters should be varied significantly to demonstrate drastic yield changes. Instead, its purpose is to illustrate how minor alterations in various laboratory settings and applications can lead to significant yield variations, such as impurities like the presence of water in solvents and substrates, or minor oxygen ingress.

We continue to value the utilisation of the radar diagram depiction and sensitivity approach by all scientists, recognising that alternative and unconventional applications of this tool contribute significantly to collective efforts aimed at enhancing the reproducibility of reactions within the scientific community.

Reviewing and exploring the diverse application of the sensitivity screen across various domains of chemical synthesis and research, we have compiled a summary of its utilisation by various academic and industrial laboratories. Drawing inspiration from the extensive adaptations since our initial publication, we have included additional recommendations aimed at implementing the utilisation of sensitivity-based assessments across diverse research fields (Fig. 3).

General parameters

While variations in stoichiometry are typically addressed in optimisation tables, the sensitivity screen offers a rapid

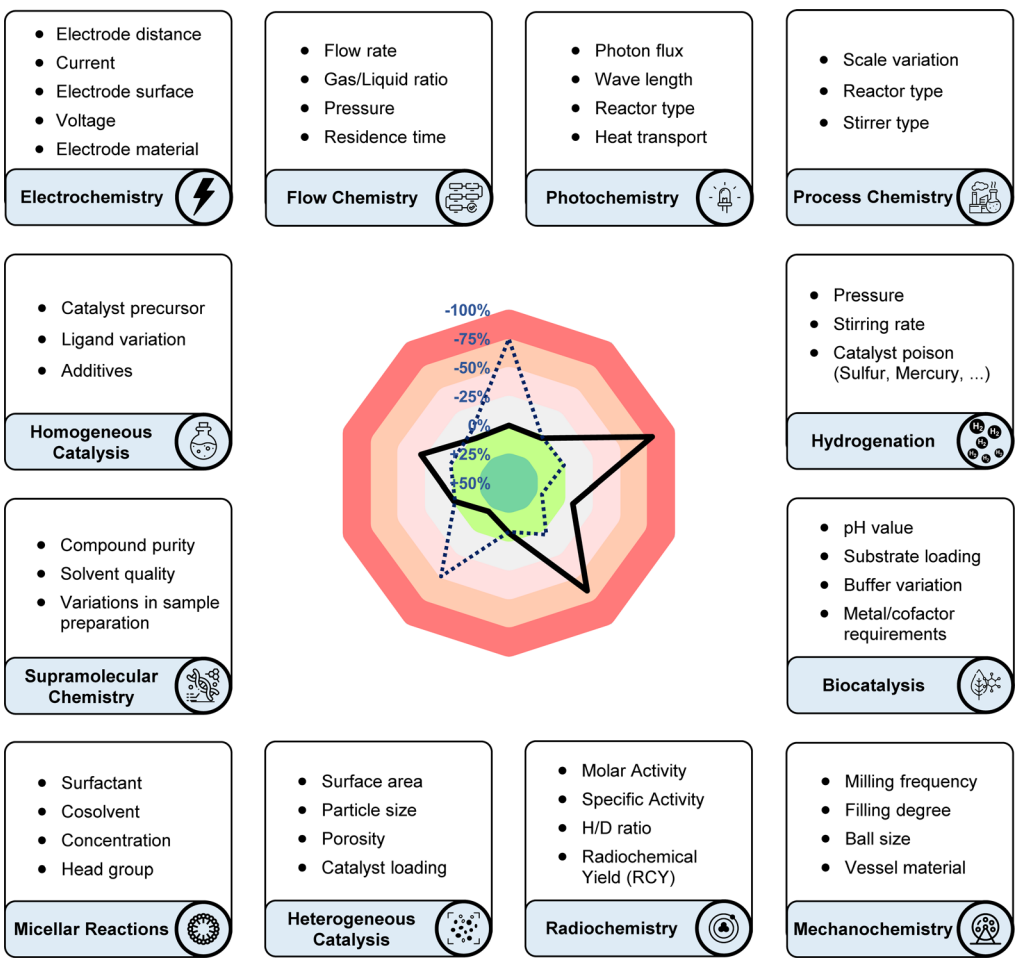


Fig. 3 Suggestions for parameters to include in a sensitivity screen across various research domains. While being established in some fields (hydrogenation, photochemistry, electrochemistry), additional recommendations for the application of the sensitivity screen in flow chemistry, mechanochemistry and biocatalysis aim to inspire scientists from these areas.

© 2024 The Author(s). Published by the Royal Society of Chemistry

assessment of their significance, particularly in multicomponent reactions.[17] Catalyst loading and the quantity of additives are intriguing parameters pertinent to scaling up processes and may have to be varied significantly depending on the intended application. Frequently neglected parameters in published laboratory procedures encompass stirring rate, reaction duration (including "overnight" reactions), and the purity of substrates and reagents.

Bull and co-workers introduced substrate purity as a new parameter.[18] They repurified their carboxylic acid starting material, enhancing its purity from 98.9% to 99.9%. This small adjustment, as revealed by their sensitivity screen, led to a notable +23% increase in yield, highlighting the significant impact even minor parameter changes can have. When Baran and co-workers established new anomeric amide reagents for electrophilic halogenation, they could demonstrate that these are very insensitive to variations in reaction parameters, even water and air intake, thus confirming their practicality.[19]

Electrochemistry

Electrochemical synthesis has experienced a renaissance in its use as a tool in organic chemistry, showcasing multifaceted utility. The availability of standardised commercial electrochemistry equipment, coupled with detailed reaction protocols, has significantly contributed to enhancing reproducibility in this field. Nevertheless, the myriad of additional parameters associated with electrochemical processes, such as variations in electrode surface area, materials, distances, supporting electrolytes, potential, or current, pose challenges in discerning significant deviations from the original procedure and those deemed negligible. Hilt and colleagues have effectively applied a sensitivity-based assessment in their methodology papers, leading to the conclusion that larger electrode surfaces significantly decrease the yield of their protocols while a smaller electrode distance can have variable effects, depending on the specific reaction system.[14,20] The use of alternating current/polarity instead of traditional direct current deserves special attention[21] since the nature of the alternating current can be widely varied in terms of frequency and nature of the potential curve, and future electrochemical methodologies might even use pulsating or variable current, parameters that can be included in a sensitivity screen.

Flow chemistry

Flow chemistry not only facilitates the continuous production of chemical reactions but also serves as an enabling technology for the transition and upscaling of newly developed methodologies, such as photochemistry and gas–liquid reactions.[22] Key parameters in flow chemistry, including flow rate, residence time, gas/liquid ratio, and pressure, play pivotal roles in process optimisation and control.

Photochemistry

Coming from a background rooted in photochemistry, our initial publication on sensitivity screening primarily focused on photochemical transformations and their corresponding parameters. Since its inception, it has been warmly embraced by the photochemical community, finding applications spanning from metallaphotoredox[23] to direct sensitisation[24] and energy transfer catalysis.[25] However, the quest for reproducibility[26] and comparative analysis in the sensitivity of photochemical reactions remains a daunting challenge. As a result, several notable approaches and critical evaluations have emerged from the scientific community. Ziegenbalg and colleagues, for instance, meticulously categorised parameters sorted by how commonly they are studied in publications, highlighting crucial factors such as reactor type, photocatalyst counter anion, heat and mass transport, temperature effects, reactor operation mode, and photon flux—some of which are often overlooked or inadequately explored.[27] Specifically, heat transport was found to play a crucial role and can dramatically change reaction outcomes when reactions are scaled up. In addition, several studies have examined and compared the design and use of commercial photoreactors, highlighting differences in irradiation uniformity and temperature control between different setups.[7,28] Despite these insightful analyses, the application of photocatalysis must navigate potential pitfalls in technology transfer,[29] particularly concerning the scalability of reactions.

Micellar reactions

The evolution of organic solvent selection in chemical processes reflects ongoing efforts to enhance reaction efficiency and sustainability.[30] This development has sparked widespread interest in alternative reaction media, with water emerging as a particularly compelling option.[31] Micellar catalysis and reactions have paved the way for exploring water as a promising alternative reaction medium.[32] Its use and optimisation introduces the key parameters surfactant and cosolvent, and collaborative studies have shown good reproducibility for several reaction classes.[33–36] We propose that conducting sensitivity assessments on these reactions, comparing different surfactants and conditions, will accelerate and simplify their industrial and academic adoption, especially since pioneering reactivity approaches on oil–water interfaces are on the rise.[37]

Mechanochemistry

Mechanochemistry shares closer ties with traditional thermal energy input methods compared to for example photo- or electrochemistry. Various parameters such as milling frequency, ball size, filling degree, and even the material composition of the vessel and balls can significantly influence reaction outcomes.[38]

Radiochemistry and isotope labelling

Radiochemical yield (RCY), molar activity (A_m) and specific activity (A_s)[39] could all be potential target values in sensitivity screen assessments for radiochemical reactions. Given the short half-life of many isotopes, the duration of reaction time becomes paramount in these instances, with sensitivity towards the timing of individual steps assisting scientists in optimising radiochemical yield across consecutive stages. Van Gemmeren and colleagues introduced solvent *H*/*D* ratio as a novel reaction

© 2024 The Author(s). Published by the Royal Society of Chemistry

parameter for β-C(sp^3)–H deuteration of free carboxylic acids, showcasing their reaction's adaptability to partially deuterated solvents.[15] In their sensitivity assessment, they combined target values for both product *H*/*D* ratio and yield, visually capturing the optimal compromise temperature in the corresponding radar diagram (Fig. 2).

Supramolecular chemistry

Meijer and co-workers have aptly summarised the most prevalent reproducibility issues in supramolecular chemistry, namely:[40] compound purity, solvent quality, and variations in sample preparation, such as slight temperature fluctuations or differing ramping speeds. Considering the sensitivity of many supramolecular applications to these minor changes, which may not necessarily affect the outcome measurably, we anticipate that the sensitivity screen will be beneficial but requires careful calibration for both the synthetic assembly of supramolecular complexes and their application.

Biocatalysis

Drawing parallels with the utilisation of additive-based robustness screens in biochemical contexts,[41,42] we anticipate significant potential for employing the sensitivity screen in chemical reactions being applied to biocatalysis, proximity labelling, and bioorthogonal chemistry settings.[43] Given the inherently higher complexity of biological systems and enzyme activity, we suggest incorporating: variations in buffer composition, pH sensitivity, metal/cofactor requirements, and substrate loading into a biochemical sensitivity screen. Successful transition from chemical methodology based screening approaches to other life sciences has so far found widespread use by the scientific community,[41,44] and the sensitivity screen has the potential to follow up on these examples.

Hydrogenation

Hydrogenation reactions are some of the most widely used transformations, reflected by the fact that 25% percent of all chemical value chains in industry include at least one hydrogenation step.[45] The hydrogen mass transport in between the different phases (gas–liquid, liquid–solid) has a pronounced impact on the rate and outcome of the reaction. Stirring rate as a single parameter has been shown to influence both yield[46] and enantioselectivity,[47] while in other cases oxygen had no effect on yield but severely lowered the *ee*.[48] Hydrogen pressure alters the concentration of dissolved hydrogen in the liquid phase, resulting in a higher reaction rate and potentially lower selectivity.[49] Catalyst poisoning can also be a serious problem, especially for industrial processes *e.g.* oil refining, and is the main reason why hydrodesulphurisation is often necessary beforehand.[50] Consequently, catalytic systems resistant to small amounts of poison are sought after,[51] and evaluating the poisoning sensitivity as a parameter represents vital information for synthetic chemists.

Homogeneous catalysis and cross-coupling reactions

Similarly, in homogeneous catalysis, variations in catalyst properties and quality can be substantial, especially when metal–ligand complexes are formed *in situ*. Gooßen and colleagues investigated the impact of different commercial palladium precatalyst batches and found considerable yield discrepancies among batches, with no single physical or spectroscopic descriptor offering reliable discrimination.[16] A depiction of parts of their study is shown utilising a sensitivity radar diagram in Fig. 2, demonstrating the significant influence of precatalyst selection on the Buchwald–Hartwig amination reaction.

Heterogeneous catalysis and surface chemistry

Given the inherently complex and unpredictable nature of heterogeneous catalytic reactions, sensitivity assessments of these reactions require careful consideration, particularly regarding the impact of reaction parameters. Catalyst samples can exhibit significant variations in particle size and properties. While precise control over these parameters may not always be feasible, conducting structure–activity relationship studies alongside sensitivity-based assessments, encompassing batches with diverse properties or from different manufacturers, facilitates straightforward comparisons of these parameters. These may encompass catalyst loading, porosity, surface area, particle size, and susceptibility to common catalyst poisons introduced by substrate impurities, particularly sulphur-containing compounds. Initiatives like rigour and reproducibility (R&R), in which Schweitzer and colleagues share best practices on reproducibility in thermal heterogeneous catalysis, are crucial, subject-specific and orthogonal approaches to the use of the sensitivity screen.[52]

Process chemistry

For industrial processes involving the simultaneous synthesis of multiple desired products, the reaction can be directed towards a desired product ratio, which may change based on the demand for the chemical product. In such cases, the operator might need to make adjustments, and a sensitivity screen can be useful for facilitating this process, allowing a straightforward visualisation of how parameters effect the new target value. The changes in overall dimensions resulting from variations in pilot scale, reactor type, and adjustable parameters can be assessed through sensitivity evaluation. For example, this assessment may involve comparing different types of stirrers, such as paddle, tooth, anchor, frame, ribbon screw, and turbine designs.[53] Analysing the characteristics of multiphase reactors, Papayannakos and co-workers have used the ratio of gas to liquid velocities as the central parameter across different reactors, providing a key feature for a chemical engineering approach towards sensitivity assessment.[54]

Conclusion and outlook

Sensitivity evaluation offers a rapid, cost-effective method for reaction evaluation, aiding in the adoption and implementation

© 2024 The Author(s). Published by the Royal Society of Chemistry

of newly developed methodologies by providing chemists and scientists from diverse fields with an overview of parameter impacts. The preceding discussion and suggestions regarding the versatile application of novel parameters in assessments across various domains are intended to inspire researchers in these fields to utilise the sensitivity screen. In doing so, they can shed light on specific challenges related to optimisation and sensitivity within their respective fields, thus providing valuable insights for the wider scientific community.

We want to emphasise that the sensitivity screen is a valuable and economic tool, but for an in-depth evaluation of reactions it should be followed up by more detailed approaches such as Design of experiments (DoE), mechanistic analysis, and advanced process optimisation. It serves as an initial, cost-effective and quick guide for chemists in these endeavours. Although standardised sensitivity assessments are desirable, the need for customisation of parameters specific to each field can hinder comparability, especially when the chemistry deviates significantly from traditional batch processes.

Our sensitivity screen has been widely embraced by chemists across various disciplines, generating a comprehensive dataset of sensitivity information. This data serves not only as a reference tool for researchers but also holds potential for informing data-driven models for sensitivity prediction in the future. We strongly believe that broader application across multiple fields will help advance the common goal of increasing transparency in newly published research and enhancing reproducibility within academia and beyond.

Data availability

This is a perspective manuscript and the original data is publicly available and can be found in the original manuscripts.

Author contributions

All authors contributed to the conceptualization and literature search, wrote the manuscript and were involved in revising, editing, and proofreading.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Generous financial support by the Deutsche Forschungsgemeinschaft (SPP 2363 - Molecular Machine Learning) and the European Research Council (Agreement No. 101098156, HighEnT) is gratefully acknowledged. We thank Dr Martin Ernst (BASF SE), Dr Felix Schäfers (INEOS Styrolution), Dr Jasper Tyler, Debanjan Rana, Dr Alessia Petti and Dr Julius Gemen (all University of Münster) for their scientific input and helpful discussions.

References

1 R. G. Bergman and R. L. Danheiser, *Angew. Chem., Int. Ed.*, 2016, **55**, 12548.
2 (*a*) I. T. Horváth, *Chem. Rev.*, 2018, **118**, 369; (*b*) W. Braje, F. Gallou, S. Handa and W. Tang, *ChemCatChem*, 2019, **11**, 5660.
3 J. Richardson, J. C. Ruble, E. A. Love and S. Berritt, *J. Org. Chem.*, 2017, **82**, 3741.
4 (*a*) B. A. de Marco, B. S. Rechelo, E. G. Tótoli, A. C. Kogawa and H. R. N. Salgado, *Saudi Pharm. J.*, 2019, **27**, 1; (*b*) J. Płotka-Wasylka and W. Wojnowski, *Green Chem.*, 2021, **23**, 8657.
5 D. C. Blakemore, L. Castro, I. Churcher, D. C. Rees, A. W. Thomas, D. M. Wilson and A. Wood, *Nat. Chem.*, 2018, **10**, 383.
6 A. Buitrago Santanilla, E. L. Regalado, T. Pereira, M. Shevlin, K. Bateman, L.-C. Campeau, J. Schneeweis, S. Berritt, Z.-C. Shi, P. Nantermet, Y. Liu, R. Helmy, C. J. Welch, P. Vachal, I. W. Davies, T. Cernak and S. D. Dreher, *Science*, 2015, **347**, 49.
7 B. Pijper, L. M. Saavedra, M. Lanzi, M. Alonso, A. Fontana, M. Serrano, J. E. Gómez, A. W. Kleij, J. Alcázar and S. Cañellas, *JACS Au*, 2024, **4**(7), 2585–2595.
8 F. Strieth-Kalthoff, F. Sandfort, M. Kühnemund, F. R. Schäfer, H. Kuchen and F. Glorius, *Angew. Chem., Int. Ed.*, 2022, **61**, e202204647.
9 (*a*) Z. Tu, T. Stuyver and C. W. Coley, *Chem. Sci.*, 2023, **14**, 226; (*b*) P. Piacentini, J. M. Fordham, E. Serrano, L. Hepp and M. Santagostino, *Org. Process Res. Dev.*, 2023, **27**, 798; (*c*) S. H. M. Mehr, M. Craven, A. I. Leonov, G. Keenan and L. Cronin, *Science*, 2020, **370**, 101; (*d*) S. M. Mennen, C. Alhambra, C. L. Allen, M. Barberis, S. Berritt, T. A. Brandt, A. D. Campbell, J. Castañón, A. H. Cherney, M. Christensen, D. B. Damon, J. Eugenio de Diego, S. García-Cerrada, P. García-Losada, R. Haro, J. Janey, D. C. Leitch, L. Li, F. Liu, P. C. Lobben, D. W. C. MacMillan, J. Magano, E. McInturff, S. Monfette, R. J. Post, D. Schultz, B. J. Sitter, J. M. Stevens, I. I. Strambeanu, J. Twilton, K. Wang and M. A. Zajac, *Org. Process Res. Dev.*, 2019, **23**, 1213; (*e*) P. S. Gromski, A. B. Henson, J. M. Granda and L. Cronin, *Nat. Rev. Chem*, 2019, **3**, 119; (*f*) S. A. Biyani, Y. W. Moriuchi and D. H. Thompson, *Chem.: Methods*, 2021, **1**, 323.
10 W. Beker, R. Roszak, A. Wołos, N. H. Angello, V. Rathore, M. D. Burke and B. A. Grzybowski, *J. Am. Chem. Soc.*, 2022, **144**, 4819.
11 L. Pitzer, F. Schäfers and F. Glorius, *Angew. Chem., Int. Ed.*, 2019, **58**, 8572.
12 (*a*) B. Zhang, T.-T. Li, Z.-C. Mao, M. Jiang, Z. Zhang, K. Zhao, W.-Y. Qu, W.-J. Xiao and J.-R. Chen, *J. Am. Chem. Soc.*, 2024, **146**, 1410; (*b*) Y. Liu, G. Luo, X. Yang, S. Jiang, W. Xue, Y. R. Chi and Z. Jin, *Angew. Chem., Int. Ed.*, 2020, **59**, 442.
13 (*a*) F. Zhang, H. S. Sasmal, C. G. Daniliuc and F. Glorius, *J. Am. Chem. Soc.*, 2023, **145**, 15695; (*b*) X.-D. Su, B.-B. Zhang, Q. Liu, J.-T. Cheng, Z.-X. Wang and X.-Y. Chen, *Org. Lett.*,

© 2024 The Author(s). Published by the Royal Society of Chemistry

2021, **23**, 8262; (*c*) L. M. Wickham, R. K. Dhungana and R. Giri, *ACS Omega*, 2023, **8**, 1060.
14 J. Fährmann and G. Hilt, *Angew. Chem., Int. Ed.*, 2021, **60**, 20313.
15 A. Uttry, S. Mal and M. van Gemmeren, *J. Am. Chem. Soc.*, 2021, **143**, 10895.
16 P. Weber, A. Biafora, A. Doppiu, H.-J. Bongard, H. Kelm and L. J. Gooßen, *Org. Process Res. Dev.*, 2019, **23**, 1462.
17 J. Hou, X. Li, K. Yan, L. Zhang, T.-P. Loh and P. Xie, *Chem. Sci.*, 2024, **15**, 1143.
18 M. A. J. Dubois, J. J. Rojas, A. J. Sterling, H. C. Broderick, M. A. Smith, A. J. P. White, P. W. Miller, C. Choi, J. J. Mousseau, F. Duarte and J. A. Bull, *J. Org. Chem.*, 2023, **88**, 6476.
19 Y. Wang, C. Bi, Y. Kawamata, L. N. Grant, L. Samp, P. F. Richardson, S. Zhang, K. C. Harper, M. D. Palkowitz, A. Vasilopoulos, M. R. Collins, M. S. Oderinde, C. C. Tyrol, D. Chen, E. A. LaChapelle, J. B. Bailey, J. X. Qiao and P. S. Baran, *Nat. Chem.*, 2024, **16**, 1539.
20 (*a*) J. Strehl and G. Hilt, *Eur. J. Org Chem.*, 2022, e202101007; (*b*) E. Babaoglu and G. Hilt, *Chem.–Eur. J.*, 2020, **26**, 8879.
21 A. G. Wills, D. L. Poole, C. M. Alder and M. Reid, *ChemElectroChem*, 2020, **7**, 2771.
22 (*a*) M. B. Plutschack, B. Pieber, K. Gilmore and P. H. Seeberger, *Chem. Rev.*, 2017, **117**, 11796; (*b*) L. Capaldo, Z. Wen and T. Noël, *Chem. Sci.*, 2023, **14**, 4230; (*c*) C. A. Hone and C. O. Kappe, *Chem.: Methods*, 2021, **1**, 454.
23 A. Yanez and A. D. M. Jeanneret, *Synlett*, 2024, **35**, 1728–1732.
24 F. Li, W. F. Zhu, C. Empel, O. Datsenko, A. Kumar, Y. Xu, J. H. M. Ehrler, I. Atodiresei, S. Knapp, P. K. Mykhailiuk, E. Proschak and R. M. Koenigs, *Science*, 2024, **383**, 498.
25 Y.-S. Jiang, S.-S. Li, X.-L. Luo, L.-N. Chen, D.-N. Chen and P.-J. Xia, *Org. Lett.*, 2023, **25**, 6671.
26 S. Cañellas, M. Nuño and E. Speckmeier, *Nat. Commun.*, 2024, **15**, 307.
27 D. Ziegenbalg, A. Pannwitz, S. Rau, B. Dietzek-Ivanšić and C. Streb, *Angew. Chem., Int. Ed.*, 2022, **61**, e202114106.
28 (*a*) T. D. Svejstrup, A. Chatterjee, D. Schekin, T. Wagner, J. Zach, M. J. Johansson, G. Bergonzini and B. König, *ChemPhotoChe*, 2021, **5**, 808; (*b*) E. Kayahan, M. Jacobs, L. Braeken, L. C. Thomassen, S. Kuhn, T. van Gerven and M. E. Leblebici, *Beilstein J. Org. Chem.*, 2020, **16**, 2484; (*c*) P. Kant, S. Liang, M. Rubin, G. A. Ozin and R. Dittmeyer, *Joule*, 2023, **7**, 1347.
29 T. Noël and E. Zysman-Colman, *Chem Catal.*, 2022, **2**, 468.
30 C. Reichardt and T. Welton, *Solvents and solvent effects in organic chemistry*, Wiley-VCH, Weinheim Germany, 2011.
31 B. H. Lipshutz, F. Gallou and S. Handa, *ACS Sustain. Chem. Eng.*, 2016, **4**, 5838.
32 G. La Sorella, G. Strukul and A. Scarso, *Green Chem.*, 2015, **17**, 644.
33 C. Krell, R. Schreiber, L. Hueber, L. Sciascera, X. Zheng, A. Clarke, R. Haenggi, M. Parmentier, H. Baguia, S. Rodde and F. Gallou, *Org. Process Res. Dev.*, 2021, **25**, 900.
34 F. Gallou, N. A. Isley, A. Ganic, U. Onken and M. Parmentier, *Green Chem.*, 2016, **18**, 14.
35 N. Borlinghaus, T. N. Ansari, L. H. von Garrel, D. Ogulu, S. Handa, V. Wittmann and W. M. Braje, *Green Chem.*, 2021, **23**, 3955.
36 R. J. Fox, J. D. Bailey, J. V. Obligacion, N. Borlinghaus, W. M. Braje, X. Li, S. Mukherjee, A. Schoen, T. B. Towne and S. Vukelić, *Org. Process Res. Dev.*, 2024, **28**, 978–1002.
37 (*a*) Y.-M. Tian, W. Silva, R. M. Gschwind and B. König, *Science*, 2024, **383**, 750; (*b*) M. Cortes-Clerget, J. Yu, J. R. A. Kincaid, P. Walde, F. Gallou and B. H. Lipshutz, *Chem. Sci.*, 2021, **12**, 4237.
38 (*a*) R. T. O'Neill and R. Boulatov, *Nat. Rev. Chem*, 2021, **5**, 148; (*b*) J. L. Howard, Q. Cao and D. L. Browne, *Chem. Sci.*, 2018, **9**, 3080; (*c*) A. M. Belenguer, G. I. Lampronti and J. K. M. Sanders, *J. Visualized Exp.*, 2018, **131**, e56824.
39 H. H. Coenen, A. D. Gee, M. Adam, G. Antoni, C. S. Cutler, Y. Fujibayashi, J. M. Jeong, R. H. Mach, T. L. Mindt, V. W. Pike and A. D. Windhorst, *Ann. Nucl. Med.*, 2018, **32**, 236.
40 T. Schnitzer, M. D. Preuss, J. van Basten, S. M. C. Schoenmakers, A. J. H. Spiering, G. Vantomme and E. W. Meijer, *Angew. Chem., Int. Ed.*, 2022, **61**, e202206738.
41 L. Anhäuser, M. Teders, A. Rentmeister and F. Glorius, *Nat. Protoc.*, 2019, **14**, 2599.
42 (*a*) K. D. Collins and F. Glorius, *Nat. Chem.*, 2013, **5**, 597; (*b*) K. D. Collins and F. Glorius, *Acc. Chem. Res.*, 2015, **48**, 619; (*c*) K. D. Collins, A. Rühling, F. Lied and F. Glorius, *Chem.–Eur. J.*, 2014, **20**, 3800; (*d*) T. Gensch, M. Teders and F. Glorius, *J. Org. Chem.*, 2017, **82**, 9154.
43 (*a*) M. A. Emmanuel, S. G. Bender, C. Bilodeau, J. M. Carceller, J. S. DeHovitz, H. Fu, Y. Liu, B. T. Nicholls, Y. Ouyang, C. G. Page, T. Qiao, F. C. Raps, D. R. Sorigué, S.-Z. Sun, J. Turek-Herman, Y. Ye, A. Rivas-Souchet, J. Cao and T. K. Hyster, *Chem. Rev.*, 2023, **123**, 5459; (*b*) E. L. Bell, W. Finnigan, S. P. France, A. P. Green, M. A. Hayes, L. J. Hepworth, S. L. Lovelock, H. Niikura, S. Osuna, E. Romero, K. S. Ryan, N. J. Turner and S. L. Flitsch, *Nat. Rev. Methods Primers*, 2021, **1**, 46; (*c*) U. Hanefeld, F. Hollmann and C. E. Paul, *Chem. Soc. Rev.*, 2022, **51**, 594.
44 C. Wan, Z. Hou, D. Yang, Z. Zhou, H. Xu, Y. Wang, C. Dai, M. Liang, J. Meng, J. Chen, F. Yin, R. Wang and Z. Li, *Chem. Sci.*, 2023, **14**, 604.
45 L. Zhang, M. Zhou, A. Wang and T. Zhang, *Chem. Rev.*, 2020, **120**, 683.
46 (*a*) P. A. Rautanen, J. R. Aittamaa and A. O. I. Krause, *Ind. Eng. Chem. Res.*, 2000, **39**, 4032; (*b*) C. Gelis, A. Heusler, Z. Nairoukh and F. Glorius, *Chem.–Eur. J.*, 2020, **26**, 14090.
47 Y. Sun, R. N. Landau, J. Wang, C. LeBlond and D. G. Blackmond, *J. Am. Chem. Soc.*, 1996, **118**, 1348.
48 D. Moock, T. Wagener, T. Hu, T. Gallagher and F. Glorius, *Angew. Chem., Int. Ed.*, 2021, **60**, 13677.
49 (*a*) T. Wagener, M. Pierau, A. Heusler and F. Glorius, *Adv. Synth. Catal.*, 2022, **364**, 3366; (*b*) Y. Xu, Y. Luo, J. Ye, Y. Deng, D. Liu and W. Zhang, *J. Am. Chem. Soc.*, 2022, **144**, 20078.
50 I. Mochida and K.-H. Choi, *J. Jpn. Pet. Inst.*, 2004, **47**, 145.

© 2024 The Author(s). Published by the Royal Society of Chemistry

51 L. Lückemeier, T. de Vos, L. Schlichter, C. Gutheil, C. G. Daniliuc and F. Glorius, *J. Am. Chem. Soc.*, 2024, **146**, 5864.
52 (*a*) D. Esposito, *Nat. Catal.*, 2022, **5**, 658; (*b*) *Addressing Rigor and Reproducibility in Thermal, Heterogeneous Catalysis*, ed. N. M. Schweitzer, R. Gounder and R. M. Rioux, Zenodo, 2023.
53 (*a*) *Handbuch Chemische Reaktoren*, ed. W. Reschetilowski, Springer Spektrum Berlin, Heidelberg, 2019; (*b*) C. Hu, *J. Flow Chem.*, 2021, **11**, 243; (*c*) S. A. Thorat, L. Chen and H. Yang, *J. Chem.*, 2022, **11**, 1.
54 A. Vonortas, A. Hipolito, M. Rolland, C. Boyer and N. Papayannakos, *Chem. Eng. Technol.*, 2011, **34**, 208.

© 2024 The Author(s). Published by the Royal Society of Chemistry

„Reaction development: A student‘s checklist“



Jasper Tyler*, Dirk Trauner*, FG*: Chem. Soc. Rev. 2025, 54, 3272.

„Reaction development: A student‘s checklist“

So You've Discovered a Reaction...

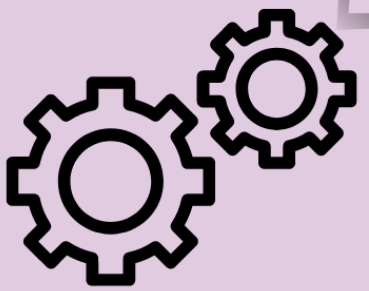
What's Next?



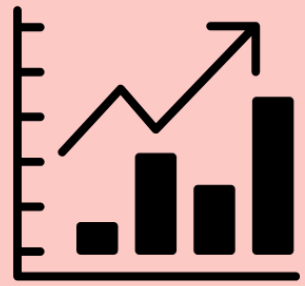
Take Stock



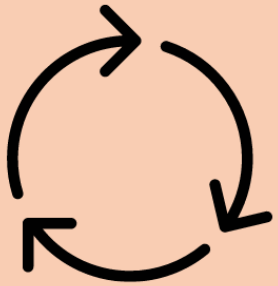
Kinetics & Thermodynamics



Mechanism



Optimisation



Catalysis



Scope



Applications

Chem Soc Rev

TUTORIAL REVIEW

Published on 06 February 2025
Licence and permissions

Check for updates

Cite this: *Chem. Soc. Rev.*, 2025, **54**, 3272

Received 18th October 2024

DOI: 10.1039/d4cs01046a

rsc.li/chem-soc-rev

Reaction development: a student's checklist†

Jasper L. Tyler, *[a] Dirk Trauner *[b] and Frank Glorius *[a]

So you've discovered a reaction. But how do you turn this new discovery into a fully-fledged program that maximises the potential of your novel transformation? Herein, we provide a student's checklist to serve as a helpful guide for synthesis development, allowing you to thoroughly investigate the chemistry in question while ensuring that no key aspect of the project is overlooked. A wide variety of the most illuminating synthetic and spectroscopic techniques will be summarised, in conjunction with literature examples and our own insights, to provide sound justifications for their implementation towards the goal of developing new reactions.

Key learning points

1. How to both conceive of and discover new reactions.
2. How to systematically investigate a reaction from initial hit to publication.
3. The principles behind some of the most prominent mechanistic experiments.
4. The importance of unbiased and standardised data generation.
5. How to explore the potential applications of novel methodology.

Introduction: how to discover a reaction

The invention, discovery, and conceptualisation of new reactions lies at the very heart of organic chemistry. Despite the immense body of work performed over the past few centuries, progress is still continually driven by the need to access synthetic tools that can explore regions of previously uncharted chemical space. Such discoveries open new horizons in the development of pharmaceuticals, agrochemicals, and functional materials, and can also unveil new reaction mechanisms that provide fundamentally new ways to think about synthesis. The methods by which we achieve this goal are constantly evolving and the wider implementation of artificial intelligence (AI) within this field is expected to modernise, if not entirely revolutionise, the strategies with which we discover new reactions. Regardless, human input in the form of creativity and chemical intuition will still be vital components in this process moving forward.

[a] *University of Muenster, Institute for Organic Chemistry, Corrensstrasse 36, 48149 Muenster, Germany. E-mail: glorius@uni-muenster.de*
[b] *Department of Chemistry, University of Pennsylvania, Philadelphia, Pennsylvania, 19104, USA. E-mail: dtrauner@upenn.edu*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4cs01046a

But how does one go about discovering – or inventing – new reactions given the vastness of chemical space, the long history of organic chemistry, and the level of sophistication it has already achieved? There are several ways to do this:

You develop it from the older literature

Scores of valuable reactions are buried deep in the scientific literature, often as a sidenote or in an unfamiliar language, and can be re-evaluated through careful reading. Where perhaps the transformation was formerly thought of as useless, recent discoveries, brought to light well after the original publication was conceived, can engender a reaction with significantly enhanced applicability.[1] Additionally, many reactions will have not been explored in depth due to technical limitations, a different focus of the initial study, product instability, or if the compound was observed as a minor byproduct.

Regardless of whether a specific mechanistic regime may have direct precedent, new discoveries that were not fully realised by the original authors can constitute vital contributions to the field. A recent example of this principle was demonstrated by Burns and coworkers in their development of the photochemical conversion of aryl azides to aminopyridines (Fig. 1(A)).[2] This transformation, originally observed in low yields as part of a complex mixture by Sundberg and coworkers in 1972,[3] was significantly improved upon in terms of the yield, selectivity, conditions and scope. This procedural

This journal is © The Royal Society of Chemistry 2025

modification transformed the reaction from an academic curiosity to viable tool for aromatic structural editing.

You discover it accidentally

Serendipity was introduced as a word into the scientific literature by R. B. Woodward, a chemist, and serendipitous discoveries have long been a cornerstone of scientific research.[4] The importance of thoroughness and careful investigation is best evidenced by the countless examples of ground-breaking research that have originated from an unexpected observation.[5,6] For example, the Friedel–Crafts reaction, discovered in 1877, was not only an unintended outcome of an attempt to synthesise amyl iodide from amyl chloride,[7] but the formation of an aryl alkylation product had even been previously observed by scientists that had been unable to explain the results and consequently overlooked the reaction.[8–10] The importance of Friedel and Crafts' thorough scientific investigation was highlighted in Olah's 1963 book ''Friedel–Crafts and Related Reactions'' where it was stated: ''*It cannot be denied that many important scientific discoveries were made by chance, but it is essential that a person engaged in the research must have a keen sense of observation and creative ability if any discovery is to be developed and not remain unknown or passed over*''.[11] A more recent example of serendipitous discovery was made by Trauner and coworkers in an attempt to catalyse 6π-electrocyclisation reactions (Fig. 1(B)).[12] Exposure of a tetraenyl carboxylic ester **1** to a Lewis acid did not afford the intended cyclohexadiene but rather bicyclo[3.1.0]hexene **2**. Further investigation determined that this transformation was general and could be subsequently applied to natural product synthesis.[13]

Although the concept of serendipity appears to be one which cannot be intentionally targeted, in recent years it has been shown that serendipity can be ''accelerated'' using a combination of high-throughput experimentation and automated workflows to effectively exploit chance in reaction discovery.[14] This was highlighted in a study by MacMillan and coworkers who randomly paired a range of ''benign'' substrates containing common functional groups using an automation platform and subjected them to photocatalytic conditions.[15] GC–MS analysis of the resulting mixtures could then be used to identify products that represent novel reactivity regimes. A similar workflow was also employed independently by Hartwig and coworkers for transition metal-catalysed cross-coupling reactions,[16] however, it must be noted that in these protocols, selection bias and chemical intuition is still essential for the choice of substrates and catalysts.[17–19]

You think of it by analogy

Perhaps the most challenging way to establish new reactions is to invent them without having a specific application in mind. However, a new reaction may follow the fundamental mechanism of an established transformation in which a key fragment is



Jasper L. Tyler

Jasper L. Tyler received his MChem degree under the supervision of Prof. Joseph P. A. Harrity at The University of Sheffield. In 2019 he moved to the University of Bristol and completed his PhD under the supervision of Professor Varinder Aggarwal. In 2023 he joined the group of Professor Frank Glorius at the University of Münster as a Humboldt research fellow. His current research focuses on skeletal editing transformations, strain-release reactivity and photochemistry.



Dirk Trauner

Dirk Trauner is a Professor of Chemistry and Pharmacology at the University of Pennsylvania. The broad objective of his research is to demonstrate the awesome power of chemical synthesis with challenging target molecules and to use it toward the establishment of synthetic biological pathways. He is well known for his pioneering contributions to optogenetics and photopharmacology, using synthetic photoswitches to confer light-sensitivity to a broad range of biological targets. Despite his extensive forays into biology and physiology he remains an organic chemist at heart, as evidenced by the synthesis (and anticipation) of more than one hundred twenty natural products.



Frank Glorius

Frank Glorius is Full Professor of Organic Chemistry at the University of Münster, Germany. His diverse and interdisciplinary research program focuses on the development of new concepts for various areas of catalysis, such as skeletal editing, photocatalysis, C–H activation, N-heterocyclic carbenes (NHCs) in organocatalysis, (asymmetric) arene hydrogenation, and as unique surface modifiers for materials science and (enantioselective) heterogeneous catalysis. In addition, his group is developing molecular tools for the study and modification of biological systems and battery systems. Finally, data science and molecular machine learning are other current topics that the group is focusing on.

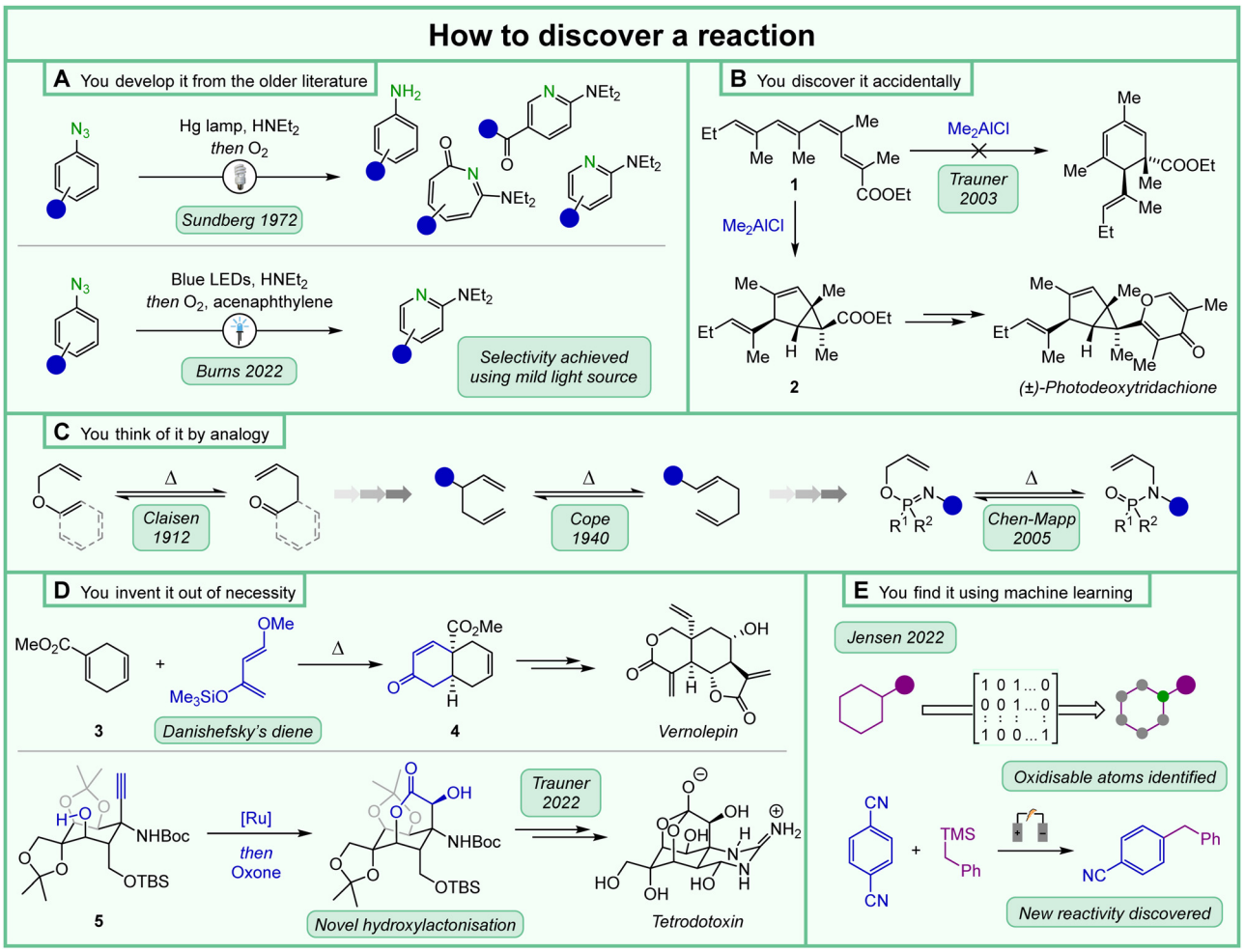


Fig. 1 How to discover a reaction.

substituted for another, to deliver an entirely different class of products. Such modifications can involve substitutions of the minimal synthon with heteroatoms, a change in the degree of unsaturation, or the recognition of isolobal relationships in transition metal complexes. Examples of this include the numerous transformations that proceed *via* a [3,3] sigmatropic shift mechanism, that was first reported by Claisen in 1912,[20] such as the Johnson- and Ireland–Claisen, Cope, oxy-Cope, and Chen-Mapp rearrangements (Fig. 1(C)).[21,22] A further example can be found in the work of Huisgen and coworkers on 1,3-dipolar cycloadditions. Here, a thorough analysis of previously known reactions (*e.g.* thermal azide–alkyne cycloadditions) and the recognition of the underlying electronic patterns, led to the invention of entirely new 1,3-dipoles, such as thiocarbonyl ylides and nitrile imines, and their associated cycloaddition reactions.[23]

You invent it out of necessity

Novel methodology can also arise when known chemistry is unable to overcome a specific barrier to the synthesis of a target compound, necessitating the invention of new classes of chemical reactions. This was indeed the situation during Danishefsky and coworkers investigations into the synthesis of sesquiterpenes such as vernolepin, where *cis*-decalin **4** was targeted as an intermediate (Fig. 1(D)).[24] It was envisaged that a highly reactive 1,3-dioxygenated diene would be capable of performing the desired Diels–Alder reaction with **3**, a relatively poor dienophile, which could then undergo elimination to access the targeted enone fragment.[25,26] The species that was most successful in facilitating this transformation, the reagent now known as "Danishefsky's diene",[27] has since been employed countless times in the preparation of natural products.[28,29]

A more modern example of this overall principle was reported during a recent synthesis of the neurotoxin tetrodotoxin, where traditional π-acid catalysis failed to give the required selectivity in the cyclisation of alkynyl alcohol **5** (Fig. 1(D)).[30] To surmount this issue, a novel ruthenium-catalysed hydroxylactonisation reaction was developed, capable of assembling a key fragment in the dioxa-adamantane core and representing a useful synthetic method for the construction of complex lactones.

You find it using machine learning

A technique that is becoming increasingly important within the field of methodology development is the use of machine learning algorithms.[31–34] Machine learning is a branch of

 This journal is © The Royal Society of Chemistry 2025

artificial intelligence that allows computers to learn patterns using a set of training data and then make predictions on new data sets. This approach is most effectively utilised when the targeted output is a limited number of well-defined parameters and, as a consequence, machine learning algorithms have been demonstrated to be a valuable tool in areas such as yield[35] and reaction condition predictions,[36] catalyst/ligand design[37] and retrosynthetic analysis.[38] However, the autonomous discovery of entirely new reaction protocols provides significant challenges such as the ability to encapsulate multiple physical and chemical properties into molecular descriptors, the accurate prediction of reaction success between coupling partners and the assessment of reaction novelty with regards to the vast amount of scientific literature.[39]

Despite this, when targeting a specific reactivity regime, employing a combination of machine learning algorithms, automated synthesis platforms,[40] and chemical intuition to simplify the problem, has allowed a handful of new chemical transformations to be discovered.[41] For example, Jensen and coworkers demonstrated that a machine learning model, trained on a data set obtained *via* automated experimentation, was able to predict whether a compound was capable of undergoing oxidation and subsequently reacting with 1,4-dicyanobenzene under electrochemical conditions, leading to the discovery of several new arylation reactions (Fig. 1(E)).[42] Although examples of such a strategy are currently few in number, the possibility of accessing entirely new reactivity modes with minimal human input is extremely attractive and could lead to a revolutionary new age of chemical discovery.[43–45]

What do you do now?

Once a reaction has been discovered, attention can then be directed toward comprehensively demonstrating the potential of the novel transformation. Herein is provided a checklist to serve as a helpful guide for reaction development. Given the significant number of variables associated with individual synthetic methodology studies, all of the themes discussed in this tutorial will not be directly applicable to every new reaction and are not a necessity for publication. Indeed, the ability to address many of the points raised is dependent on factors such as time and resource availability. However, understanding the fundamental principles underlying each topic, and assessing their potential applicability to the newly discovered reaction, may provide the key to maximising the impact and broadening the readership of your research.

1. Take stock

Have you made the desired compound?

Structural assignment is one of the central components of synthetic methodology projects. The gold standard of organic molecule characterisation is often stated to be X-ray crystallography, which provides the potential to simultaneously observe both the molecular formula, connectivity and (absolute) configuration of the species under investigation. However, this technique is not entirely infallible,[46,47] and its usage can be constrained as a result of cost, restricted access to the requisite equipment or the non-crystalline nature of the compound. In most cases, organic compound assignment is performed primarily using a combination of nuclear magnetic resonance (NMR) spectroscopy and high-resolution mass spectrometry (HRMS), which can also be complicated by overlapping peaks and fragmentation pathways, respectively. This being said, the constantly increasing levels of NMR spectrometer resolution and the development of novel spectroscopic and ionisation techniques has allowed increasingly more accurate structural assignments.[48,49] As no individual technique is infallible, it is essential that when faced with a molecule that is unknown within the current literature, a combination of as many available characterisation techniques is employed to ensure the correct conclusions are drawn.

For highly complex molecules, such as newly discovered natural products, computational calculations can also be a powerful method to verify a predicted structure.[50] An example of such a technique being used to assist characterisation includes Rychnovsky's study of hexacyclinol, which led to a complete reassignment of the original structure on the basis of calculated ^{13}C chemical shifts (Fig. 2(A)).[51–53] In cases where little information is known about the nature of the analyte, machine learning algorithms that can generate predictions of molecular structure, based on spectroscopic data, has the potential of greatly expediting the process of structural assignment.[54,55] Although significant progress has been made in recent years, this technique is yet to be widely employed, due in part to the limited quantity of reliable data sets that new models can be trained on.[56]

Are all reaction conditions necessary?

It is highly likely that the conditions that were first employed during the discovery of the reaction are not those that could be considered as the most efficient. Therefore, huge savings on time and energy can be made by performing selective test reactions that can establish what specific conditions are actually necessary before undertaking an extensive optimisation study. The most important factor here is time, a chemists most precious resource. Although one can simply use TLC analysis or quench a series of reactions at various time points to determine when the maximum yield is reached, it may be more profitable to employ *in situ* monitoring, which can give a more precise picture of the reaction profile.[57,58]

This was highlighted by Coldham and O'Brien who employed *in situ* IR spectroscopy to determine that the deprotonation of Boc-2-phenylpyrrolidine (**6**) at −78 °C only goes to partial completion due to the slow rotation of the carbamate group which sterically inhibits the approach of the base in rotamer **6-a** (Fig. 2(B)).[59] To facilitate the interconversion of the starting material rotamers, the reaction was performed at 0 °C, both improving the yield of functionalised product **7** and establishing that the reaction time could be reduced to 5 minutes.

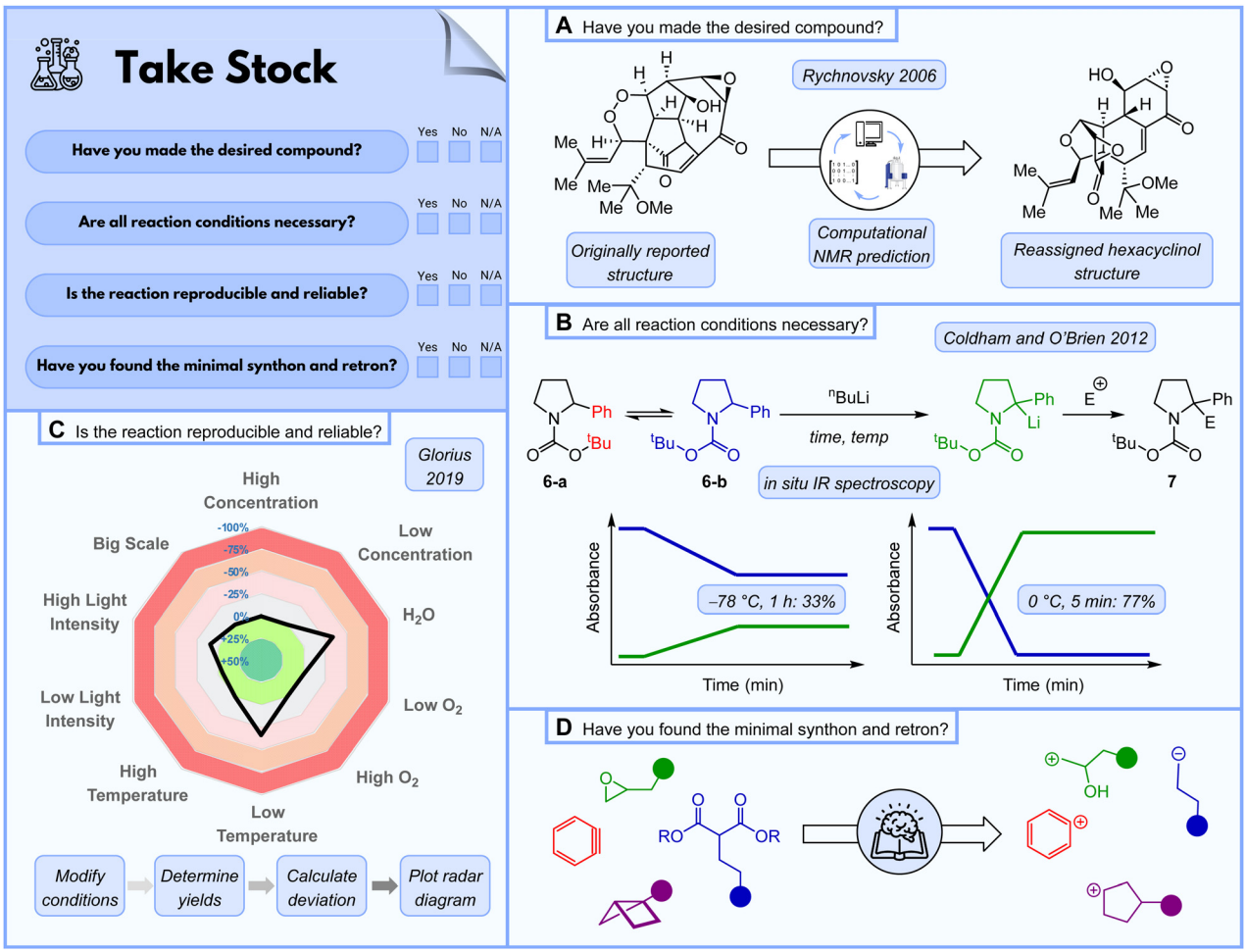


Fig. 2 Take stock checklist.

Is the reaction reproducible and reliable?

The most ideal transformation can only have limited applications if the reaction is not reproducible and reliable, *i.e.* the protocol can be performed multiple times, and by different chemists, whilst still obtaining a consistent result in terms of yield and purity.[60] It may also be highly beneficial to determine the sensitivity of any new transformation to variations in the conditions, which can provide insights on how transferable the reaction might be to an industrial setting. In this regard, Glorius and coworkers have developed a user-friendly and systematic sensitivity screen protocol that allows the investigation and standardised reporting of a reaction's sensitivity (Fig. 2(C)).[61] By performing 10 test experiments that probe the reaction sensitivity to small changes in various parameters such as scale, temperature, concentration and the presence of additives, then reporting the data in a radar diagram, one can easily identify the conditions that impact the transformation most significantly. This can ultimately give vital information on which conditions should be most carefully controlled and assist in the troubleshooting of any previously discovered issues.

Have you found the minimal synthon and retron?

Finally, it is worth considering what happens when the reaction is stripped away to the minimal synthons, *i.e.*, to what extent can one simplify the substrate. Carrying out this analysis can indicate whether simpler starting materials or reagents can be employed under the same conditions to broaden its applicability (Fig. 2(D)). Conversely, the minimal retron needs to be defined and its occurrence should be checked in chemical databases. This will yield target molecules of interest to which the reaction could be applied, demonstrating its synthetic value. Performing this thought experiment may also present the possibility of applying intramolecular reactivity to an intermolecular system in order to maximise the potential for diversification. However, when adapting a transformation from an intramolecular to an intermolecular system (or *vice versa*) one must consider that, although the targeted product may be obtained, this may be associated with a change in the reaction rate, the rate-limiting step or even may involve a switch in reaction mechanism.

2. Kinetics and thermodynamics

Can the rate law be determined?

The rate law is a mathematical description of a reaction's rate as a function of the concentration of the reactants present in the rate limiting step and the rate constant (k), typically in the

 This journal is © The Royal Society of Chemistry 2025

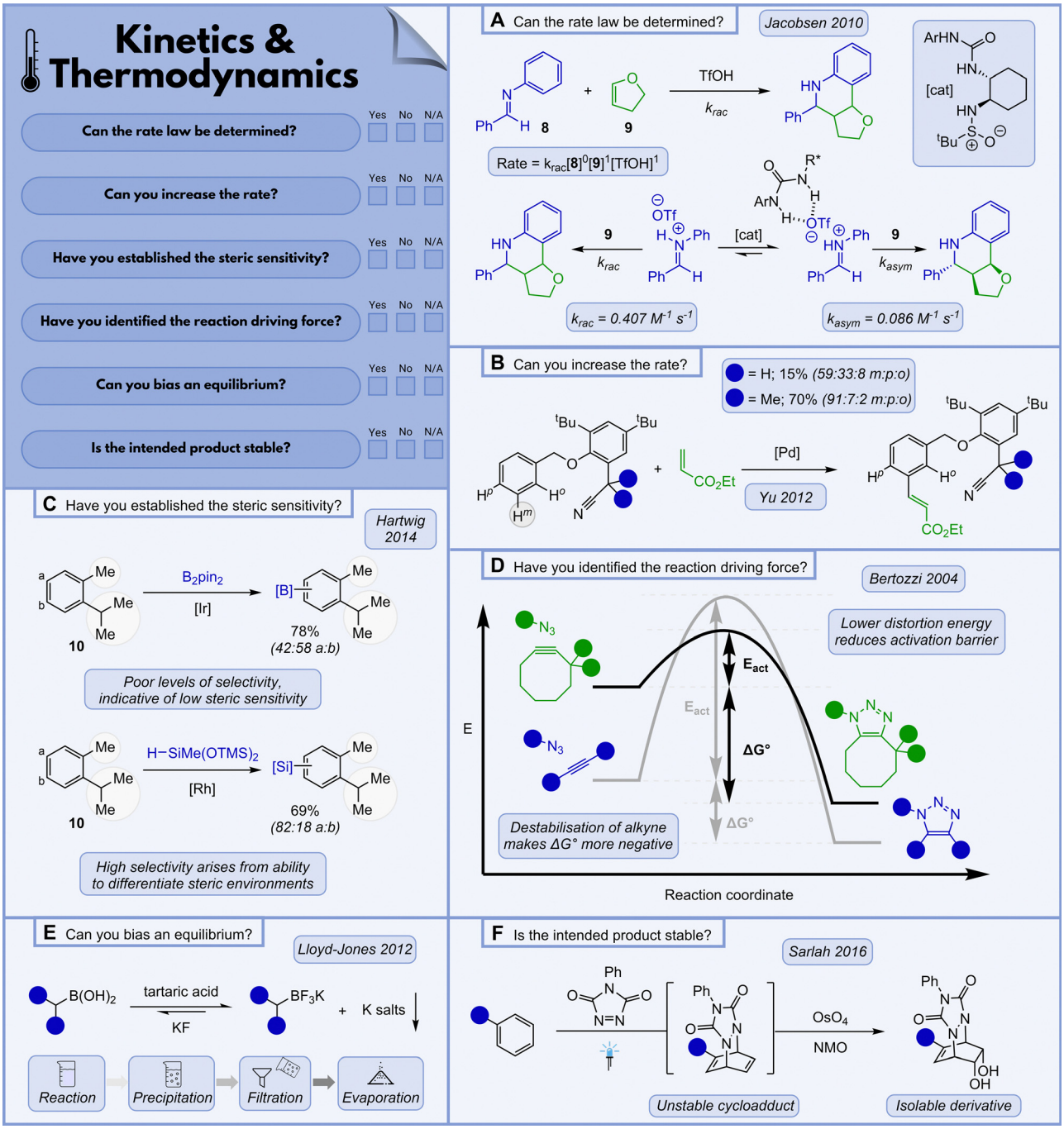


Fig. 3 Kinetics & thermodynamics checklist.

form; rate = k[reactant A]^x[reactant B]^y.[62] Deriving the rate law and determining the overall order of a reaction (the value of x and y) is an often-neglected part of reaction discovery.[63] However, obtaining the differential rate equation can provide substantial amounts of information that can help guide decisions on reaction stoichiometry and give insight into the mechanism in operation.[64–67]

During an investigation into the enantioselective Povarov reaction, Jacobsen and coworkers employed kinetic analysis to establish that in the absence of the chiral catalyst the reaction was first-order in both alkene (**9**) and acid but zeroth-order in imine (**8**), indicating that this species undergoes quantitative protonation to generate the resting state of the achiral catalyst (Fig. 3(A)).[68] Additionally, these rate laws were used to calculate the second-order rate constants for the symmetric and asymmetric reactions which showed that the chiral catalyst bound iminium species reacted considerably slower than when unbound, indicating that the high enantioselectivity observed arises due to the catalyst binding being particularly favourable.

Can you increase the rate?

Regardless of whether the rate of reaction has been determined, it may be worth considering how it could be maximised. Of course, an increase in temperature will raise the average kinetic energy of the reactants, providing a greater proportion of molecules with the minimum energy necessary to react, but this may be deleterious to the outcome of temperature sensitive transformations.

In addition to conditions, manipulation of the substrates themselves may provide the most effective way to directly influence reaction rate. The introduction of transient interactions such as hydrogen bonds, halogen bonds or Lewis acid–base coordination can bring together reacting molecules and orientate them as to improve the probability of a productive collision.[69–71] Such interactions can also be used to pre-activate reactants and lower the activation barrier.[72] For intramolecular reactions, the pre-organisation of substrates can impact the rate of a reaction drastically. One way to achieve this is to exploit the Thorpe-Ingold (or *gem*-dimethyl) effect.[73,74] Here, the introduction of geminal disubstitution in linear molecules restricts rotational freedom, with steric repulsion disfavouring the linear conformer and thereby increasing the population of the conformation pre-disposed towards cyclisation. This powerful effect has been exploited in many of the seminal systems for directed C–H activation developed by Yu and coworkers.[75,76] In one such example, remote *meta*-C–H bond functionalisation was targeted using a nitrile-containing directing group to facilitate the formation of the metallocycle, however, functionalisation was sluggish and non-selective using the initially designed template (Fig. 3(B)).[77] It was then observed that the efficiency of cyclisation could be greatly improved by installing two alkyl substituents adjacent to the nitrile group, harnessing the Thorpe-Ingold effect and facilitating Pd C–H insertion.

Have you established the steric sensitivity?

Determining the steric sensitivity of a reaction can help justify differing reaction rates between substrates, indicate the limit of reactivity and may even provide valuable mechanistic insights. Steric hinderance may also be critical to achieving the desired reactivity, impeding a deleterious side reaction or blocking alternative reactive sites that could lead to isomeric mixtures. Although computational calculations of steric parameters such as cone angle, buried volume, solid angle and *A*-value can supply valuable information as to whether a reaction is kinetically feasible,[78] competition experiments between substrates or measuring selectivity towards different reaction sites on the same molecule can be a quick and accessible alternative to establishing the role of steric hinderance in a transformation. A representative example of this was illustrated by Hartwig and coworkers in their report of a rhodium-catalysed C–H silylation of arenes, where the regioselectivity of this transformation was principally dictated by the steric control of the substituents *ortho* and *meta* to the reacting C–H bond (Fig. 3(C)).[79] To directly compare the level of steric sensitivity with an analogous C–H borylation reaction, electronically unbiased substrate **10** was subjected to both reaction conditions individually, with the silylation procedure showing a significantly increased selectivity for functionalisation occurring *para* to the largest group.

As well as providing a potential strategy for improving selectivity, increasing the steric bulk of substrates can also be advantageous with reactions that have early transition states, where attractive dispersive interactions are felt the strongest. This has been the case, for instance, in radical addition reactions to form highly congested, contiguous quaternary stereocenters.[80]

Have you identified the reaction driving force?

The thermodynamic favourability of a chemical process is given by the standard Gibbs free energy change, ΔG°. The value of ΔG° can be calculated directly from the relative changes of enthalpy (H) and entropy (S) in the system using the equation $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$, where a negative value of ΔG° indicates a thermodynamically favourable process. To maximise the enthalpic favourability (ΔH) of a reaction, one can consider the use of a reagent which can generate highly stable byproducts. For example, the formation of P═O double bonds ($\approx$575 kJ mol^{-1})[81] is a common strategy to exploit thermodynamic principles in organic chemistry and is one of the key driving forces behind transformations such as the Mitsunobu, Appel, Wittig, and Staudinger reactions.[82]

An alternative, but related, strategy for the maximisation of a systems overall change in enthalpy is to weaken the bonds of the starting materials. This is most often achieved through the confinement of the breaking bond in a strained ring, thus destabilising the reagent relative to the product and enhancing its reactivity (Fig. 3(D)).[83] Examples that have been highly successful in this regard include the strain-promoted azide–alkyne cycloaddition (SPAAC) reaction developed by the Bertozzi group[84] and the strain-release heteroatom functionalisation of bicyclo[1.1.0]butanes reported by Baran and coworkers.[85–91] However, it must also be noted that as well as providing a thermodynamic driving force, employing strained molecules can also lower a reactions activation barrier (a kinetic effect), due to the reduction in the distortion energy required to access the transition state.[92–94]

Temperature will have an effect on the reaction driving force *via* entropy since $\partial\Delta G/\partial T = -T\Delta S^\circ$. Hence, reactions that proceed with a decrease in entropy, such as cycloadditions, will be thermodynamically more favourable at lower temperatures, whereas fragmentations will be more favourable upon heating. Pressure will also affect the driving force *via* the reaction volume given that $\partial\Delta G/\partial P = \Delta V$. Consequently, reactions that occur with a decrease in volume, such as cycloadditions, will be more favourable at higher pressures.

Can you bias an equilibrium?

The identification of a reversible process in a novel reaction can provide several new avenues of investigation to improve the transformation. Le Chatelier's principle states that if a dynamic equilibrium is perturbed by changing reaction conditions, the position of the equilibrium will shift to counteract the change

 This journal is © The Royal Society of Chemistry 2025

and re-establish the equilibrium. In the case of ring-closing metathesis, the equilibrium is often driven toward product formation through the generation of ethylene, a volatile gas which can escape from the reaction mixture.[95] This creates a decrease in the concentration of the products that will ultimately shift the position of the equilibrium towards the ring-closed compound. The counterpart to this strategy is to precipitate the product or byproduct from the reaction solution. Such an approach was succinctly demonstrated by Lloyd-Jones and coworkers during their organotrifluoroborate synthesis from boronic acids and KF (Fig. 3(E)).[96] Here, addition of L-(+)-tartaric acid was used to precipitate potassium bitartrate from solution, driving the equilibrium toward trifluoroborate formation and facilitating byproduct removal *via* filtration.

Is the intended product stable?

The stability of a given reaction product is often characterised by its ability to resist decomposition in air, at room temperature, under ambient light, and is accordingly labelled as "bench-stable". This, however, does not take into account stability with respect to purification and the reaction conditions, both of which should be determined for novel structures. Product stability over the course of the reaction can be rapidly assessed by subjecting the requisite compound to the same conditions as the starting materials experience and monitor any depletion over time.

Decomposition during purification can often be hard to identity as the byproducts formed may elude identification. The simplest method to access this is to quantify product formation both before and after purification, using techniques such as quantitative liquid/gas chromatography and NMR analysis, with a discrepancy between the two values indicating mass loss. Once instability has been identified, one can attempt to alleviate degradation by using more mild reaction conditions or an alternative purification strategy such as avoiding nucleophilic or protic solvents, neutralising or using a different column chromatography stationary phase or avoiding chromatography entirely by precipitating the molecule from solution. If no method can be identified that effectively suppresses decomposition, then the derivatisation of the reaction products may be the most profitable method to access a stable adduct. This strategy was utilised by Sarlah and coworkers in 2016, where it was shown that the cycloadducts from a previously observed arene dearomatising [4+2] cycloaddition reaction were not amenable to purification (Fig. 3(F)).[97] To alleviate this issue, the resulting dienes were subjected to dihydroxylation conditions to access isolable diols and hence rendering the transformation synthetically useful.

3. Mechanism

Does the proposed mechanism require evidence?

For many new reaction protocols, it is highly likely that a combination of chemical intuition and literature precedent can provide an approximation of the mechanism in operation. In such cases, the mechanistic "evidence" can simply come from the appropriate citations of previous literature reports that have established similar pathways. However, if an entirely novel mechanism is postulated, or if the proposed mechanism does not fit all of the observed experimental results, it may be necessary to probe this proposal by finding evidence that supports your assertions and elucidates the finer details of the reaction coordinate. Most mechanistic experiments are only relevant to certain classes of reactions and provide a limited amount of information, therefore, it is of huge benefit to be aware of a broad variety of techniques. A selection of the more widely exploited types of experiments are detailed both below and in Fig. 4, along with key references that can deliver a more in-depth treatise.

One of the most widely used and sensitive tools for studying reaction mechanisms are kinetic isotope effect (KIE) experiments, in which the ratio of rate constants for the reactions involving substrates containing the light (k_L) and the heavy (k_H) isotope (typically hydrogen and deuterium) is calculated.[98] A system in which the breaking of a bond to an isotopically labelled atom occurs in the rate-determining step will show a discrepancy in relative rates due to the increased bond strength for heavier isotopes, resulting in a KIE > 1. Kinetic isotope effects can also be measured rapidly and without direct kinetic measurements by analysing the product distribution in a competition experiment between isotopically labelled substrates.[99] For example, Stahl and coworkers, upon subjecting both the deuterated and non-deuterated substrates to their reaction conditions, observed a significant primary KIE during an investigation into a newly developed copper-catalysed benzylic etherification reaction (Fig. 4(A)).[100] This result was fully consistent with the density functional theory calculations which suggested that the hydrogen-atom transfer step was associated with the greatest energetic barrier on the reaction pathway.

If a redox process is anticipated, then cyclic voltammetry studies can provide the ground state redox potentials of the starting materials and reagents, to indicate the likelihood of an oxidation or reduction process occurring.[101] Similarly, if energy transfer is postulated, then calculating the triplet energy (E_T) of the substrate can indicate whether this process is feasible, specifically that E_T (sensitiser) exceeds E_T (substrate).[102,103] Stern–Volmer fluorescence quenching studies, where the emission spectrum of an excited molecule is monitored upon subjection to a substrate, can determine which species in a given reaction can directly participate in energy/electron transfer with a sensitizer or photocatalyst.[104] Determination of the quantum yield of a photochemical reaction (Φ, the number of molecules participating in the reaction per absorbed photon), can also confirm the existence of a radical chain mechanism if Φ is calculated to be >1.[104–106]

As well as signifying whether a given substrate can effectively absorb light at the emitted wavelength, UV-visible spectroscopy can provide key information on the mechanism.[104] For example, Melchiorre and coworkers, through analysis of substrate absorption spectroscopy, demonstrated that enamines and benzyl bromides are capable of generating electron

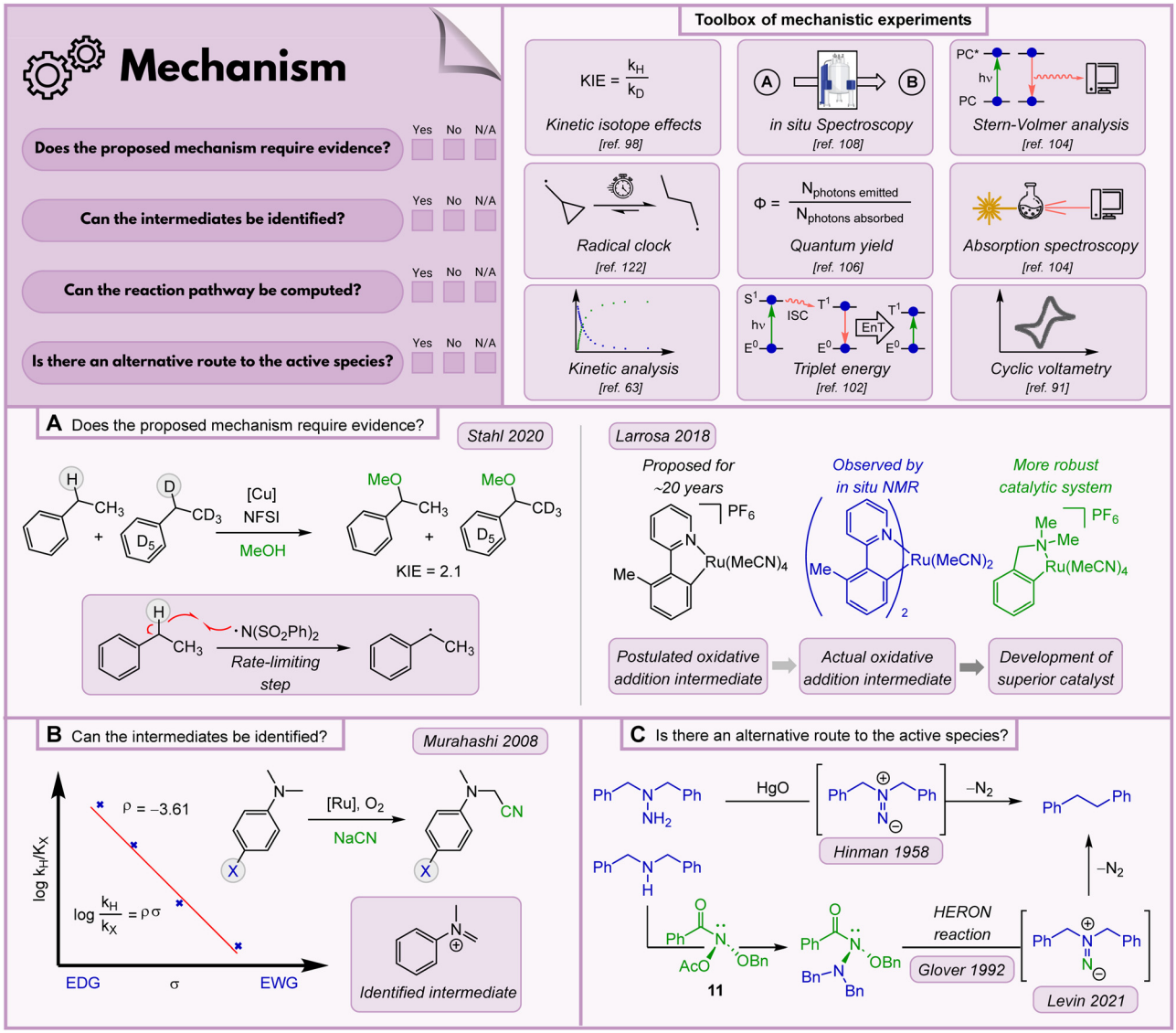


Fig. 4 Mechanism checklist.

donor–acceptor (EDA) complexes that undergo electron transfer upon excitation with visible light.[107] This observation was key to the elucidation of the unique mechanism displayed in their asymmetric aldehyde α-alkylation transformation. Perhaps the technique with the potential to provide the greatest quantity of data in a single experiment is the use of *in situ* monitoring techniques such as UV-visible, NMR, fluorescence and IR spectroscopy.[108–110] This often provides important kinetic data as well as the direct observation of the formation and interconversion of reaction intermediates. This was highlighted in a study from Larrosa and coworkers where *in situ* NMR spectroscopy was used to observe the formation of a previously unknown bis-cycloruthenated complex, later confirmed to be the key intermediate required for oxidative addition (Fig. 4(A)).[111] The importance of performing such mechanistic experiments was highlighted in this report as the information obtained, with regards to the mechanism, was directly used to design a novel pre-catalyst that could more easily access the active catalytic species.

Can the intermediates be identified?

If an intermediate on the proposed reaction pathway is a stable species, then its direct isolation can be targeted to prove its existence. Alternatively, one could attempt to synthesise the given intermediate and subject this compound to the reaction conditions. However, it is more likely that the key intermediates involved in the reaction are transient high energy species which severely complicates handling. Therefore, a variety of mechanistic experiments can be performed to gain further insight into the identity of the intermediates. For example, the extent and nature of charge build-up in the transition state of the rate-limiting step can be estimated through the Hammett equation.[112] By calculating the rate of reaction for a variety of substrates with varying electronic properties, and plotting the

 This journal is © The Royal Society of Chemistry 2025

data according to the substituents corresponding σ values (constants that describe electronic properties), the gradient of this graph provides a value for rho (ρ).[113–120] A positive ρ value means that electron density is increased in the transition state and *vice versa*, providing valuable information on the existence of charged species on the reaction pathway. This type of study has been performed countless times but a representative example from Murahashi and coworkers demonstrates this principle succinctly as the calculation of a large and negative ρ value confirmed the existence of a key iminium intermediate in the mechanism of a ruthenium-catalysed oxidative cyanation of amines (Fig. 4(B)).[121]

If a radical intermediate is proposed then a radical clock experiment can be performed in which a substrate containing an intramolecular pathway for the suspected radical intermediate is subjected to the reaction conditions.[122] Typically, cyclopropane ring-opening is exploited as this transformation can be more facile than the intramolecular or intermolecular steps that follow radical formation. Hence, the identification of cyclopropane ring-opening products provides strong evidence for radical formation. A distinct method to obtain similar information is to introduce trapping agents into the reaction and characterise the resulting adducts. The addition of external electrophiles and nucleophiles can be used to trap anions and cations respectively, whereas single electron species can be trapped with radical scavengers such as 2,2,6,6-tetramethylpiperidin-1-yloxyl (TEMPO) and 2,6-di-*tert*-butyl-4-methylphenol (BHT).[123,124] With the knowledge gained from these experiments on the nature of the reaction intermediates, a revision of the reaction conditions may be required to further promote the desired reactivity.

Can the reaction pathway be computed?

With a solid understanding of the reaction mechanism, now backed-up with evidence obtained from mechanistic investigation, it may be beneficial to compute the reaction pathway. This primarily consists of calculating the relative energies of the proposed intermediates and locating the transition states between them which can provide strong indications as to the feasibility of these species forming under the reaction conditions.[125]

As well as the viability of the major pathway, one can also compute any alternative pathways to disprove other potential competing mechanisms and rationalise the regio-, chemo- or stereoselectivity.[125] For example, in the 2013 report from Siegel and Houk on metal-free C(sp^2)–H oxidation, quantum mechanical calculations were used to discover that the transformation proceeds through a 'reverse oxygen-rebound' pathway, making it mechanistically distinct from previous oxygen-rebound mechanisms observed for metal–oxo oxidants.[126] Further calculations were also employed to rationalise the high level of aryl C(sp^2)–H *vs.* benzylic C(sp^3)–H selectivity.

Is there an alternative route to the active species?

Finally, one should critically evaluate the proposed mechanism and ask: can the key reactive intermediate be accessed by an alternative route? This could lead to a simplification of the reaction protocol and the transformation becoming more broadly applicable. If an already established protocol using the same intermediate has been reported in the literature, then a justification must be made as to what an alternative method can add to the field. It may be that the conditions have been significantly improved in terms of yield, selectivity, toxicity of reagents or functional group tolerance. For instance, the extrusion of N_2 from an isodiazene intermediate to deliver the nitrogen deleted product has been reported several times dating back to the 1950s, but the reaction was never deemed of particular synthetic use due to the harsh and hazardous conditions required (Fig. 4(C)).[127,128] However, Levin and coworkers successfully employed anomeric amide **11** that had been previously demonstrated by Glover and coworkers to participate in S_N2 reactions with amines, before undergoing a HERON (Heteroatom Rearrangement On Nitrogen) reaction, to deliver the desired isodiazene intermediate.[129] Employing this strategy, pioneered by Glover,[130–132] provided a much milder protocol to access the active species which could subsequently extrude N_2 gas and thus allowed access to a viable tool for nitrogen atom deletion.[133]

Another example are vinylidene carbenes, which can either stem from a 1,1-dibromo alkene and an organolithium compound (Corey–Fuchs reaction) or from a carbonyl compound and a diazophosphonate (Ohira–Bestmann reaction), and can rearrange to the same alkyne.[134]

4. Optimisation

Has the yield been fully maximised?

Optimising a reaction for yield is a process that, albeit important in the context of reaction development, can often be laborious and time-consuming. It is therefore vital to maximise efficiency by rationalising the choice of specific conditions based on previous observations and a mechanistic understanding of the transformation. As well as the more obvious reaction parameters (time, temperature, solvent, stoichiometry and concentration), aspects such as mixing rates, solution pH, the introduction of additives and the addition rate/order of reagents should be considered (Fig. 5). However, there are many parameters that are only applicable to certain classes of reactions. For example, in light-induced reactions, the specific choice of wavelength and light intensity can be varied,[135] for electrochemical reactions, current density/voltage, electrode material and inter-electrode distance must be considered[136–138] and in flow systems, residence time and flow rates should be optimised.[139]

It must also be noted that the techniques described above are typically used for the yield optimisation of a single ''model'' substrate. Although this is often due to a single molecule being targeted, especially in an industrial setting, the optimal reaction conditions that have been located may lack generality for other substrates or classes of compound.[140,141] In order to address this, Doyle and coworkers recently reported the design

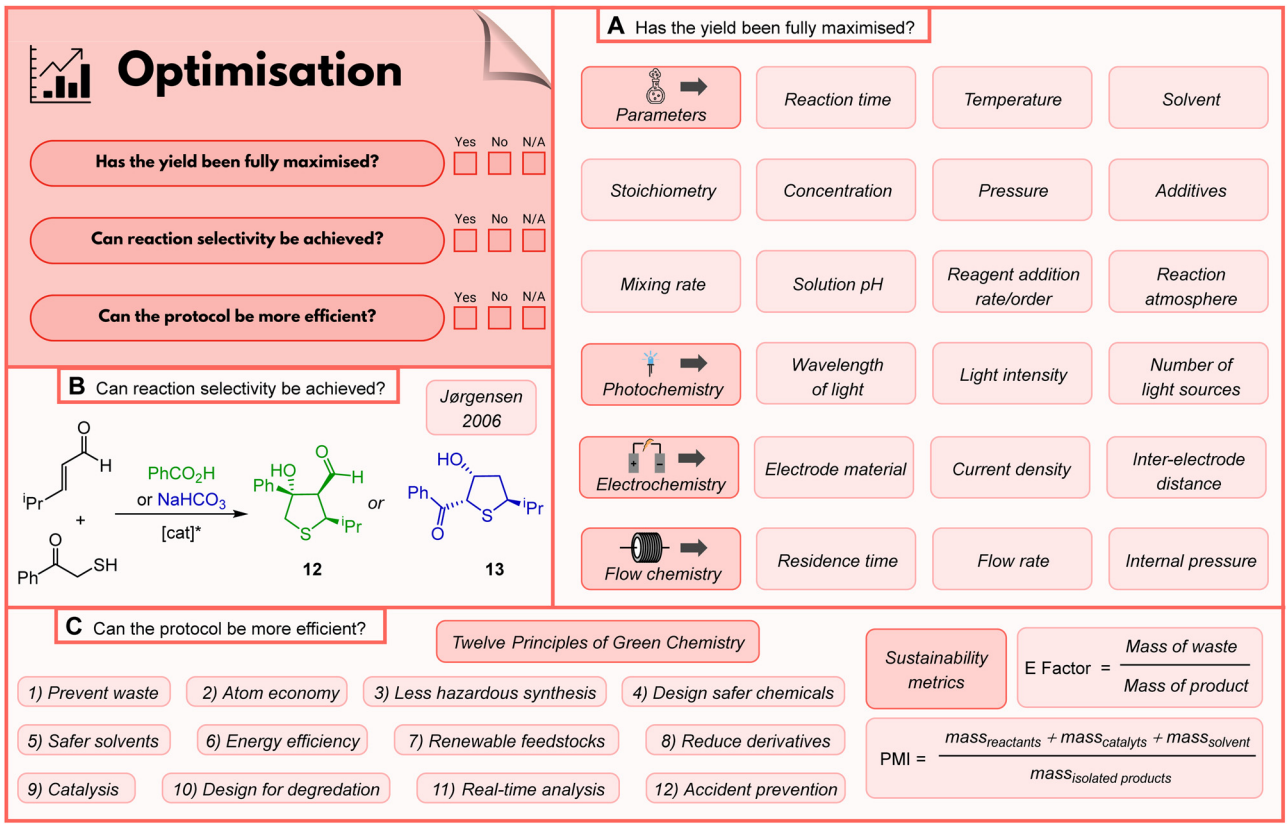


Fig. 5 Optimisation checklist.

and application of reinforcement learning optimisation models that attempt to identify the most generally applicable conditions for a given substrate scope without prior knowledge of the reaction system.[142] Not only does the implementation of a statistical model reduce the total number of experiments being performed, but the identification of a general reaction protocol may avoid undertakings such as substrate re-optimisation or the use of multiple conditions for a single transformation.

Can reaction selectivity be achieved?

When performing an optimisation for yield, it is highly likely that this involves improving the selectivity of the desired reaction over the formation of any byproducts. This is an aspect of chemoselectivity, which can also be examined through the tolerance of the reaction to a variety of functional groups. In cases in which byproducts are observed, one must consider whether these species are of synthetic interest themselves, as obtaining conditions that allow selective access to multiple distinct compounds under similar conditions can add significant value to your protocol.[143,144] This principle of divergent reactivity has been reported countless times but one such example from Jørgensen and coworkers demonstrated that the selectivity of their organocatalytic cascade synthesis of tetrahydrothiophenes could be switched from aldehyde **12** to ketone **13**, initially observed as a side-product during the optimisation study, by employing a basic additive (Fig. 5(B)).[145]

As well as chemoselectivity, regio- and stereoselectivity should be considered in order to render the reaction synthetically useful. Improving regio- and diastereoselectivity is often achieved by introducing selective steric interactions, for instance, through tuning bond lengths. This has been famously achieved in aldol additions where a change from lithium to boron enolates generally leads to higher diastereoselectivities.[146]

When maximising enantioselectivity, the non-selective background reaction must be disfavoured relative to the reaction that occurs within the chiral environment. The most profitable strategy in this regard is the use of chiral catalysts in systems where the desired reactivity does not occur to a measurable extent in its absence (see Section 5).[70,147–150] However, even when the catalyst and substrate interact, mixtures of enantiomers can be generated when the energy difference between the two corresponding transition states is not sufficiently high. This being said, a $\Delta\Delta G^{\ddagger}$ of $\sim$2.7 kcal mol^{-1} is all that is required to achieve 99% enantiomeric excess at room temperature.[151,152]

Can the protocol be more efficient?

The most likely bottlenecks when undertaking a synthetic methodology project are the setup (and subsequent running) of reactions and product purification. It is therefore essential to scrutinise the efficiency of both these aspects to maximise productivity and avoid needless time loss. These considerations are closely related to reaction sustainability, a factor whose

 This journal is © The Royal Society of Chemistry 2025

importance cannot be overstated, as the justification of chemical transformations in the context of environmental impact will be perhaps the most important aspect of future reaction development.[153,154] The specifications for sustainable or "green" chemistry have been succinctly outlined in the "12 Principles of Green Chemistry" (Fig. 5(C)), developed by Anastas and Warner in 1998.[155] Despite being established more than 25 years ago, these principles are still considered as the guiding framework for the design of new reaction protocols due to their holistic approach, considering all aspects of any given transformation, from the method of obtaining the raw materials, through to the toxicity and biodegradability of the waste byproducts.[156]

Practical steps that can be taken towards meeting these criteria include performing syntheses in benign solvents such as water, or under solvent free conditions, to reduce a large quantity of reaction waste.[157,158] Additionally, running a reaction in air, using solar light, performing the synthesis at room temperature or employing catalysts (see Section 5), can drastically improve sustainability.[159] With respect to product purification, methods which maximise efficiency are also likely to reduce environmental impact. If feasible, non-chromatographic purification methods such as recrystallisation, liquid–liquid extraction, distillation, sublimation, and trituration can be employed as more expedient and less wasteful approaches to compound isolation. Finally, it is imperative to note that before claiming that the newly developed transformation can be labelled as green or sustainable, all aspects of environmental impact must be fully assessed. This can be achieved by determining quantities such as overall atom economy,[160] reaction mass efficiency,[161] process mass intensity (PMI)[162–165] and complete environmental factor (cEF)[166,167] and comparing theses values to analogous syntheses (Fig. 5(C)).[168,169]

5. Catalysis

Can the reaction be catalysed?

The design and implementation of catalysts lies at the heart of countless industrial chemical processes, and innovation in catalysis is central to solving many of society's most important issues.[170] Although it may not seem practical in many cases, it is always worth considering whether a newly discovered transformation can be rendered catalytic. Indeed, many stoichiometric reactions known to chemists for decades have been rendered catalytic through pioneering work to redesign the reaction system. For example, the Mitsunobu reaction, first reported in 1967,[171,172] was not demonstrated in a catalytic system until 2006 when Toy and coworkers succinctly showed that stoichiometric iodobenzene diacetate (PIDA) could be employed as an external oxidant to recycle the azo reagent (diethyl azodicarboxylate).[173] Since this seminal report, several research groups have made significant improvements to the efficiency of the recycling protocol, reaching a pinnacle in 2019 with Denton and coworkers development of a novel phosphorus-based organocatalyst that can facilitate completely redox-free catalytic Mitsunobu reactions (Fig. 6(A)).[174]

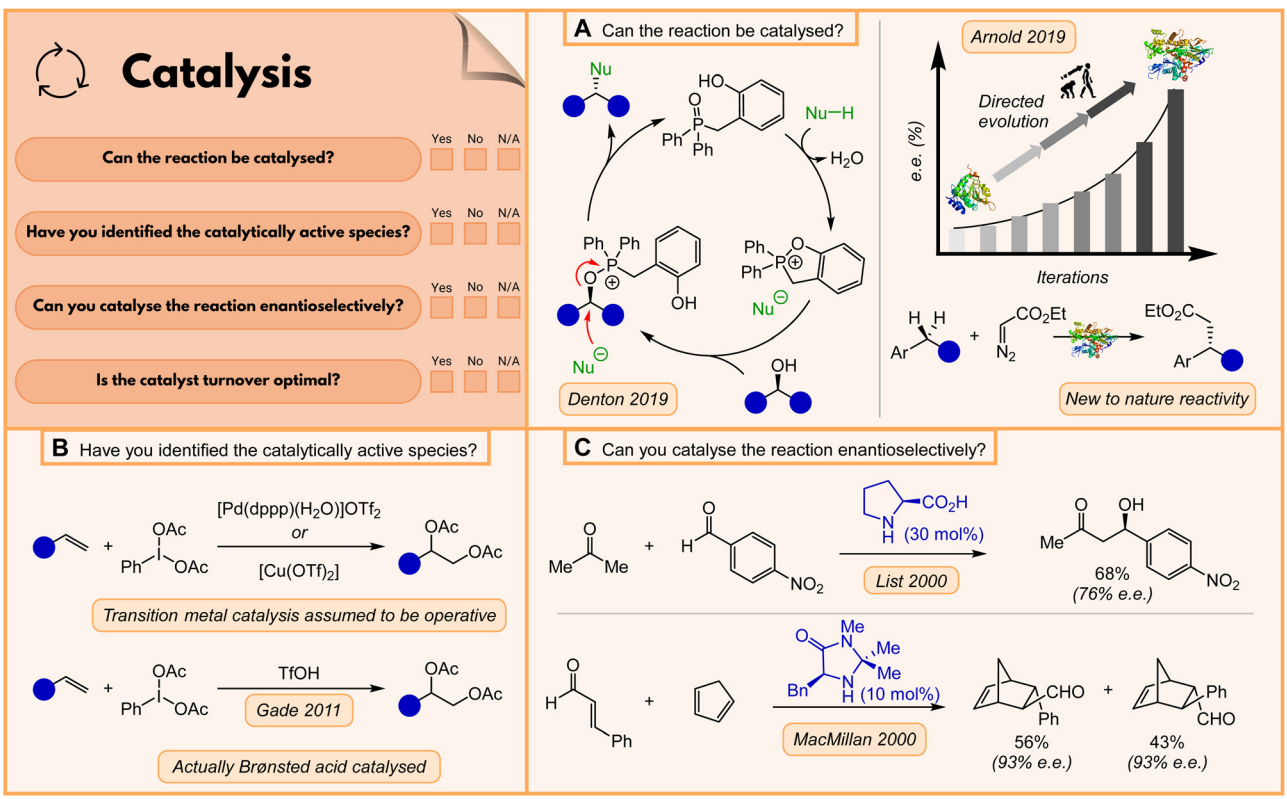


Fig. 6 Catalysis checklist.

The enormous potential of catalysis is best evidenced by natures use of enzymes to accelerate reactions crucial to an organism's function.[150,175] However, it is highly unlikely that a newly developed transformation can be directly catalysed by a native enzyme. This being said, researchers, most notably Arnold and coworkers, have established that enzymes can be fine-tuned to target the catalysis of a reaction entirely different from an enzymes original function.[176] This is achieved through a technique known as ''directed evolution'', where the genes encoding an enzyme are subjected to iterative rounds of mutagenesis and only mutated enzymes that show improved reaction efficiency are taken forward to the next round (Fig. 6(A)). For instance, Arnold succinctly demonstrated that a cytochrome P450 monooxygenase could be modified *via* directed evolution to promote C–H alkylation reactions.[177] Here, turnover and enantioselectivity were targeted in parallel to access a highly efficient catalytic system, capable of operating under exceptionally mild conditions. This overarching strategy demonstrates the breadth of transformations that could potentially be applicable to enzyme catalysis.[150,175]

Have you identified the catalytically active species?

Although often overlooked, in many catalytic systems, the reagent added to a reaction mixture acts only as a precatalyst and is activated *in situ* to unveil the active species. However, determining the true nature of this species can directly lead to a better understanding of the reaction manifold which, in turn, can aid the development of more potent and efficient catalytic systems.[41,178] A case study that succinctly demonstrates this concept was reported by Gade and coworkers during their mechanistic investigation of previously reported Pd-catalysed and Cu-catalysed diacetoxylation reactions of alkenes (Fig. 6(B)).[179] Through carefully designed experiments, the screening of various metal complexes, salts and Lewis acids, as well as employing kinetic and *in situ* spectroscopic studies, it was discovered that Brønsted acids were solely responsible for the catalytic activity displayed by the system. In fact, they determined that the only role of the metal complexes, that were previously invoked as the catalytically active species,[180,181] was to simply generate TfOH in solution. Using this information, the reaction conditions were re-examined and a protocol was developed that was equally as efficient as the previously reported transformation, whilst completely eliminating the use of precious metals.

Can you catalyse the reaction enantioselectively?

Regardless of the nature of the transformation, one should always consider the possibility of employing a chiral catalyst, considering that even the simplest systems, for example those catalysed by acid, can be rendered enantioselective by employing chiral sources of H^+ such as chiral phosphoric acids (CPAs).[182] In fact, every conceivable reaction can be rendered asymmetric through the desymmetrisation of a C_s or C_i symmetric substrate with enantiotopic functional groups.

Classically, enantioselectively catalysed reactions have been most effectively achieved with homogeneous metal catalysts, however, in the last two decades, organocatalysis has rapidly become a vital part of the synthetic chemist's toolbox due to key advantages such as catalyst stability, low toxicity and reductions in associated reaction cost.[69,183–185] This field has been spearheaded by the work of List[186] and MacMillan[187] who, in 2000, independently reported different enantioselective protocols, catalysed by organic compounds derived from natural sources and was subsequently followed by an upsurge of reports of novel and innovative systems in the years that followed.

Is the catalyst turnover optimal?

Turnover number (TON) refers to the number of moles of substrate that a mole of catalyst can convert to the desired product before becoming inactive. Although much debate has arisen from the apparent fundamental flaw of comparing a condition specific value across systems that employ differing conditions, when analysing a single protocol, TON can still provide valuable information on reaction efficiency.[188] This principle was neatly demonstrated by Stahl and coworkers, where a high TON was targeted in the optimisation of a quinone-based ligand for C–H arylation, leading to an efficient sp^2–sp^2 coupling protocol that could be performed using 0.15 mol% of Pd catalyst.[189] The ability to perform reactions at lower catalyst loadings is not only beneficial due to the reduction in cost associated with employing a precious metal with a finite natural supply but also from a sustainability viewpoint, can decrease the environmental impact associated with both obtaining toxic metals and removing them from fine chemicals.

6. Scope

Is the substrate scope diverse?

Generally speaking, a substrate scope is the term given to the extent of the area of chemical space in which a transformation is relevant, but often specifically refers to the collection of discrete substrates that have been tested under the newly developed conditions. Current trends in reaction development clearly show that the number of scope entries in synthetic methodology publications are increasing drastically over time.[190] However, it is often the case that many of the compounds included in scope studies add very little value to the publication as they are closely related to other entries in terms of functionality and steric environment due to selection bias, leading to a large but non-diverse scope.[191] Furthermore, the strong bias towards ''positive'', high-yielding data points, whilst failing to include ''negative'' results from failed experiments, introduces considerable reporting bias.[192,193]

In order to diverge from this current trend, new strategies, and a community-wide effort, are needed to develop methods for substrate scope standardisation which can alleviate both selection and reporting biases.[194] One such approach that addresses this issue, pioneered by Doyle and coworkers, involves the featurisation of all potential substrates using DFT calculated parameters such as electronegativity and buried volume, followed by dimension reduction and hierarchical clustering, to generate a 2D map of chemical space (Fig. 7(A)).[195]

 This journal is © The Royal Society of Chemistry 2025

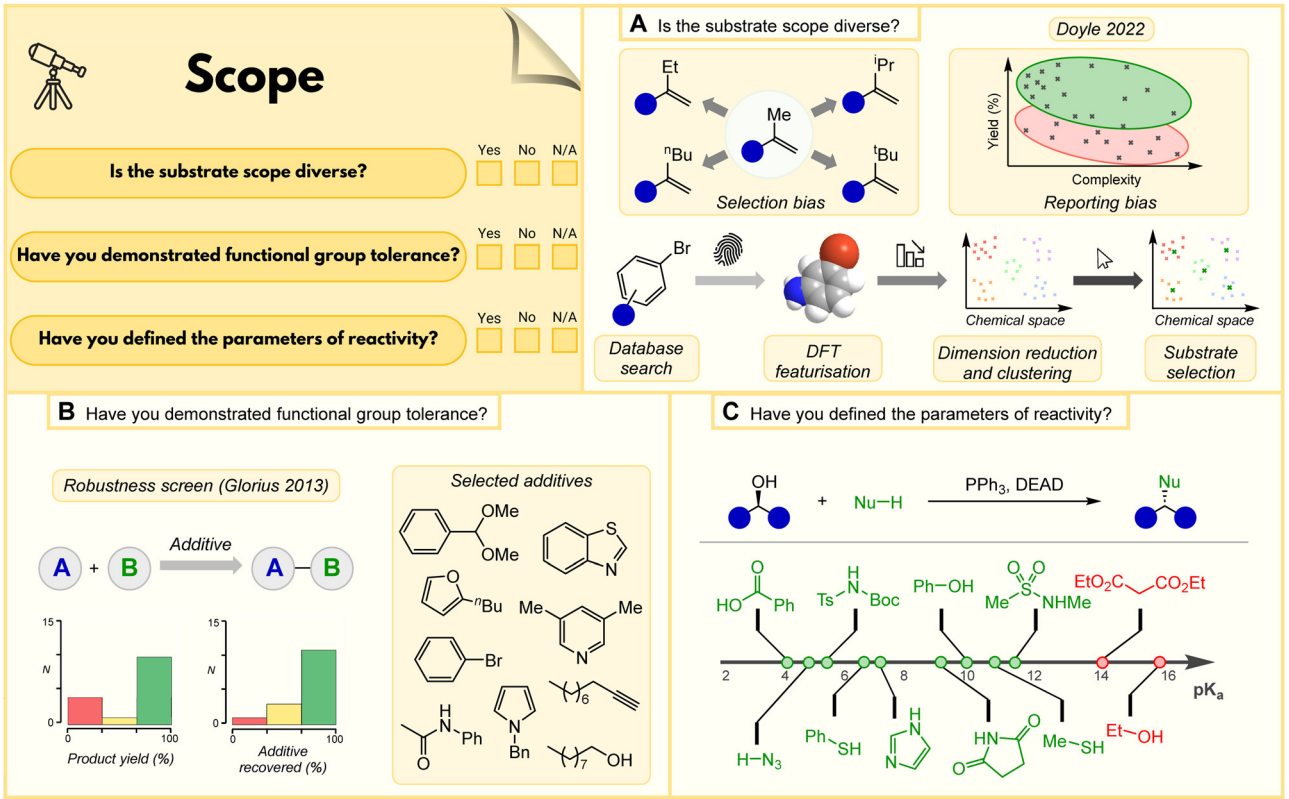


Fig. 7 Scope checklist.

From here, substrates that are representative of each cluster in unique areas of chemical space can be selected, providing a broad coverage from the fewest number of experiments.

Have you demonstrated functional group tolerance?

Within a substrate scope it is necessary to determine the limits of reactivity, especially with respect to the classes of functional groups that are compatible and incompatible with the transformation. As discussed above, the drive to improve the appearance of publications has led to considerable levels of both selection and reporting bias, therefore providing an incomplete picture of the overall reaction.[190–193] One strategy to alleviate these issues, developed by Glorius and coworkers, is to apply a "robustness screen" in which a selection of standardised additives that bear a broad range of functional groups are subjected to the conditions of the transformation (Fig. 7(B)).[196] By analysing both reaction inhibition and additive decomposition, an expedient evaluation can be made on the tolerance of the conditions to each functionality. However, in its current form, the "robustness screen" does not account for the location of functional groups relative to the reactive sight or the conformation of molecules, factors which can have a profound influence on the tolerance of certain functionalities. This can be illustrated by the observations made initially by Schrock[197] and Grubbs[198] and later studied by Fürstner,[199] of alkene metathesis inhibition *via* stabilising chelate interactions between carbonyl lone pairs and the metal atom of the carbene complex. This inhibition effect was demonstrated to only occur when the carbonyl is positioned relative to the alkene that allows the formation of a 5- or 6-membered metallacycle.

Have you defined the parameters of reactivity?

Despite the amount of data generated during an investigation of substrate scope, a minority of synthetic publications discuss the results thoroughly in terms of improving the understanding of the reaction manifold. Perhaps this is not surprising given that publication pressure has brought about both selection and reporting bias and so researchers are hesitant to state the true limits of reactivity lest it reflect poorly on the newly developed protocol. However, the information that can be obtained should not be overlooked. For example, it may be possible to define a pK_a range for which a reactant will facilitate reactivity, as has been demonstrated for the Mitsunobu reaction (Fig. 7(C)).[200–203] Similarly, the limits of substrate electro- or nucleophilicity can be estimated by measuring the outcome of reactants that represent different positions on the Mayr scale.[204] Finally, and most importantly, attempting to apply an entirely new substrate class to the established reaction conditions may not only help improve the generality and broad applicability of the transformation but, if substantially interesting, may constitute an entirely new project.

7. Applications

Can it be used to access target compounds?

Now that the reaction is fully understood, it is important to assess the value of the discovery by demonstrating how this

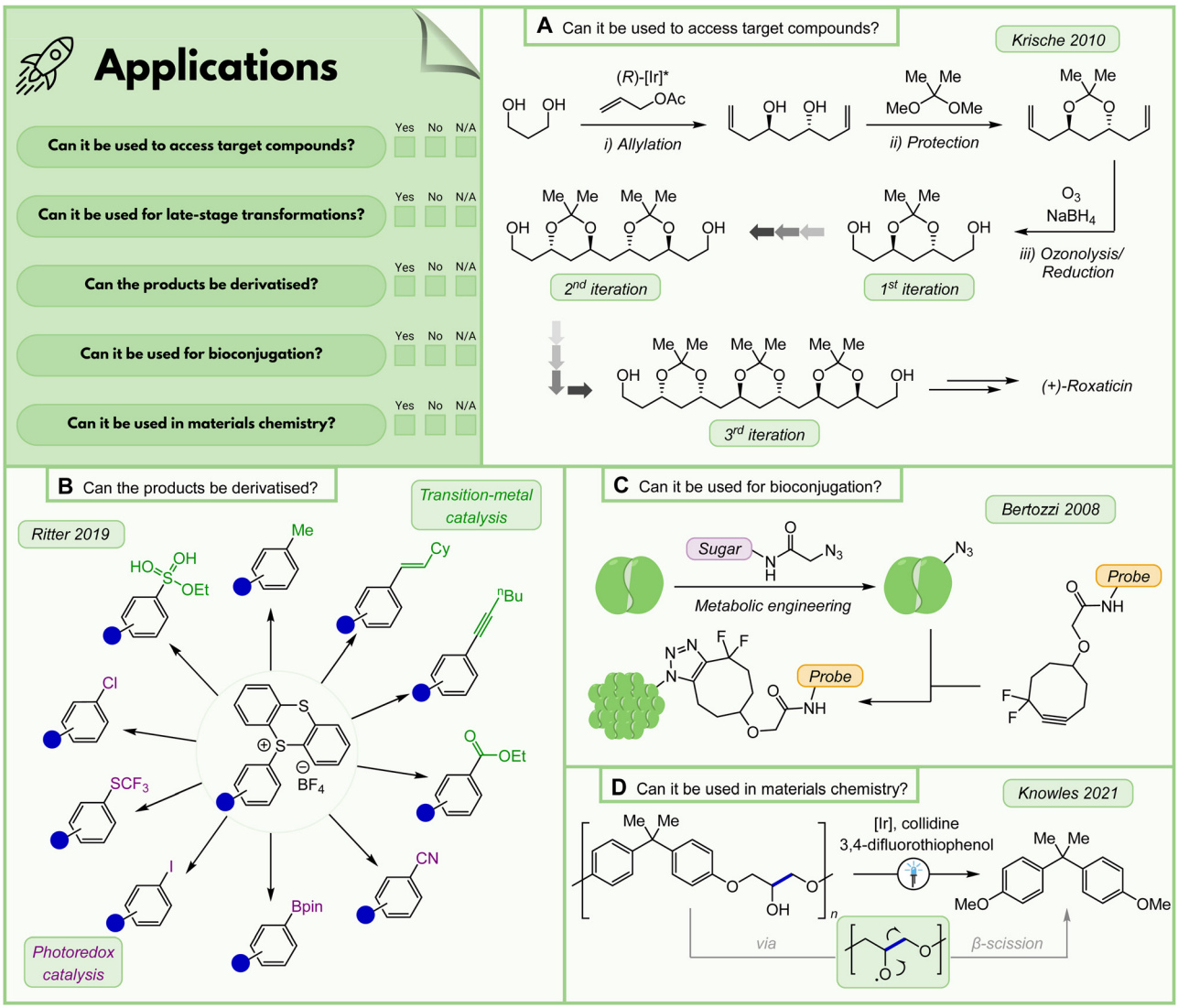


Fig. 8 Applications checklist.

new transformation can be employed to reach specific target molecules. If the reaction in question can be used to generate a frequently observed structural motif within biologically active molecules, it may be applicable to the synthesis of a wide variety of natural products or drugs. For example, in 2008, Krische and coworkers reported an enantioselective allylation of primary alcohols *via* iridium-catalysed transfer hydrogenation (Fig. 8(A)).[205,206] It was later demonstrated that this reaction could be performed iteratively by regeneration of the primary alcohol *via* alkene ozonolysis and subsequent reduction.[207] This approach, coupled with the ability to assess either enantiomer of the secondary alcohol by employing the *R* or *S* stereoisomer of the chiral catalyst,[208] allows the synthesis of stereodefined 1,3-polyols, fragments that are regularly encountered in polyketides.[209]

Can it be used for late-stage transformations?

Late-stage functionalisation can add additional substituents to existing molecules through highly chemoselective transformations, affording the potential for rapid diversification.[210,211] Similarly, reactions that facilitate a change in the topology of ring systems, referred to as skeletal editing, have recently received much attention.[212,213] They promise to have a large impact on structure–activity relationship studies as they facilitate the direct interconversion between structural motifs, such as (hetero)aromatic moieties, that would traditionally necessitate a complete *de novo* synthesis.[133,214–216]

Can the products be derivatised?

If a reaction installs a linchpin functional group, then it may be most profitable to display the utility of the transformation by derivatising a chosen substrate and accessing a broad range of new functionalities. Being able to generate a wide variety of compounds from common starting materials represents one of the most efficient methods to explore new areas of chemical space and is commonly exploited in structure–activity relationship studies and diversity-oriented syntheses for library generation.

 This journal is © The Royal Society of Chemistry 2025

This strategy has been utilised countless times but an example worth highlighting was published by Ritter and coworkers in 2019.[217] In this report, an aromatic C–H thianthrenation reaction was developed and shown to be highly site-selective and functional group tolerant (Fig. 8(B)). In order to emphasise the versatility of the resulting thianthrenium salts, a derivatisation study was performed on the functionalised analogue of the insecticide pyriproxyfen which was successfully transformed into an array of functionalised molecules *via* either transition-metal or photoredox catalysis to access both the corresponding aryl metal and aryl radical intermediates.

Can it be used for bioconjugation?

If the conditions that have been developed for the reaction in question are appropriately mild, then it may be possible to apply the chemistry to the field of bioconjugation. Specifically, this involves the tethering of a biomolecule, often a protein, with a small molecule that can act as a therapeutic agent, warhead or label, providing a highly efficient method to accumulate the desired compound in a specific area in a biological system.[218] Perhaps the most common method used in this field is to append the requisite small molecule with a functional group capable of undergoing a click-type reaction with an amino acid side-chain.[219] However, reactions that involve functionalities not observed in protein structures can still be a valuable asset to the field of bioconjugation. It is now common practice to append a linker molecule onto a biomolecule (*in vitro*) that contains a suitably reactive functional handle from which the desired compound can be attached, an approach regularly exploited in the synthesis of antibody–drug conjugates.[218,219] Additionally, the ability to selectively incorporate artificial amino acids into proteins has allowed the introduction of non-native functional groups directly into the biomolecule at a single position.[220] These techniques have allowed 1,3-dipolar cycloadditions between azides and alkynes to become one of the premier methods for bioconjugation, despite neither of these functional groups being present in native amino acids or other biomolecules.[221] For example, in a 2008 study, Bertozzi and coworkers deployed their strain-promoted [3+2] azide–alkyne cycloaddition reaction to non-invasively label zebrafish embryos *in vivo* (Fig. 8(C)).[222] The developing cells were treated with an unnatural sugar to metabolically label their cell-surface with azides, which acts as a linchpin for the coupling with a fluorescent probe compatible with imaging techniques. This, along with countless other related studies, demonstrate the wide variety of synthetic transformations that can find applications within this fertile and highly valuable area of chemistry.

Can it be used in materials chemistry?

Finally, one should consider whether the transformation could be utilised to assemble molecules that have applications in materials chemistry. In most instances, this will not be applicable as organic materials used for applications such as biomaterials[223] are typically oligomers and polymers and, in fields such as organic photovoltaics,[224] high molecular weight conjugated chromophores are required. However, the doping of semiconductors,[225] the functionalisation of metal surfaces[226,227] and the assembly of hydrogen-bonded organic frameworks (HOFs)[228] represent examples of fields where novel low molecular weight discrete organic molecules can be applied. It is therefore highly valuable to be aware of the design criteria of these systems in order to access the potential applicability of any newly synthesised molecules to these areas of materials chemistry.

Alternatively, it may be profitable to apply a new transformation to the degradation of a material.[229] Synthetic polymeric materials, such as plastics, have been accumulating in the environment for decades as a result of global consumption and the limited ability to efficiently recycle such compounds.[230–233] As a result, developing new methods for plastic depolymerisation, where polymers are selectively cleaved back to their constituent monomer units, is rapidly becoming one of the most important areas in sustainable chemistry research.[234,235] For example, Knowles and coworkers have recently developed a strategy for the chemical recycling of high molecular weight hydroxylated polymers.[236–238] This depolymerisation protocol occurs through the photocatalytic formation of alkoxy radicals, *via* proton-coupled electron transfer (PCET), which subsequently undergo β-scission of vicinal C–C bonds, generating well-defined and isolable monomer derivatives. Such strategies highlight the value of methodologies that facilitate the selective cleavage of fragments commonly encountered in polymers, such as amide, ester, carbamate, alkene and ether bonds, under mild conditions and, consequently, demonstrating the ability of a new transformation to help achieve this goal can greatly improve the potential impact of the studies undertaken.

Conclusions

When the synthetic results have been collected and a comprehensive understanding of the system is obtained, it is now time to share the results with the chemistry community. This will most likely be in the form of a scientific journal article or a patent which will have considerably different content criteria, depending on the destination of the submission. However, common to all publications should be a concerted effort to communicate concisely and with as little bias as possible. As repeatedly stated in this tutorial, reporting negative results and standardising data will be central to the digitisation of organic synthesis which, by definition, will expedite greater publication transparency.[193,239,240] Given that the future of synthetic chemistry is likely to be data driven,[43–45,241,242] the reporting of results can greatly impact the training data-sets that new machine learning models will be reliant upon, and with the slow uptake of this ideal in the field of synthetic method development, individual efforts are required to normalise non-bias reporting. All this being said, the ability to sell one's own chemistry still plays a vital part in publication writing and has impacts on both the size of the readership and the likelihood of enticing other researchers to take up related areas of investigation.

In addition to this tutorial review, we provide a handy reference that summarises many of the queries that one could ask themselves in the course of methodology development, as well as the checklists shown above, and can be found in the ESI.†

Data availability

No primary research results, software or code have been included and no new data were generated or analysed as part of this review.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Generous financial support by the Alexander von Humboldt Foundation (J. L. T.), is gratefully acknowledged. The authors thank Dr Alessia Petti, Felix Katzenburg, Florian Boser, Debanjan Rana for helpful discussions and Molly Kirkham for graphical assistance.

References

1 J. Woo, A. H. Christian, S. A. Burgess, Y. Jiang, U. F. Mansoor and M. D. Levin, *Science*, 2022, **376**, 527–532.
2 S. C. Patel and N. Z. Burns, *J. Am. Chem. Soc.*, 2022, **144**, 17797–17802.
3 R. J. Sundberg, S. R. Suter and M. Brenner, *J. Am. Chem. Soc.*, 1972, **94**, 513–520.
4 R. M. Roberts, *Serendipity, Accidental discoveries in science*, Wiley, New York, 1989.
5 H. C. Brown, *Science*, 1980, **210**, 485–492.
6 L. E. Overman, *Tetrahedron*, 2009, **65**, 6432–6446.
7 C. Friedel and J.-M. Crafts, *Ann. Chim. Phys.*, 1884, **1**, 449–532.
8 C. Friedel and J.-M. Crafts, *Compt. Rend.*, 1877, **84**, 1450–1454.
9 C. Friedel and J.-M. Crafts, *Compt. Rend.*, 1877, **84**, 1392–1395.
10 A. A. Ashdown, *Ind. Eng. Chem.*, 1927, **19**, 1063–1065.
11 G. A. Olah, *Friedel-Crafts and Related Reactions*, Interscience Publishers, 1963.
12 A. Miller and D. Trauner, *Synlett*, 2006, 2295–2316.
13 A. K. Miller and D. Trauner, *Angew. Chem., Int. Ed.*, 2003, **42**, 549–552.
14 K. D. Collins, T. Gensch and F. Glorius, *Nat. Chem.*, 2014, **6**, 859–871.
15 A. McNally, C. K. Prier and D. W. C. MacMillan, *Science*, 2011, **334**, 1114–1117.
16 D. W. Robbins and J. F. Hartwig, *Science*, 2011, **333**, 1423–1427.
17 E. S. Isbrandt, R. J. Sullivan and S. G. Newman, *Angew. Chem., Int. Ed.*, 2019, **58**, 7180–7191.
18 F. Strieth-Kalthoff, C. Henkel, M. Teders, A. Kahnt, W. Knolle, A. Gómez-Suárez, K. Dirian, W. Alex, K. Bergander, C. G. Daniliuc, B. Abel, D. M. Guldi and F. Glorius, *Chem*, 2019, **5**, 2183–2194.
19 M. N. Hopkinson, A. Gómez-Suárez, M. Teders, B. Sahoo and F. Glorius, *Angew. Chem., Int. Ed.*, 2016, **55**, 4361–4366.
20 L. Claisen, *Ber. Dtsch. Chem. Ges.*, 1912, **45**, 3157–3166.
21 A. M. Martín Castro, *Chem. Rev.*, 2004, **104**, 2939–3002.
22 N. Graulich, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2011, **1**, 172–190.
23 M. Breugst and H.-U. Reissig, *Angew. Chem., Int. Ed.*, 2020, **59**, 12293–12307.
24 S. Danishefsky, P. Schuda and K. Kato, *J. Org. Chem.*, 1976, **41**, 1081–1088.
25 S. Danishefsky, P. F. Schuda, T. Kitahara and S. J. Etheredge, *J. Am. Chem. Soc.*, 1977, **99**, 6066–6075.
26 S. Danishefsky, T. Kitahara, P. F. Schuda and S. J. Etheredge, *J. Am. Chem. Soc.*, 1976, **98**, 3028–3030.
27 S. Danishefsky and T. Kitahara, *J. Am. Chem. Soc.*, 1974, **96**, 7807–7808.
28 S. Danishefsky, *Acc. Chem. Res.*, 1981, **14**, 400–406.
29 S. Danishefsky, M. Hirama, K. Gombatz, T. Harayama, E. Berman and P. Schuda, *J. Am. Chem. Soc.*, 1978, **100**, 6536–6538.
30 D. B. Konrad, K.-P. Rühmann, H. Ando, B. E. Hetzler, N. Strassner, K. N. Houk, B. S. Matsuura and D. Trauner, *Science*, 2022, **377**, 411–415.
31 P. S. Gromski, A. B. Henson, J. M. Granda and L. Cronin, *Nat. Rev. Chem.*, 2019, **3**, 119–128.
32 L. Bustillo, T. Laino and T. Rodrigues, *Chem. Sci.*, 2023, **14**, 10378–10384.
33 C. W. Coley, N. S. Eyke and K. F. Jensen, *Angew. Chem., Int. Ed.*, 2020, **59**, 22858–22893.
34 H. J. Kulik and M. S. Sigman, *Acc. Chem. Res.*, 2021, **54**, 2335–2336.
35 M. Fitzner, G. Wuitschik, R. Koller, J.-M. Adam and T. Schindler, *ACS Omega*, 2023, **8**, 3017–3025.
36 N. I. Rinehart, R. K. Saunthwal, J. Wellauer, A. F. Zahrt, L. Schlemper, A. S. Shved, R. Bigler, S. Fantasia and S. E. Denmark, *Science*, 2023, **381**, 965–972.
37 B. Meyer, B. Sawatlon, S. Heinen, O. A. von Lilienfeld and C. Corminboeuf, *Chem. Sci.*, 2018, **9**, 7069–7077.
38 C. W. Coley, W. H. Green and K. F. Jensen, *Acc. Chem. Res.*, 2018, **51**, 1281–1289.
39 A. Tkatchenko, *Nat. Commun.*, 2020, **11**, 4125.
40 N. S. Eyke, B. A. Koscher and K. F. Jensen, *Trends Chem.*, 2021, **3**, 120–132.
41 W. Bort, I. I. Baskin, T. Gimadiev, A. Mukanov, R. Nugmanov, P. Sidorov, G. Marcou, D. Horvath, O. Klimchuk, T. Madzhidov and A. Varnek, *Sci. Rep.*, 2021, **11**, 3178.
42 A. F. Zahrt, Y. Mo, K. Y. Nandiwale, R. Shprints, E. Heid and K. F. Jensen, *J. Am. Chem. Soc.*, 2022, **144**, 22599–22610.
43 A. F. de Almeida, R. Moreira and T. Rodrigues, *Nat. Rev. Chem.*, 2019, **3**, 589–604.

 This journal is © The Royal Society of Chemistry 2025

44 I. W. Davies, *Nature*, 2019, **570**, 175–181.
45 P. M. Pflüger and F. Glorius, *Angew. Chem., Int. Ed.*, 2020, **59**, 18860–18865.
46 S. Mithani, G. Weeratunga, N. J. Taylor and G. I. Dmitrienko, *J. Am. Chem. Soc.*, 1994, **116**, 2209–2210.
47 S. J. Gould, N. Tamayo, C. R. Melville and M. C. Cone, *J. Am. Chem. Soc.*, 1994, **116**, 2207–2208.
48 R. Powers, *Expert Opin. Drug Discovery*, 2009, **4**, 1077–1098.
49 V. Exarchou, M. Krucker, T. A. van Beek, J. Vervoort, I. P. Gerothanassis and K. Albert, *Magn. Reson. Chem.*, 2005, **43**, 681–687.
50 J. Wu, P. Lorenzo, S. Zhong, M. Ali, C. P. Butts, E. L. Myers and V. K. Aggarwal, *Nature*, 2017, **547**, 436–440.
51 G. Saielli and A. Bagno, *Org. Lett.*, 2009, **11**, 1409–1412.
52 J. A. Porco, S. Su, X. Lei, S. Bardhan and S. D. Rychnovsky, *Angew. Chem., Int. Ed.*, 2006, **45**, 5790–5792.
53 S. D. Rychnovsky, *Org. Lett.*, 2006, **8**, 2895–2898.
54 S. Grimme, *Angew. Chem., Int. Ed.*, 2013, **52**, 6306–6312.
55 A. Howarth, K. Ermanis and J. M. Goodman, *Chem. Sci.*, 2020, **11**, 4351–4359.
56 I. Cortés, C. Cuadrado, A. Hernández Daranas and A. M. Sarotti, *Front. Nat. Prod.*, 2023, **2**, 1122426.
57 J. L. Tyler, A. Noble and V. K. Aggarwal, *Angew. Chem., Int. Ed.*, 2022, **61**, e202214049.
58 K. L. Skubi, J. B. Kidd, H. Jung, I. A. Guzei, M.-H. Baik and T. P. Yoon, *J. Am. Chem. Soc.*, 2017, **139**, 17186–17192.
59 N. S. Sheikh, D. Leonori, G. Barker, J. D. Firth, K. R. Campos, A. J. H. M. Meijer, P. O'Brien and I. Coldham, *J. Am. Chem. Soc.*, 2012, **134**, 5300–5308.
60 M. Baker, *Nature*, 2016, **533**, 452–454.
61 L. Pitzer, F. Schäfers and F. Glorius, *Angew. Chem., Int. Ed.*, 2019, **58**, 8572–8576.
62 K. A. Connors, *Chemical kinetics, The study of reaction rates in solution*, Wiley-VCH, New York, NJ, 1990.
63 D. G. Blackmond, *J. Am. Chem. Soc.*, 2015, **137**, 10852–10866.
64 A. P. Smalley and M. J. Gaunt, *J. Am. Chem. Soc.*, 2015, **137**, 10632–10641.
65 E. F. van der Eide and W. E. Piers, *Nat. Chem.*, 2010, **2**, 571–576.
66 J. V. Obligacion, S. P. Semproni, I. Pappas and P. J. Chirik, *J. Am. Chem. Soc.*, 2016, **138**, 10645–10653.
67 D. G. Blackmond, *Angew. Chem., Int. Ed.*, 2005, **44**, 4302–4320.
68 H. Xu, S. J. Zuend, M. G. Woll, Y. Tao and E. N. Jacobsen, *Science*, 2010, **327**, 986–990.
69 A. G. Doyle and E. N. Jacobsen, *Chem. Rev.*, 2007, **107**, 5713–5743.
70 R. S. J. Proctor, A. C. Colgan and R. J. Phipps, *Nat. Chem.*, 2020, **12**, 990–1004.
71 S. M. Banik, A. Levina, A. M. Hyde and E. N. Jacobsen, *Science*, 2017, **358**, 761–764.
72 A. Guerrero-Corella, A. Fraile and J. Alemán, *ACS Org. Inorg. Au*, 2022, **2**, 197–204.
73 M. E. Jung and G. Piizzi, *Chem. Rev.*, 2005, **105**, 1735–1766.
74 Q. Xiao and P. E. Floreancig, *Org. Lett.*, 2008, **10**, 1139–1142.
75 M. Wasa and J.-Q. Yu, *J. Am. Chem. Soc.*, 2008, **130**, 14058–14059.
76 H. Jiang, J. He, T. Liu and J.-Q. Yu, *J. Am. Chem. Soc.*, 2016, **138**, 2055–2059.
77 D. Leow, G. Li, T.-S. Mei and J.-Q. Yu, *Nature*, 2012, **486**, 518–522.
78 D. P. White, J. C. Anthony and A. O. Oyefeso, *J. Org. Chem.*, 1999, **64**, 7707–7716.
79 C. Cheng and J. F. Hartwig, *Science*, 2014, **343**, 853–857.
80 L. Rummel and P. R. Schreiner, *Angew. Chem., Int. Ed.*, 2024, **63**, e202316364.
81 J. Clayden, N. Greeves and S. G. Warren, *Organic chemistry*, Oxford University Press, Oxford, New York, 2012.
82 J. Xue, Y.-S. Zhang, Z. Huan, J.-D. Yang and J.-P. Cheng, *J. Am. Chem. Soc.*, 2023, **145**, 15589–15599.
83 J. Turkowska, J. Durka and D. Gryko, *Chem. Commun.*, 2020, **56**, 5718–5734.
84 N. J. Agard, J. A. Prescher and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2004, **126**, 15046–15047.
85 E. Kim and H. Koo, *Chem. Sci.*, 2019, **10**, 7835–7851.
86 K. Tokunaga, M. Sato, K. Kuwata, C. Miura, H. Fuchida, N. Matsunaga, S. Koyanagi, S. Ohdo, N. Shindo and A. Ojida, *J. Am. Chem. Soc.*, 2020, **142**, 18522–18531.
87 M. Macias-Contreras, H. He, K. N. Little, J. P. Lee, R. P. Campbell, M. Royzen and L. Zhu, *Bioconjugate Chem.*, 2020, **31**, 1370–1381.
88 S. Singh, I. S. Dubinsky-Davidchik and R. Kluger, *Org. Biomol. Chem.*, 2016, **14**, 10011–10017.
89 N. J. Agard, J. M. Baskin, J. A. Prescher, A. Lo and C. R. Bertozzi, *ACS Chem. Biol.*, 2006, **1**, 644–648.
90 L. Li and Z. Zhang, *Molecules*, 2016, **21**, 1393.
91 J. M. Lopchuk, K. Fjelbye, Y. Kawamata, L. R. Malins, C.-M. Pan, R. Gianatassio, J. Wang, L. Prieto, J. Bradow, T. A. Brandt, M. R. Collins, J. Elleraas, J. Ewanicki, W. Farrell, O. O. Fadeyi, G. M. Gallego, J. J. Mousseau, R. Oliver, N. W. Sach, J. K. Smith, J. E. Spangler, H. Zhu, J. Zhu and P. S. Baran, *J. Am. Chem. Soc.*, 2017, **139**, 3209–3226.
92 F. M. Bickelhaupt and K. N. Houk, *Angew. Chem., Int. Ed.*, 2017, **56**, 10070–10086.
93 P. Vermeeren, T. A. Hamlin and F. M. Bickelhaupt, *Chem. Commun.*, 2021, **57**, 5880–5896.
94 I. Fernández and F. M. Bickelhaupt, *Chem. Soc. Rev.*, 2014, **43**, 4953–4967.
95 S. Monfette and D. E. Fogg, *Chem. Rev.*, 2009, **109**, 3783–3816.
96 A. J. J. Lennox and G. C. Lloyd-Jones, *Angew. Chem., Int. Ed.*, 2012, **51**, 9385–9388.
97 E. H. Southgate, J. Pospech, J. Fu, D. R. Holycross and D. Sarlah, *Nat. Chem.*, 2016, **8**, 922–928.
98 M. Gómez-Gallego and M. A. Sierra, *Chem. Rev.*, 2011, **111**, 4857–4963.
99 E. M. Simmons and J. F. Hartwig, *Angew. Chem., Int. Ed.*, 2012, **51**, 3066–3072.
100 H. Hu, S.-J. Chen, M. Mandal, S. M. Pratik, J. A. Buss, S. W. Krska, C. J. Cramer and S. S. Stahl, *Nat. Catal.*, 2020, **3**, 358–367.

101 N. Elgrishi, K. J. Rountree, B. D. McCarthy, E. S. Rountree, T. T. Eisenhart and J. L. Dempsey, *J. Chem. Educ.*, 2018, **95**, 197–206.
102 F. Strieth-Kalthoff, M. J. James, M. Teders, L. Pitzer and F. Glorius, *Chem. Soc. Rev.*, 2018, **47**, 7190–7202.
103 J. Großkopf, T. Kratz, T. Rigotti and T. Bach, *Chem. Rev.*, 2022, **122**, 1626–1653.
104 L. Buzzetti, G. E. M. Crisenza and P. Melchiorre, *Angew. Chem., Int. Ed.*, 2019, **58**, 3730–3747.
105 M. Julliard and M. Chanon, *Chem. Rev.*, 1983, **83**, 425–506.
106 M. A. Cismesia and T. P. Yoon, *Chem. Sci.*, 2015, **6**, 5426–5434.
107 E. Arceo, I. D. Jurberg, A. Alvarez-Fernández and P. Melchiorre, *Nat. Chem.*, 2013, **5**, 750–756.
108 A. Madani and B. Pieber, *ChemCatChem*, 2023, **15**, e202201583.
109 R. Chung and J. E. Hein, *Top. Catal.*, 2017, **60**, 594–608.
110 M. Hunger and J. Weitkamp, *Angew. Chem., Int. Ed.*, 2001, **40**, 2954–2971.
111 M. Simonetti, D. M. Cannas, X. Just-Baringo, I. J. Vitorica-Yrezabal and I. Larrosa, *Nat. Chem.*, 2018, **10**, 724–731.
112 L. P. Hammett, *Chem. Rev.*, 1935, **17**, 125–136.
113 C. S. Yeung and V. M. Dong, *Chem. Rev.*, 2011, **111**, 1215–1292.
114 B. W. Michel, L. D. Steffens and M. S. Sigman, *J. Am. Chem. Soc.*, 2011, **133**, 8317–8325.
115 A. K. Cook and M. S. Sanford, *J. Am. Chem. Soc.*, 2015, **137**, 3109–3118.
116 R. S. Sanchez and F. A. Zhuravlev, *J. Am. Chem. Soc.*, 2007, **129**, 5824–5825.
117 A. I. Konovalov, A. Lishchynskyi and V. V. Grushin, *J. Am. Chem. Soc.*, 2014, **136**, 13410–13425.
118 J. M. Richter, B. W. Whitefield, T. J. Maimone, D. W. Lin, M. P. Castroviejo and P. S. Baran, *J. Am. Chem. Soc.*, 2007, **129**, 12857–12869.
119 A. R. Dick, J. W. Kampf and M. S. Sanford, *J. Am. Chem. Soc.*, 2005, **127**, 12790–12791.
120 N. R. Deprez and M. S. Sanford, *J. Am. Chem. Soc.*, 2009, **131**, 11234–11241.
121 S. Murahashi, T. Nakae, H. Terai and N. Komiya, *J. Am. Chem. Soc.*, 2008, **130**, 11005–11012.
122 D. Griller and K. U. Ingold, *Acc. Chem. Res.*, 1980, **13**, 317–323.
123 K. U. Ingold and D. A. Pratt, *Chem. Rev.*, 2014, **114**, 9022–9046.
124 E. C. Gentry, L. J. Rono, M. E. Hale, R. Matsuura and R. R. Knowles, *J. Am. Chem. Soc.*, 2018, **140**, 3394–3402.
125 G.-J. Cheng, X. Zhang, L. W. Chung, L. Xu and Y.-D. Wu, *J. Am. Chem. Soc.*, 2015, **137**, 1706–1725.
126 C. Yuan, Y. Liang, T. Hernandez, A. Berriochoa, K. N. Houk and D. Siegel, *Nature*, 2013, **499**, 192–196.
127 W. D. Hinsberg and P. B. Dervan, *J. Am. Chem. Soc.*, 1978, **100**, 1608–1610.
128 R. L. Hinman and K. L. Hamm, *J. Am. Chem. Soc.*, 1959, **81**, 3294–3297.
129 J. J. Campbell and S. A. Glover, *J. Chem. Soc., Perkin Trans. 2*, 1992, 1661–1663.
130 S. A. Glover, *Tetrahedron*, 1998, **54**, 7229–7271.
131 S. A. Glover, A. Rauk, J. M. Buccigross, J. J. Campbell, G. P. Hammond, G. Mo, L. E. Andrews and A.-M. E. Gillson, *Can. J. Chem.*, 2005, **83**, 1492–1509.
132 S. A. Glover and A. A. Rosser, *J. Phys. Org. Chem.*, 2015, **28**, 215–222.
133 S. H. Kennedy, B. D. Dherange, K. J. Berger and M. D. Levin, *Nature*, 2021, **593**, 223–227.
134 D. Habrant, V. Rauhala and A. M. P. Koskinen, *Chem. Soc. Rev.*, 2010, **39**, 2007–2017.
135 L. Buglioni, F. Raymenants, A. Slattery, S. D. A. Zondag and T. Noël, *Chem. Rev.*, 2022, **122**, 2752–2906.
136 C. Kingston, M. D. Palkowitz, Y. Takahira, J. C. Vantourout, B. K. Peters, Y. Kawamata and P. S. Baran, *Acc. Chem. Res.*, 2020, **53**, 72–83.
137 M. Dörr, M. M. Hielscher, J. Proppe and S. R. Waldvogel, *ChemElectroChem*, 2021, **8**, 2621–2629.
138 D. M. Heard and A. J. J. Lennox, *Angew. Chem., Int. Ed.*, 2020, **59**, 18866–18884.
139 M. B. Plutschack, B. Pieber, K. Gilmore and P. H. Seeberger, *Chem. Rev.*, 2017, **117**, 11796–11893.
140 C. C. Wagen, S. E. McMinn, E. E. Kwan and E. N. Jacobsen, *Nature*, 2022, **610**, 680–686.
141 C. N. Prieto Kullmer, J. A. Kautzky, S. W. Krska, T. Nowak, S. D. Dreher and D. W. C. MacMillan, *Science*, 2022, **376**, 532–539.
142 J. Y. Wang, J. M. Stevens, S. K. Kariofillis, M.-J. Tom, D. L. Golden, J. Li, J. E. Tabora, M. Parasram, B. Shields, D. Primer, B. Hao, D. Del Valle, S. DiSomma, A. Furman, G. G. Zipp, S. Melnikov, J. Paulson and A. Doyle, *Nature*, 2024, **626**, 1025–1033.
143 S. Krautwald and E. M. Carreira, *J. Am. Chem. Soc.*, 2017, **139**, 5627–5639.
144 G. Zhan, W. Du and Y.-C. Chen, *Chem. Soc. Rev.*, 2017, **46**, 1675–1692.
145 S. Brandau, E. Maerten and K. A. Jørgensen, *J. Am. Chem. Soc.*, 2006, **128**, 14986–14991.
146 D. A. Evans, J. V. Nelson, E. Vogel and T. R. Taber, *J. Am. Chem. Soc.*, 1981, **103**, 3099–3111.
147 T. G. Saint-Denis, R.-Y. Zhu, G. Chen, Q.-F. Wu and J.-Q. Yu, *Science*, 2018, **359**, eaao4798.
148 A. Fanourakis, P. J. Docherty, P. Chuentragool and R. J. Phipps, *ACS Catal.*, 2020, **10**, 10672–10714.
149 S.-H. Xiang and B. Tan, *Nat. Commun.*, 2020, **11**, 3786.
150 C. K. Winkler, J. H. Schrittwieser and W. Kroutil, *ACS Cent. Sci.*, 2021, **7**, 55–71.
151 J. F. Hartwig, *Organotransition metal chemistry, From bonding to catalysis*, University Science Books, California, 2010.
152 A. J. Sterling, S. Zavitsanou, J. Ford and F. Duarte, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2021, **11**, e1518.
153 I. T. Horváth, *Chem. Rev.*, 2018, **118**, 369–371.
154 H. Flerlage and J. C. Slootweg, *Nat. Rev. Chem.*, 2023, **7**, 593–594.
155 P. T. Anastas and J. C. Warner, *Green chemistry Theory and practice*, Oxford University Press, Oxford, 2000.
156 P. Anastas and N. Eghbali, *Chem. Soc. Rev.*, 2010, **39**, 301–312.

 This journal is © The Royal Society of Chemistry 2025

157 C. J. Clarke, W.-C. Tu, O. Levers, A. Bröhl and J. P. Hallett, *Chem. Rev.*, 2018, **118**, 747–800.
158 R. A. Sheldon and J. M. Woodley, *Chem. Rev.*, 2018, **118**, 801–838.
159 T. P. Yoon, M. A. Ischay and J. Du, *Nat. Chem.*, 2010, **2**, 527–532.
160 B. M. Trost, *Science*, 1991, **254**, 1471–1477.
161 A. D. Curzons, D. N. Mortimer, D. J. C. Constable and V. L. Cunningham, *Green Chem.*, 2001, **3**, 1–6.
162 E. R. Monteith, P. Mampuys, L. Summerton, J. H. Clark, B. U. W. Maes and C. R. McElroy, *Green Chem.*, 2020, **22**, 123–135.
163 C. Jimenez-Gonzalez, C. S. Ponder, Q. B. Broxterman and J. B. Manley, *Org. Process Res. Dev.*, 2011, **15**, 912–917.
164 D. Cespi, E. S. Beach, T. E. Swarr, F. Passarini, I. Vassura, P. J. Dunn and P. T. Anastas, *Green Chem.*, 2015, **17**, 3390–3400.
165 E. C. Sherer, A. Bagchi, B. Kosjek, K. M. Maloney, Z. Peng, S. A. Robaire, R. P. Sheridan, E. Metwally and L.-C. Campeau, *Org. Process Res. Dev.*, 2022, **26**, 1405–1410.
166 R. A. Sheldon, *Green Chem.*, 2007, **9**, 1273–1283.
167 R. A. Sheldon, *Green Chem.*, 2017, **19**, 18–43.
168 F. Roschangar, R. A. Sheldon and C. H. Senanayake, *Green Chem.*, 2015, **17**, 752–768.
169 T. Welton, *Proc. Math. Phys. Eng. Sci.*, 2015, **471**, 20150502.
170 E. Roduner, *Chem. Soc. Rev.*, 2014, **43**, 8226–8239.
171 O. Mitsunobu and M. Yamada, *Bull. Chem. Soc. Jpn.*, 1967, **40**, 2380–2382.
172 O. Mitsunobu, M. Yamada and T. Mukaiyama, *Bull. Chem. Soc. Jpn.*, 1967, **40**, 935–939.
173 T. Y. S. But and P. H. Toy, *J. Am. Chem. Soc.*, 2006, **128**, 9636–9637.
174 R. H. Beddoe, K. G. Andrews, V. Magné, J. D. Cuthbertson, J. Saska, A. L. Shannon-Little, S. E. Shanahan, H. F. Sneddon and R. M. Denton, *Science*, 2019, **365**, 910–914.
175 C. M. Clouthier and J. N. Pelletier, *Chem. Soc. Rev.*, 2012, **41**, 1585–1605.
176 H. Renata, Z. J. Wang and F. H. Arnold, *Angew. Chem., Int. Ed.*, 2015, **54**, 3351–3367.
177 R. K. Zhang, K. Chen, X. Huang, L. Wohlschlager, H. Renata and F. H. Arnold, *Nature*, 2019, **565**, 67–72.
178 I. Šolić, H. X. Lin and R. W. Bates, *Tetrahedron Lett.*, 2018, **59**, 4434–4436.
179 Y.-B. Kang and L. H. Gade, *J. Am. Chem. Soc.*, 2011, **133**, 3658–3667.
180 Y. Li, D. Song and V. M. Dong, *J. Am. Chem. Soc.*, 2008, **130**, 2962–2964.
181 J. Seayad, A. M. Seayad and C. L. L. Chai, *Org. Lett.*, 2010, **12**, 1412–1415.
182 M. Terada, *Synthesis*, 2010, 1929–1982.
183 D. W. C. MacMillan, *Nature*, 2008, **455**, 304–308.
184 S. Bertelsen and K. A. Jørgensen, *Chem. Soc. Rev.*, 2009, **38**, 2178–2189.
185 A. H. Cherney, N. T. Kadunce and S. E. Reisman, *Chem. Rev.*, 2015, **115**, 9587–9652.
186 B. List, R. A. Lerner and C. F. Barbas, *J. Am. Chem. Soc.*, 2000, **122**, 2395–2396.
187 K. A. Ahrendt, C. J. Borths and D. W. C. MacMillan, *J. Am. Chem. Soc.*, 2000, **122**, 4243–4244.
188 S. Kozuch and J. M. L. Martin, *ACS Catal.*, 2012, **2**, 2787–2794.
189 C. A. Salazar, K. N. Flesch, B. E. Haines, P. S. Zhou, D. G. Musaev and S. S. Stahl, *Science*, 2020, **370**, 1454–1460.
190 M. C. Kozlowski, *Org. Lett.*, 2022, **24**, 7247–7249.
191 T. Gensch and F. Glorius, *Science*, 2016, **352**, 294–295.
192 M. P. Maloney, C. W. Coley, S. Genheden, N. Carson, P. Helquist, P.-O. Norrby and O. Wiest, *Org. Lett.*, 2023, **25**, 2945–2947.
193 F. Strieth-Kalthoff, F. Sandfort, M. Kühnemund, F. R. Schäfer, H. Kuchen and F. Glorius, *Angew. Chem., Int. Ed.*, 2022, **61**, e202204647.
194 D. Rana, P. M. Pflüger, N. P. Hölter, G. Tan and F. Glorius, *ACS Cent. Sci.*, 2024, **10**, 899–906.
195 S. K. Kariofillis, S. Jiang, A. M. Żurański, S. S. Gandhi, J. I. Martinez Alvarado and A. G. Doyle, *J. Am. Chem. Soc.*, 2022, **144**, 1045–1055.
196 K. D. Collins and F. Glorius, *Nat. Chem.*, 2013, **5**, 597–601.
197 J. Feldman, J. S. Murdzek, W. M. Davis and R. R. Schrock, *Organometallics*, 1989, **8**, 2260–2265.
198 G. C. Fu and R. H. Grubbs, *J. Am. Chem. Soc.*, 1992, **114**, 7324–7325.
199 A. Fürstner and K. Langemann, *J. Am. Chem. Soc.*, 1997, **119**, 9130–9136.
200 K. C. K. Swamy, N. N. B. Kumar, E. Balaraman and K. V. P. P. Kumar, *Chem. Rev.*, 2009, **109**, 2551–2651.
201 T. Tsunoda, Y. Yamamiya, Y. Kawamura and S. Itô, *Tetrahedron Lett.*, 1995, **36**, 2529–2530.
202 O. Mitsunobu, *Synthesis*, 1981, 1–28.
203 S. Fletcher, *Org. Chem. Front.*, 2015, **2**, 739–752.
204 H. Mayr and M. Patz, *Angew. Chem., Int. Ed. Engl.*, 1994, **33**, 938–957.
205 I. S. Kim, M.-Y. Ngai and M. J. Krische, *J. Am. Chem. Soc.*, 2008, **130**, 14891–14899.
206 I. S. Kim, M.-Y. Ngai and M. J. Krische, *J. Am. Chem. Soc.*, 2008, **130**, 6340–6341.
207 S. B. Han, A. Hassan, I. S. Kim and M. J. Krische, *J. Am. Chem. Soc.*, 2010, **132**, 15559–15561.
208 J. Feng, Z. A. Kasun and M. J. Krische, *J. Am. Chem. Soc.*, 2016, **138**, 5467–5478.
209 S. G. Aiken, J. M. Bateman, H.-H. Liao, A. Fawcett, T. Bootwicha, P. Vincetti, E. L. Myers, A. Noble and V. K. Aggarwal, *Nat. Chem.*, 2023, **15**, 248–256.
210 T. Cernak, K. D. Dykstra, S. Tyagarajan, P. Vachal and S. W. Krska, *Chem. Soc. Rev.*, 2016, **45**, 546–576.
211 L. Guillemard, N. Kaplaneris, L. Ackermann and M. J. Johansson, *Nat. Rev. Chem.*, 2021, **5**, 522–545.
212 J. Jurczyk, J. Woo, S. F. Kim, B. D. Dherange, R. Sarpong and M. D. Levin, *Nat. Synth.*, 2022, **1**, 352–364.
213 M. Peplow, *Nature*, 2023, **618**, 21–24.
214 J. Jurczyk, M. C. Lux, D. Adpressa, S. F. Kim, Y.-H. Lam, C. S. Yeung and R. Sarpong, *Science*, 2021, **373**, 1004–1012.

215 J. C. Reisenbauer, O. Green, A. Franchino, P. Finkelstein and B. Morandi, *Science*, 2022, **377**, 1104–1109.

216 G. L. Bartholomew, F. Carpaneto and R. Sarpong, *J. Am. Chem. Soc.*, 2022, **144**, 22309–22315.

217 F. Berger, M. B. Plutschack, J. Riegger, W. Yu, S. Speicher, M. Ho, N. Frank and T. Ritter, *Nature*, 2019, **567**, 223–228.

218 N. Stephanopoulos and M. B. Francis, *Nat. Chem. Biol.*, 2011, **7**, 876–884.

219 M. S. Kang, T. W. S. Kong, J. Y. X. Khoo and T.-P. Loh, *Chem. Sci.*, 2021, **12**, 13613–13647.

220 J. Y. Axup, K. M. Bajjuri, M. Ritland, B. M. Hutchins, C. H. Kim, S. A. Kazane, R. Halder, J. S. Forsyth, A. F. Santidrian, K. Stafin, Y. Lu, H. Tran, A. J. Seller, S. L. Biroc, A. Szydlik, J. K. Pinkstaff, F. Tian, S. C. Sinha, B. Felding-Habermann, V. V. Smider and P. G. Schultz, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 16101–16106.

221 J. Kaur, M. Saxena and N. Rishi, *Bioconjugate Chem.*, 2021, **32**, 1455–1471.

222 S. T. Laughlin, J. M. Baskin, S. L. Amacher and C. R. Bertozzi, *Science*, 2008, **320**, 664–667.

223 X. Zhang, C. Gong, O. U. Akakuru, Z. Su, A. Wu and G. Wei, *Chem. Soc. Rev.*, 2019, **48**, 5564–5595.

224 J. Luke, E. J. Yang, C. Labanti, S. Y. Park and J.-S. Kim, *Nat. Rev. Mater.*, 2023, **8**, 839–852.

225 A. D. Scaccabarozzi, A. Basu, F. Aniés, J. Liu, O. Zapata-Arteaga, R. Warren, Y. Firdaus, M. I. Nugraha, Y. Lin, M. Campoy-Quiles, N. Koch, C. Müller, L. Tsetseris, M. Heeney and T. D. Anthopoulos, *Chem. Rev.*, 2022, **122**, 4420–4492.

226 C. M. Crudden, J. H. Horton, I. I. Ebralidze, O. V. Zenkina, A. B. McLean, B. Drevniok, Z. She, H.-B. Kraatz, N. J. Mosey, T. Seki, E. C. Keske, J. D. Leake, A. Rousina-Webb and G. Wu, *Nat. Chem.*, 2014, **6**, 409–414.

227 J. C. Love, L. A. Estroff, J. K. Kriebel, R. G. Nuzzo and G. M. Whitesides, *Chem. Rev.*, 2005, **105**, 1103–1169.

228 P. Li, M. R. Ryder and J. F. Stoddart, *Acc. Mater. Res.*, 2020, **1**, 77–87.

229 H. Li, A. Bunrit, N. Li and F. Wang, *Chem. Soc. Rev.*, 2020, **49**, 3748–3763.

230 C. Jehanno, M. M. Pérez-Madrigal, J. Demarteau, H. Sardon and A. P. Dove, *Polym. Chem.*, 2019, **10**, 172–186.

231 Y.-M. Siu, J. Roane and M. J. Krische, *J. Am. Chem. Soc.*, 2021, **143**, 10590–10595.

232 Y. Lu, S. K. Woo and M. J. Krische, *J. Am. Chem. Soc.*, 2011, **133**, 13876–13879.

233 F. Perez, A. R. Waldeck and M. J. Krische, *Angew. Chem., Int. Ed.*, 2016, **55**, 5049–5052.

234 J. H. Hsu, T. E. Ball, S. Oh, E. E. Stache and B. P. Fors, *Angew. Chem., Int. Ed.*, 2024, **63**, e202316578.

235 H. Jiang and E. E. Stache, *Chem*, 2023, **9**, 2373–2375.

236 S. T. Nguyen, L. R. Fries, J. H. Cox, Y. Ma, B. P. Fors and R. R. Knowles, *J. Am. Chem. Soc.*, 2023, **145**, 11151–11160.

237 S. T. Nguyen, E. A. McLoughlin, J. H. Cox, B. P. Fors and R. R. Knowles, *J. Am. Chem. Soc.*, 2021, **143**, 12268–12277.

238 S. T. Nguyen, P. R. D. Murray and R. R. Knowles, *ACS Catal.*, 2020, **10**, 800–805.

239 R. Mercado, S. M. Kearnes and C. W. Coley, *J. Chem. Inf. Model.*, 2023, **63**, 4253–4265.

240 S. Herres-Pawlis, F. Bach, I. J. Bruno, S. J. Chalk, N. Jung, J. C. Liermann, L. R. McEwen, S. Neumann, C. Steinbeck, M. Razum and O. Koepler, *Angew. Chem., Int. Ed.*, 2022, **61**, e202203038.

241 N. Artrith, K. T. Butler, F.-X. Coudert, S. Han, O. Isayev, A. Jain and A. Walsh, *Nat. Chem.*, 2021, **13**, 505–508.

242 F. Strieth-Kalthoff, F. Sandfort, M. H. S. Segler and F. Glorius, *Chem. Soc. Rev.*, 2020, **49**, 6154–6168.

 This journal is © The Royal Society of Chemistry 2025

Supplementary Information (SI) for Chemical Society Reviews.
This journal is © The Royal Society of Chemistry 2025

Reaction Development: A Student's Checklist

- Take Stock
- Kinetics & Thermodynamics
- Mechanism
- Optimisation
- Catalysis
- Scope
- Applications

Take Stock

Have you made the desired compound?
- Have you obtained a range of characterisation data?
- Have you obtained the X-ray crystal structure?

Have you determined where all the atoms go?
- Can you write out a balanced equation?
- Where do all of the atoms in the product come from?
- What are the byproducts and side-products?
- Is there involvement from solvent, water or oxygen?

Are all reaction conditions necessary?
- How long does it take to reach full conversion?
- Are you operating at the lowest feasible temperature?
- What is the lowest equivalency of reagent that can be used?
- Are all reagents required to achieve product formation?

Is the reaction reproducible and robust?
- Can you reproduce the yield of the reaction?
- Can the purification conditions be reliably reproduced?
- Have you performed a sensitivity screen?

Have you found the minimal synthon and retron?
- Can the substrate or reagents be simplified?
- Can the reactivity be extended to intermolecular systems?
- Is the minimal retron found in natural products?

Kinetics & Thermodynamics

Can the rate law be determined?
- What reagents are involved in the rate-determining step?
- Can you calculate the rate constant?

Can you increase the rate?
- What is the effect on the rate of increasing temperature?
- Can you harness the Thorpe-Ingold effect?

Have you established the steric sensitivity?
- What are the steric limits of the reaction?
- Can you favour reactivity using intramolecularisation?

Have you identified the reaction driving force?
- What is the thermodynamic driving force?
- Can you estimate the value of ΔH and ΔS?
- Can you run the reaction at a higher pressure?
- Can you destabilise the starting materials?

Can you bias an equilibrium?
- Can you stabilise an intermediate at equilibrium?
- Can you evolve or precipitate the product or byproduct?

Is the intended product stable?

Mechanism

Do you have evidence for the mechanism?
- Can you measure kinetic isotope effects?
- Can you monitor the reaction with in situ spectroscopy?
- Are the bond forming steps concerted or stepwise?
- Can you conceive of a radical clock experiment?
- Can you measure the quantum yield of the reaction?
- Can you estimate the frontier molecular orbitals?
- Can you probe the reaction with UV-vis spectroscopy?
- Can you calculate the triplet energy of the substrates?
- Can you perform a Hammett analysis?
- Can you analyse the kinetic profile of the reaction?
- Can you measure redox potentials with cyclic voltammetry?
- Can you perform a Stern-Volmer quenching study?

Have the intermediates be identified?
- Can an intermediate be isolated or detected by spectroscopy?
- Can you intercept an intermediate using a trapping agent?

Can the reaction pathway be computed?
- Are the proposed transition state energies feasible?
- Can you rationalise selectivity using computation?

Is there a better route to the active species?
- Can the route to the active species be simplified?
- How does the method compare to similar transformations?

Optimisation

Has the yield been fully maximised?

Have you optimised the...
- Reaction time
- Temperature
- Stoichiometry
- Concentration
- Catalyst loading
- Solvent
- Mixing rate
- Reagent addition rate
- Pressure
- Solution pH

Have you considered...
- Reagent addition order
- Additives
- Reaction atmosphere
- Wavelength of light
- Light intensity
- Number of light sources
- Waste production
- Electrode material
- Current density
- Inter-electrode distance
- Residence time
- Flow rate
- Sustainability
- Reaction cost

Can reaction selectivity be achieved?
- Can chemo-/regio/or diastereoselectivity be achieved?
- Can the reaction be rendered enantioselective?
- Can the conditions access divergent reactivity?

Can the protocol be more efficient?
- Can you compare PMI and E-Factor to similar syntheses?
- Can the reaction be run in air or at ambient temperature?
- Can any step of the procedure be automated?
- Can the reaction be performed with less toxic reagents?

Catalysis

Can the reaction be catalysed?
- Can the reaction be catalysed by an enzyme?
- Can a reagent be turned over using a redox agent?
- Can the reaction be catalysed in the excited state?
- Have you confirmed the existence of a catalytic cycle?

Have you identified the catalytically active species?
- Can you detect the active species using spectroscopy?
- Can you find a relationship between pre and active catalysts?
- Are nanoparticles responsible for catalytic activity?
- Can a better strategy to form the active catalyst be conceived?

Can you catalyse the reaction enantioselectively?
- Can the reaction be used for desymmeterisation?
- Can an organocatalyst be employed?
- Is there a relationship between sterics and enantioselectivity?
- Have you identified the enantiodetermining step?

Is the catalyst turnover optimal?
- Can Turnover Number (TON) be calculated?
- Can catalyst deactivation pathways be suppressed?
- Can you reduced the use of precious metals?
- Is there a relationship between ligand identity and TON?

Scope

Is the substrate scope diverse?
- Have you explored a wide range of substrates?
- Have you attempted to eliminate selection and reporting bias?
- Can complex substrates be tolerated?
- Have a variety of steric environments been explored?
- Have you reported the 'negative' results?
- Can a standardised scope selection model be used?

Have you shown functional group tolerance?
- Can the reaction tolerate nucleophiles or electrophiles?
- Can you tolerate substrates that can be oxidised or reduced?
- Can you tolerate Lewis acids and Lewis bases?
- Can you tolerate fragments found frequently in drugs?
- Can the reaction be performed in a range of solvents?
- Have you employed a robustness screen?

Have you defined the parameters of reactivity?
- Can you define a pKa range in which reactivity is observed?
- Is there a steric limit to reactivity?
- Can you define a redox window for the starting materials?
- Have you considered the mechanistic implications of the scope?
- Can the Mayr scale be used to define the limits of electrophilicity, nucleophilicity and radical philicity?

Applications

Can it be used to access target compounds?
- Can the reaction be applied to natural product synthesis?
- Can the reaction be used to access motifs found in drugs?
- Can you access building blocks for functional materials?

Can it be used for late-stage transformations?
- Can the reaction tolerate dense functionality?
- Can the reaction be used for skeletal editing?

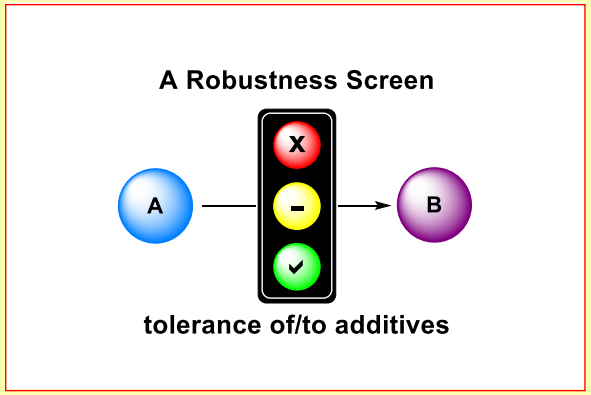
Can the products be derivatised?
- Do you form a linchpin group that can be further reacted?
- Can functional group interconversion be demonstrated?
- Can the reaction be used as part of a cascade?

Can it be used for bioconjugation?
- Are the conditions suitably mild for in vivo functionalisation?
- Can the reaction target a native amino acid side chain?
- Can the reaction be run in a cell or in an organism?
- Can click-type reactivity between biomolecules be achieved?

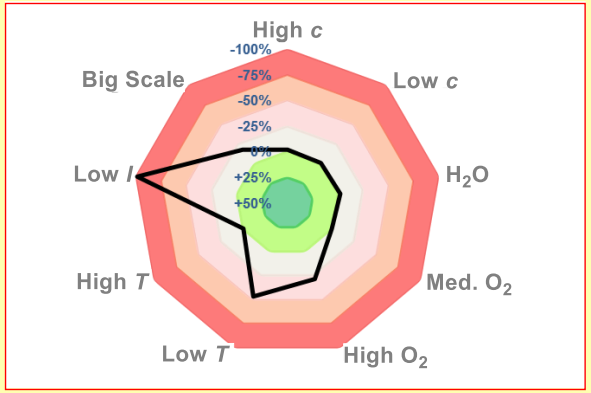
Can it be used in materials chemistry?
- Can the reaction be used to depolymerise plastics?
- Can the functionalisation of a plastic be achieved?

Note: Given the significant number of variables associated with individual synthetic methodology studies, all of the themes displayed above are not directly applicable to every new reaction or are a necessity for publication.

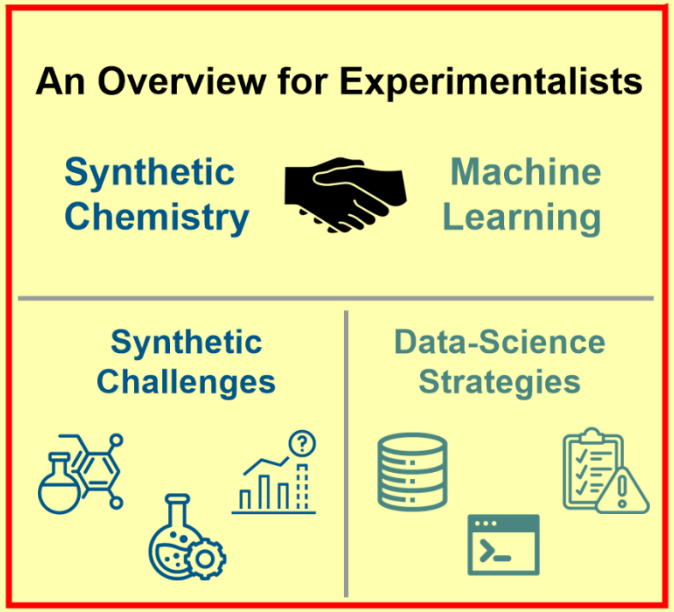
Review: N. Hölter, F. Katzenburg, F. Boser, F. Glorius, *Angew. Chem. Novit* **2026**, *2*, *e70032.*
Machine Learning for Synthetic Organic Chemistry: Methods, Applications, and Best Practices



Robustness Screen:
Collins, Glorius, *Nature Chem.* **2013**, *5, 597*

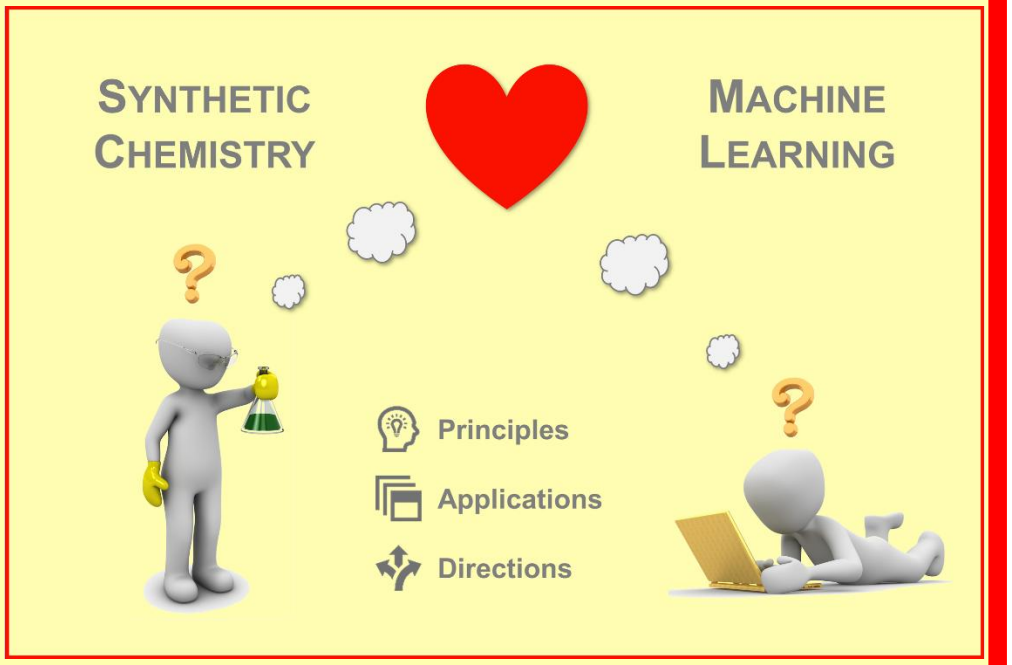


Sensitivity Screen:
Pitzer, Schäfers, Glorius, *Angew. Chem. Int. Ed.* **2019**, *58,* 8572.



DATA!!

Tutorial Review: F. Strieth-Kalthoff, F. Sandfort, M. H. S. Segler, F. Glorius, *Chem. Soc. Rev.* **2020**, *49*, 6154-6168.
Machine Learning the Ropes: Principles, Applications and Directions in Synthetic Chemistry



Check for updates

Angewandte Chemie Novit

Angewandte Novit Chemie GDCh German Chemical Society

REVIEW OPEN ACCESS

Machine Learning for Synthetic Organic Chemistry: Methods, Applications, and Best Practices

Niklas Hölter | Felix Katzenburg | Florian Boser | Frank Glorius

Organisch-Chemisches Institut, Universität Münster, Münster, Germany

Correspondence: Frank Glorius (glorius@uni-muenster.de)

Received: 25 March 2026 | **Revised:** 17 June 2026

Keywords: artificial intelligence | data science | digital chemistry | molecular machine learning | synthetic chemistry

ABSTRACT

The growing integration of artificial intelligence (AI) and machine learning (ML) is transforming experimental chemistry laboratories. Especially in synthetic chemistry, researchers routinely handle complex and high-dimensional data, fostering meaningful synergies between chemistry and data science. This review is intended as a practical overview that connects the everyday challenges of synthetic chemists with the digital tools available to address them. It does not seek to explain theoretical foundations of ML or to provide a comprehensive survey of all recent studies in the field. Rather, our goal is to highlight emerging technologies, discuss key considerations for their application, and present a selection of illustrative examples. To begin, we outline the prerequisites for successfully applying data science in synthetic chemistry. Next, we give a realistic overview of strategies and bottlenecks in predictive modeling of molecular properties, reaction outcomes and reaction conditions. We further highlight data-driven approaches that can be applied in the development of new chemical reactions and synthetic methodologies, including all relevant stages from reaction discovery and optimization to substrate scope evaluation and mechanistic analyses. Finally, we briefly discuss the transformative role of large language models and agentic workflows in synthetic chemistry, focusing on opportunities and challenges in the laboratories of the future.

1 | Introduction

In synthetic chemistry, artificial intelligence (AI) and machine learning (ML) have recently begun to move toward laboratory practice. This momentum builds on a long history of AI research that dates back to the 1956 Dartmouth computer science workshop, often considered the birth of AI research [1]. Early chemistry applications soon followed, including the DENDRAL system for mass-spectral structure elucidation (1965) [2, 3] and Corey and Wipke's first proposal for computer-assisted synthesis planning (1969) [4]. However, it took the field another 60 years to solve technical challenges, which ultimately led to a visible hype phase, fueled by recent general headline AI advances. Despite the momentum, this hype eventually seemed to cool down in synthetic chemistry, where many AI models proved hard to deploy in real laboratories due to data limitations, workflow integration hurdles, and computational cost. Now, with ongoing developments in the field, modern data science techniques are slowly becoming useful tools for everyday laboratory practice. Ultimately, this transition toward real impact was recognized by the 2024 Nobel Prize in Chemistry for computational protein design and AI-enabled protein-structure prediction [5, 6].

Recent advances in AI, molecular ML, cheminformatics, and laboratory automation offer new opportunities to accelerate synthetic chemistry. Molecular ML and data-science tools are

Niklas Hölter and Felix Katzenburg contributed equally to this work. The order of authorship was determined by coin toss.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Angewandte Chemie Novit* published by Wiley-VCH GmbH on behalf of Gesellschaft Deutscher Chemiker (GDCh; the German Chemical Society).

moving beyond proof-of-concept demonstrations and are entering chemists' everyday workflows [7–9]. Growing repositories of high-quality data, accessible software, and networked laboratory equipment now allow chemists to predict molecular properties before synthesis [10–12], design optimization campaigns that converge in fewer iterations [13–15], plan synthesis routes [16, 17] and delegate routine operations to automated platforms [17–19]. In practice, this translates to efficient molecular design, optimization campaigns that reach optimal conditions in tens rather than hundreds of reactions, and less time spent on experimental planning and execution. The resulting efficiency increase reduces resource consumption, cuts the number of failed reactions, and frees researcher time for tasks that machines cannot yet perform [20]. Coupled with high-throughput experimentation, these recent advances in data science and ML have been particularly consequential and enabling for synthetic chemistry, a field in which combinatorial reaction-condition spaces, nonadditive molecular interactions, and poor inter-laboratory reproducibility have long challenged even expert practitioners.

Given these developments, the question is no longer whether to use molecular ML, but when and how to apply it effectively [21]. This review examines how ML can support the central tasks of synthetic chemistry. It is organized into four parts. The first part introduces the general prerequisites and best practices of molecular ML. We then describe how ML can predict relevant molecular properties and guide molecular design, before outlining how data-driven tools can support the key stages of synthetic method development. Finally, we discuss how large language models (LLMs) and agentic systems can be integrated into chemical research. Rather than surveying model architectures or providing a comprehensive account of prior work, we focus on how these methods address concrete synthetic problems. We organize discussions around common challenges in synthesis and map each to the ML strategies and technologies best suited to address it. For each case, we highlight the key decisions, data requirements, integration with experimental workflows, and point out actionable insights and common pitfalls. Selected recent examples illustrate the principles, while technical details are deferred to focused reviews.

2 | Prerequisites for the Successful Application of ML in Synthetic Chemistry

A common misconception is that adopting ML tools for synthetic problems poses enormous entry barriers and requires years of training in computer science or statistics. Likewise, mathematical backgrounds might intimidate interested users, leading to the assumption that meaningful engagement with ML is reserved only for specialists. This perception can discourage otherwise capable researchers from exploring ML approaches, even when these tools could offer practical and accessible ways to enhance experimental design, optimize workflows, and extract deeper insights from data. In reality, most end users do not need to develop or even implement models from scratch, and with a solid foundation in data science concepts and basic programming skills, it is entirely feasible to start applying existing tools effectively.

2.1 | A Digital Skillset for Experimental Chemists

The growing use of data science in synthetic chemistry quietly reshapes the skill set expected of experimental chemists. The undoubtedly most central skill is the ability to write, read, and understand simple code. While this requirement seems like a recent paradigm shift, it has been proposed since the late 1960s, albeit then in disciplines such as physical and theoretical chemistry [22–24]. Today, basic programming in a high-level language like Python is not only considered the entry point to ML but is becoming routine practice in many disciplines. Treating manual or AI-assisted coding as a core skill for synthetic chemists is therefore not radical, but overdue [25, 26].

With basic coding skills, the entry barrier to cheminformatics and molecular ML drops significantly. Typically, cheminformatics comes first, providing tools to handle molecular data in digital form. This includes understanding common identifiers, file formats, and routine tasks such as data parsing, cleaning, standardization, removal of salts or solvents, deduplication, and simple table operations. Closely related to cheminformatics are basic data science operations such as exploratory and statistical data analysis. Chemists must learn to explore and visualize data, distinguish correlation from causation, and identify biases, artifacts and noise early on. This foundation enables building clean, well-annotated datasets from experimental or public sources. From here, learning to train predictive ML models requires only a small additional effort. At more advanced levels, also software-engineering best practices such as version control, clear documentation, unit testing, and adherence to FAIR (findable, accessible, interoperable, and reusable) [27] data principles become relevant. However, all these stages are nowadays heavily facilitated by LLMs, which have become indispensable assistants for coding tasks and even enable the creation of complex software with beginner-level skills (see Section 5). Last, a solid foundation of digital skills also increasingly enables synthetic chemists to operate systems beyond ML algorithms, for example, running and analyzing simple quantum-chemical (QM) calculations at scale or interfacing with programmable laboratory hardware, such as liquid handlers, analytical devices, or sensors [28, 29].

To summarize, in a digital era, programming and data science are becoming increasingly important scientific skills and mastering them is as qualifying as mastering any instrument in the laboratory. Digital training should not be confined to specialist tracks; it should be integrated throughout the undergraduate curriculum and tied to real laboratory problems [29, 30]. However, we state that digital skills alone are not enough. Chemical expertise remains essential for three main reasons, which we will discuss in detail in this review. First, chemists define the hypotheses, scope, and success criteria of the campaign and select computational methods that are suited to address a given problem. Second, experimental data is often scarce, costly, or biased. In such small-data settings, domain knowledge is crucial for selecting meaningful descriptors (see Section 2.2) and applying physically informed constraints that make models reliable. Third, chemists always serve as the final filter and evaluate the model's outputs in a chemical context (see Section 2.4). It remains up to the chemists to ensure that ML campaign results translate into meaningful, and most importantly, interpretable experimental results.

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.2 | Selecting Molecular Representations and Model Architectures

When planning an ML-supported project in synthetic chemistry, a basic understanding of ML foundations is naturally indispensable. As a detailed introduction of fundamental data science concepts is beyond the scope of this review, we refer to several excellent reviews addressing the basics of ML, molecular representations, model architectures and algorithms [31–36]. Here, we instead focus on decision guidelines on which technique to use for which scenarios, while outlining advantages and limitations of individual methods.

As discussed in the next chapter (see Section 2.3), high-quality data is the crucial factor determining the scope and limitations of any ML model. However, molecular input data such as structures or reaction schemes is often non-numeric. Therefore, it needs to be converted into a machine-readable format—a molecular representation—to be processed by ML models. Simplified, molecular representations differ in their level of abstraction. Very universal representations, such as molecular graphs, structure-based fingerprints [37, 38], or text-based string formats such as SMILES [39] or SELFIES [40] are closely related to a molecular structure and broadly applicable, but do not directly encode decisive molecular properties as information that is relevant to a given task. In contrast, tailored representations such as vectors of physical descriptors or problem-specific fingerprints [41, 42] often contain information that is directly related to the prediction target. In turn, they cannot be considered general and transferable to other tasks. Therefore, specialized deep learning model architectures such as graph neural networks (GNNs) or transformer models [43] are typically required when working with very general graph- or text-based representations (vide infra). Vectorized representations such as molecular fingerprints or physical descriptors can usually be processed by both traditional low-capacity models such as multivariate linear regression (MLR), support vector machines (SVM), random forests (RFs), or gradient boosting (GB) models, as well as modern deep learning models, like neural networks.

As a rule of thumb, there is an inherent trade-off between data volume and representation specificity: the less data available, the more expert knowledge must be infused into the representation to ensure a predictive and generalizable model. Consequently, expert-informed representations are typically preferred in combination with simple ML models when only small amounts, typically around 10^1 to 10^3 datapoints, are available. Building prior knowledge of reaction mechanisms and electronic structure into the representation means the model does not need to infer this information from molecular structure alone. Examples of specific, engineered molecular representations include experimentally derived or calculated descriptors (e.g., NMR shifts, dipole moments, or steric parameters) or specialized geometrical descriptors capturing 3D structures of conformational ensembles of molecules. Even though this type of featurization often requires expert-domain knowledge for feature engineering, for example, to define the features and/or to run QM calculations, the combination with low-capacity ML models mitigates overfitting risks and allows for direct interpretation of feature importance. These advantages still render expert-selected (QM) descriptors a popular choice for ML modeling in small data regimes [44–46]. It is noteworthy that recent developments in the areas of semiempirical methods [47, 48], ML interatomic potentials (MLIPs) [49], end-to-end programmatic [50–54] or even agentic workflows (see Section 5.3) have significantly facilitated the calculation of QM-descriptors. Furthermore, an increasing number of QM-based descriptor libraries for different compound classes is now also deposited in repositories such as the MolSSI descriptor library project [55], improving their findability, accessibility and reusability.

In contrast, so-called feature learning becomes feasible with growing dataset sizes and the usage of powerful deep learning models, often in regimes with $> 10^4$ datapoints available. Feature learning describes that modern deep learning models with many parameters can automatically extract task-relevant information from universal representations such as molecular graphs or text-based representations. Thereby, feature learning eliminates the need for expert-guided feature engineering, but the trade-off is data hunger: without explicit problem-specific descriptors, these models typically require substantially more examples to be trained. On the other hand, deep learning models can often outperform smaller models, assuming they were trained on large chemical datasets. Furthermore, hybrid approaches exist in which molecular graphs are augmented with chemical knowledge in the form of (QM) descriptors attached to the graph, promising better generalizability with less training data [56–60].

While these simplified guidelines are often successfully applied in practice, it has recently been observed that the traditional link between model complexity and required dataset size is fading. Modern techniques allow powerful, complex models to be used even when data is scarce. An established strategy is transfer learning, where instead of training a large model from scratch on a small dataset, the model is first pretrained on a related task where plenty of data is available [61]. This pretrained model, which has already learned to extract valuable information, is then fine-tuned on a small dataset to adapt it to the specific task of interest. Typically, this approach works best when the pretraining task is closely related to the final application. However, pushing the concept further resulted in the development of so-called foundation models. Here, the strategy shifts from task-specific pretraining toward creating a single, powerful model with a comprehensive, almost encyclopedic knowledge base. This is achieved by training on massive, heterogeneous datasets that cover a wide spectrum of tasks and information within a domain, for example, combining molecular structures, corresponding molecular properties and chemical reactions. The ultimate objective is to instill a fundamental understanding of the underlying principles and relationships of the field. Consequently, this pre-packaged expertise serves as a versatile starting point for fine-tuning for a variety of specific downstream applications [62]. As foundation models typically (need to) operate on a very general molecular representation, these are typically models that operate on graph- [63] or text-based [64] molecular data. Last, it is important to conclude that neither molecular representation nor model choice is universally optimal. In every ML campaign, different configurations of ML models, algorithms, hyperparameters, and featurization techniques need to be compared and benchmarked to obtain an optimal model.

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

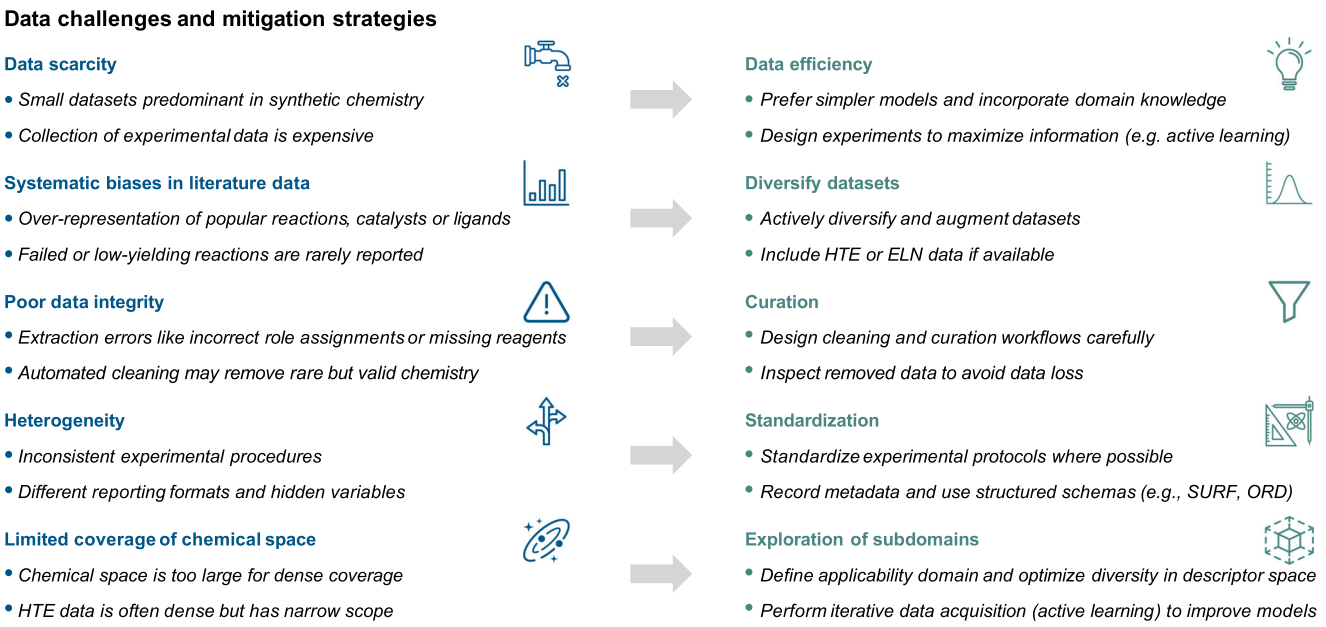


FIGURE 1 | Key data challenges in ML for synthetic chemistry and practical strategies to address them.

2.3 | The Critical Role of Reliable Data

ML models reflect the data they are trained on. They learn patterns, including systematic errors, artifacts, and biases present in the training set. As a result, data quality strongly determines predictive performance. This dependence is often described as the "garbage in, garbage out" principle. If the training data is flawed, the predictions will be flawed as well.

Considering the size of available chemical datasets, recent scaling analyses suggest that most are far below the regime where the achievable performance of large models saturates. Gadepally, Coley, and coworkers investigated the behavior of chemical language models for generative modeling and graph neural networks for learned interatomic potentials as model and dataset sizes were varied over many orders of magnitude (up to ten million datapoints) [65]. Performance improved consistently as the number of training examples increased, with no clear sign of saturation in the explored range. Consequently, many studies that rely on a few thousand molecules or reactions are likely operating in a data-limited regime. At the same time, it is important to note that most experimental chemistry is conducted in such a data-limited domain, where datasets are constrained in scope, diversity, and consistency due to the time and cost required to run the experiments. In this setting, simpler models are often advantageous because they are less prone to overfitting, while performance depends more on data quality than on architectural complexity, with individual data points and experimental noise having a disproportionate impact [46].

The requirements for data quality and quantity resulting from this double bind pose fundamental challenges for synthetic chemistry (Figure 1). The main sources of reaction and molecular property data are experiments that, in many cases, are not exhaustively controlled or standardized [66]. Additionally, the vast majority of historic chemical data is not only affected by varying experimental practices, but also by systematic biases as previously formalized by our group [67], as well as the groups of Grzybowski, Lee and Norquist [68–70]. Specifically, published data across different domains often reflects conscious and subconscious human choices. Literature-derived reaction datasets over-represent popular reagents, conditions, and transformations. They often follow power-law distributions in which a small fraction of substrates, ligands or catalysts dominates most reported reactions. Models trained on such data inherit these preferences and struggle to generalize beyond well-established chemistry. Furthermore, literature datasets are strongly skewed toward successful outcomes. Failed or low-yielding experiments are rarely reported, yet they are essential for defining decision boundaries and calibrating uncertainty. When negative data is absent, models become overconfident and conservative in exploration. Several reaction-yield studies show that adding even modest amounts of failed reactions to reduce this reporting bias markedly improves predictive performance and generalization [67, 69, 71].

In contrast to these systematic influences, experimental errors and noise are not dominant limitations. ML models are generally robust against or even benefit from low to moderate levels of random noise [72]. A more serious issue is data integrity. Large reaction datasets contain substantial fractions of missing atoms, incorrect stereochemistry, or misassigned reagents. Therefore, automated cleaning procedures are essential to ensure data quality [73, 74] but can, at the same time, exacerbate this problem by removing rare but chemically valid transformations, confusing novelty with error [67].

Taken together, these systematic data biases have led to limited success when training models on data extracted from databases such as SciFinder or Reaxys. Grzybowski and coworkers reported that feed-forward networks, graph neural networks, and transformer-based models all failed to meaningfully outperform a simple popularity baseline for predicting yields and reaction conditions in Suzuki–Miyaura couplings. This trend

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

was independent of the choice of molecular representation for a diverse and carefully curated literature dataset of over 16 000 reactions [70]. In a related study, Vuilleumier and coworkers manually collected a literature dataset of around 2200 Ni-catalyzed C—C and C—N couplings following C—O bond activation [75]. Although reporting bias could be partially mitigated by considering optimization tables, the predictive performance of even the best-performing model, an RF with DFT descriptors, remained modest, with an coefficient of determination (R^2) between −0.30 and 0.66 across different data splits. More strikingly, predictive performance collapsed when data from entire papers were held out as test sets, underscoring the difficulty of out-of-sample predictions and the need for careful evaluation of model generalizability (see Section 2.4).

Another widely used source of reaction data is the USPTO corpus, which contains around 1.8 million reactions automatically extracted from United States patents [76]. Differently filtered and cleaned versions of the dataset exist with the most commonly used subsets USPTO_MIT [77] (stereochemical information removed) and USPTO_STEREO [78] (stereochemical information retained) covering around 480 000 and 1 million reactions, respectively. Its scale and open accessibility made it a central source of training data for reaction prediction and retrosynthesis models [16, 43]. At the same time, USPTO data inherits many limitations of literature-derived datasets, including extraction errors, incomplete reaction information, high variability in reported yields and strong biases toward popular and robust transformations.

Consequently, models trained on literature data have shown only limited success in predicting reaction yield or selectivity (see Section 3.2). In contrast, high-throughput experimentation (HTE) has become a more reliable source of homogeneous reaction data in organic synthesis. Following the pioneering work of Doyle, Dreher, and coworkers on Buchwald–Hartwig couplings [79], the number of HTE datasets used for ML has grown rapidly [80–84]. Using HTE data often improves model performance because yield and selectivity distributions better reflect the explored chemical space and include many low-yielding reactions [67]. Furthermore, experimental procedures and analysis workflows are highly standardized, which reduces uncontrolled variability. At the same time, most HTE studies follow combinatorial designs with limited substrate and condition diversity. For datasets with only a few thousand reactions, this design bias can cause models to learn patterns tied to the experimental layout rather than chemically meaningful features (see Section 3.2 for a more detailed discussion) [85, 86]. Careful evaluation of molecular representations and rigorous out-of-sample testing are therefore essential [87]. Moreover, in many HTE workflows, yields are estimated from LC–UV peak area percentages or LC–MS signal intensity relative to an internal standard rather than accurately determined from calibration curves. These values report assay response, not absolute yield, and can introduce systematic deviations. General and accurate calibration-free quantification methods are therefore highly desirable to further improve data quality [88, 89].

Beyond literature and HTE datasets, electronic laboratory notebooks (ELNs) have recently emerged as another source of reaction data for quantitative reactivity prediction. They occupy an intermediate position between literature datasets and dedicated HTE campaigns, as ELNs record reactions at the time of execution and often include detailed conditions and failed experiments. Yet, entries are rarely standardized across projects. Inconsistent formatting and missing condition data can limit model performance without careful curation. Industrial ELNs better reflect the broad exploratory space of medicinal chemistry and cover many reaction classes. Their diversity exceeds that of most single-transformation HTE datasets, although common reactions such as amide couplings and cross-couplings remain overrepresented. To probe how the experimental data source affects model performance, Wiest and coworkers trained yield-prediction models on two datasets of Buchwald–Hartwig couplings: 3960 reactions generated by high-throughput experimentation and 1000 reactions extracted from AstraZeneca ELNs [90]. Whereas the HTE dataset contained all possible combinations of the considered ligands, bases, additives and substrates, the ELN data sampled only a small fraction of a much larger chemical space spanning roughly 470 million possible combinations of reactants, ligands, bases, and solvents. An RF model using simple physicochemical and structural features derived with RDKit achieved strong random-split performance on the HTE data (R^2 = 0.913) but showed limited predictive power on the ELN set (R^2 = 0.266). These results indicate that sparsity and heterogeneity hinder yield regression on ELN data. However, focusing on out-of-sample performance, ELN-trained models may generalize across broader a substrate space and are often more useful for classification tasks, such as identifying likely failures or ranking conditions [91].

Building on lessons learned from training models across different data regimes, the groups of Coley and Denmark have formulated best practices for the design and acquisition of new datasets [92, 93]. Their considerations can be generally applied to smaller datasets collected in batch as well as large HTE campaigns. Rational dataset design for ML models of chemical reactivity requires the explicit definition of the intended domain of applicability, including which substrates, catalysts, and conditions the model should generalize to. Dataset size alone does not compensate for poor coverage of this space, as redundant or sparsely distributed sampling limits extrapolative performance. Diversity should therefore be defined in the chosen molecular or reaction representation space. In this regard, descriptor-based or fingerprint-based clustering provides a more effective strategy for selecting informative experiments than intuition or exhaustive combinatorics. Experimental variables unrelated to the chemical hypothesis, such as inconsistent procedures, analytical artifacts, or unrecorded condition changes, should be controlled or standardized to avoid confounding the predicted variable.

A further consideration applies across all the data sources discussed above: they capture the performance of chemical reactions almost exclusively through the end-point yield. Other dimensions, such as reaction kinetics, are arguably more diagnostic of the underlying chemistry. These dimensions are, however, also more difficult to capture experimentally. High-throughput methods have only recently begun to make them accessible at a comparable scale (see Section 4.4) [94–98].

Last, the limited accessibility and quality of historical data underscore the need for common ontologies and open data repositories

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 | Datasets repeatedly referenced throughout this review. For each dataset, we list its reaction composition, the predicted target property, the number of reactions, and the original reference.

Entry	Dataset	Composition	Prediction Target	Size	Reference(s)
1	Doyle–Dreher	Buchwald–Hartwig couplings for all combinations of 15 aryl halides, 4 Buchwald-type ligands, 3 bases, and 22 isoxazole additives. Reactions were performed via HTE on a 0.1 mmol scale.	Reaction yield (UV peak area relative to an internal standard)	3955	[79]
2	Denmark	Thiol additions to *N*-acyl imines catalyzed by BINOL-derived chiral phosphoric acids, covering pairwise combinations of 5 imines and 5 thiols (25 products) with 43 catalysts. Reactions were performed on a 0.1 mmol scale.	Enantioselectivity ($\Delta\Delta G^{\ddagger}$, kcal/mol)	1075	[102]
3	Richardson–Sach	Suzuki–Miyaura couplings screened in an automated nanomole-scale flow platform, across a matrix of 11 ligands (+1 blank), 7 bases (+1 blank), and 4 solvents for combinations of quinoline electrophiles and indazole boron nucleophiles. Reactions run at ∼0.4 μmol scale.	Reaction yield (UV area%)	5760	[103]
4	NiCOlit	Nickel-catalyzed C—O bond activation cross-couplings (forming C—C and C—N bonds) manually extracted from 45 published papers, with diverse coupling partners, substrates, ligands, bases, additives, and solvents reported with full experimental conditions. Includes both scope and optimization data.	Reaction yield (different quantification methods)	2003	[75]
5	Pistachio	Reactions automatically text-mined from US, EU, and World patent office data, including paragraph text, reaction classification, and component roles.	—	∼22.5 million[a]	[104]

[a]Version-dependent.

for synthetic chemistry. Data standards such as Simple User-Friendly Reaction Format (SURF) [99] or the Open Reaction Database (ORD) Schema [100], together with repositories like the ORD [100] and the Chemotion repository [101] are early initiatives that support the publication of FAIR reaction data [27]. While improvements in the user-friendliness and accessibility of these platforms will further increase their uptake, a culture shift in the research community is necessary to normalize and reward the sharing of both positive and negative experimental results. In this context, publishers and funding agencies must accept responsibility in a process that will require changes in publication and evaluation practices.

Several datasets recur throughout this review. We summarize them in Table 1 and refer to it whenever one of them is referenced, to avoid redundancy and make it easier to follow links between studies that share a dataset.

2.4 | Critically Evaluating ML Models Beyond Performance Metrics

Before turning to evaluation, we briefly define a model's success and failure, as both terms are used in distinct senses throughout this review. In a technical sense, a model fails when it cannot meaningfully process an input at all because the compound lies outside its applicability domain (see below), for example, due to unsupported atoms, bonds, molecular sizes or a mismatched molecular representation. In a second, more common sense, a model fails when it returns an output that is simply too inaccurate or unreliable to benefit the user in a task. A successful model is accordingly one whose predictions are applicable and reliable enough for the task at hand. Success is thus defined relative to a concrete decision and an acceptable error tolerance, not in absolute terms.

Most evaluations of ML models in synthetic chemistry report global performance metrics to assess and judge the quality of predictive models. These metrics provide a convenient summary of predictive performance and enable head-to-head comparisons of algorithms and model configurations. However, global metrics alone do not capture how a model performs across different substrate classes or reactivity regimes, nor do they indicate whether selected features reflect chemically meaningful relationships. The design of the training and test sets used to obtain evaluation results must be carefully considered while employing validation strategies that explicitly probe robustness and generalization.

2.4.1 | Choosing the Right Evaluation Metric

The choice of evaluation metrics should reflect the underlying task and data structure. For common regression problems in

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

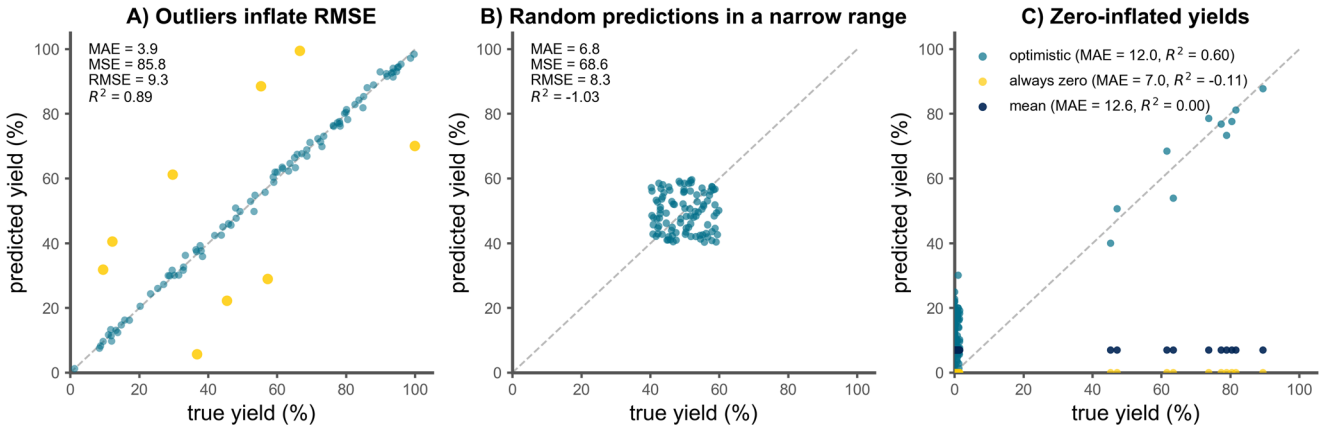


FIGURE 2 | Complementary regression metrics expose different model failure modes. Predicted versus true yield in each panel; the dashed line marks perfect agreement. (A) A few outliers leave MAE low but sharply inflate MSE and RMSE. (B) Random predictions in a narrow range (40%–60%) give a small MAE yet a negative R^2, revealing no explained variance. (C) On a zero-inflated dataset, always predicting 0% yields a lower MAE than an optimistic model that captures high yields, though the optimistic model achieves the higher R^2.

synthesis, such as yield or selectivity prediction, both mean absolute errors (MAE) and (root) mean squared errors (RMSE/MSE) should be reported. As RMSE squares each error before averaging, large deviations contribute disproportionately to its value. Consequently, MAE captures the error of a representative prediction, whereas RMSE punishes large errors and outliers. R^2 indicates the explained variance, but should be interpreted with caution, as its value can be over-optimistic when the data's dynamic range is large, and it may not be valid for true nonlinear models. Some possible pitfalls and trade-offs of using these popular metrics are illustrated in Figure 2. Outliers inflate RMSE rapidly, even when a model performs well on the bulk of the data (Figure 2A). A model making random predictions on a dataset with yields between 40% and 60% can still achieve a small MAE, but its limitations surface in a low or even negative R^2 (Figure 2B). The reverse case is equally instructive. A model that always predicts 0% yield on a zero-inflated dataset can reach a lower MAE than a model that captures higher yields well but is overly optimistic about zero-yielding examples (Figure 2C). The second model nonetheless achieves the higher R^2. These contrasting failure modes show why no single metric suffices. Reporting several complementary metrics, chosen to reflect the structure of the dataset and the decision the model is meant to support, guards against the misleading verdict that any one metric can deliver in isolation.

If the ranking of candidates is more relevant than absolute values, a rank-based metric such as the Spearman correlation coefficient or Kendall's tau can be more informative. For classification tasks, an accuracy metric is only meaningful for balanced datasets, which are rarely encountered in chemistry. In the more common case of imbalanced classes, metrics such as Matthews correlation coefficient (MCC), balanced accuracy, F1 score, or precision-recall area-under-curve (PR-AUC) provide a more realistic assessment. In all settings, model performance should not be judged by a single metric. A small set of complementary metrics, discussed in the context of the prediction problem and prospective applications, provides a more robust picture of predictive utility [105].

2.4.2 | Applicability Domain of ML Models

During model development and evaluation, performance is assessed on predefined training, validation and test splits. After deployment, validation becomes far more costly as predictions for new compounds or reactions require experimental confirmation. The value of a model, therefore, lies in its ability to prioritize the most promising experiments. This value disappears when the model is applied to examples outside the chemical space covered by the training data, where its performance can deteriorate sharply. Thus, the applicability, reliability and decidability of a model should be evaluated for the intended use case and chemical space (Figure 3) [106].

Applicability describes whether an input is compatible with a model and depends on the model architecture and the training data. The compound class must be compatible with the chosen featurization or embedding. For example, models developed for small organic molecules may fail for metal complexes or polymers due to unsupported bond types or molecular size limits. Here, "failure" refers to the technical sense defined above: the input is incompatible with the model rather than merely predicted inaccurately. In addition, input descriptors should lie within the range covered by the training data (Figure 3A), or close structural analogs should be present for sequence- or graph-based models. If extrapolation beyond the training domain is intended, the model's ability to explore new chemical space must be evaluated carefully as outlined below [106].

Reliability reflects whether sufficient and relevant data support a prediction. Uncertainty associated with low reliability stems from insufficient coverage in sparse regions of the feature or latent space (Figure 3B). Evaluating model performance via cross-validation with nearest neighbors can help estimate prediction reliability [106].

Decidability describes the uncertainty associated with a specific prediction. It indicates how much confidence can be placed in the predicted value once applicability and reliability are satisfied.

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

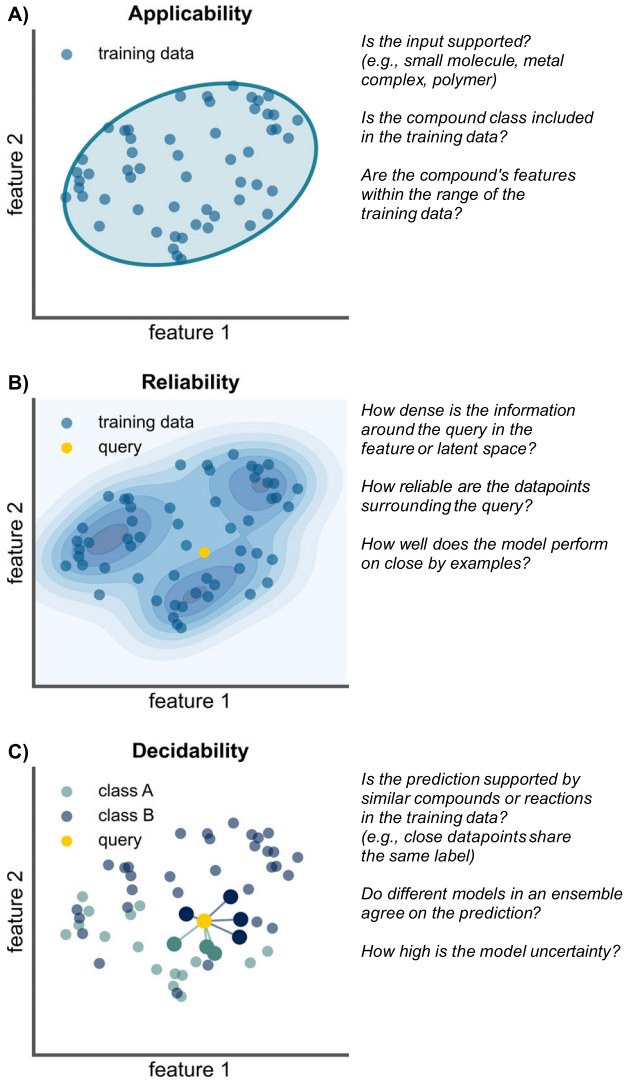


FIGURE 3 | Assessing a model's applicability domain in chemical space. (A) Applicability indicates whether a query lies within the model domain defined by the model-compatible input and training data. (B) Reliability reflects the extent of local data support in the feature space and can be estimated by the data density around a query. (C) Decidability captures confidence in the predicted outcome. It can be estimated from model ensembles, from calibrated confidence scores, or from predicted class probabilities.

This uncertainty can be estimated from agreement within model ensembles, from calibrated confidence scores, or from predicted class probabilities (Figure 3C).

The quantification of model uncertainty itself is an active area of ML research whose detailed discussion is outside the scope of this work. In the context of in silico chemistry, this topic has been reviewed recently by Jacob, Kirchner, and coworkers [107]. A key distinction for understanding and reducing model uncertainty, which we highlight for its relevance to experimental science, is the distinction between epistemic and aleatoric uncertainty. Epistemic (or systematic) uncertainty arises from a lack of knowledge. In the context of chemistry, it can arise from limited training data that excludes key compound classes or from a featurization that fails to capture important characteristics, such as a ligand's conformational flexibility. Aleatoric (or statistical) uncertainty, on the other hand, stems from the variability in the outcome of an experiment due to inherent random effects like detector noise. While both types play a role in almost every predictive task, their estimation and differentiation are key for reducing model uncertainty. Epistemic uncertainty can typically be reduced (and therefore identified) by collecting more representative data, feature engineering or changing the ML algorithm or architecture. Aleatoric uncertainty, on the other hand, is a characteristic of the data acquisition process that cannot be reduced by improving featurization or optimizing the model. Distinguishing between both types, therefore, helps determine whether improving predictions requires better data and models or more precise experiments (e.g., more accurate measurements and repeated observations) [108, 109].

In the context of reaction prediction, the importance of reliability in asymmetric catalysis has recently been examined by Reid and coworkers using eight selectivity datasets of varying size (240 to 12 619 reactions) spanning organo- (chiral phosphoric acids, imidodiphosphorimidates, thioureas, and secondary amines) and metal catalysis (palladium, rhodium and magnesium catalysts) [110]. Both homogeneous datasets, where the reactant transformations are consistent, and heterogeneous datasets featuring multiple reaction types were considered. The authors used these case studies to quantify the extent to which predictive accuracy depends on local similarity in reaction space. They represented reactions as concatenated extended connectivity fingerprints (ECFP4) [37] and constructed reduced training sets that contained only reactions within a defined distance of the query. A RF regressor was then trained on these restricted subsets and used to predict selectivity as the difference in Gibbs free energy of activation ($\Delta\Delta G^{\ddagger}$). The analysis showed that most of the predictive power arises from nearest and next-nearest neighbors. Training only on structurally similar reactions achieved nearly the same accuracy as using the full dataset. Conversely, adding chemically diverse but distant reactions improved out-of-sample predictions only marginally, by approximately 0.25–0.5 kcal/mol, when the target reaction lacked close analogs in the training set. These results demonstrate that, for many selectivity models, performance is largely governed by local data density rather than global coverage.

2.4.3 | Evaluation and Adversarial Controls

ML models readily exploit confounding variables, experimental design artifacts, and other spurious correlations instead of chemically meaningful relationships. Adversarial controls are therefore necessary to avoid overoptimistic performance estimates and poor generalization. Guidelines for such controls, designed to test model interpretability, uncover hidden correlations, and investigate the dependence of performance on key features, have been proposed by Keiser, Walsh as well as Rodrigues and coworkers [34, 105, 111].

Logical and interpretability checks help reveal discrepancies with qualitative chemical understanding. Large errors may point to data quality issues or systematic bias, while feature importances and attributions should align with chemically relevant trends. To test for confounders and evaluate model complexity, alternative

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

models should be trained. If comparable accuracy is achieved with metadata or layout variables, the main model may be "right for the wrong reasons" and not rely on chemically relevant features. Models should further be compared against simple baselines such as predicting the mean or the most frequent class. When using more sophisticated models like deep learning architectures, comparison to less complex models like RFs or SVMs can simplify training and evaluate whether performance is data-limited [34, 111].

Moreover, deliberately replacing key molecular descriptors by a one-hot encoding (OHE) or randomizing features helps to investigate whether a model merely learns the structure of the dataset or the identity of compounds. Note that this simple control has proven to be particularly important in case of combinatorial datasets, such as HTE-derived reaction datasets (see Sections 2.3 and 3.2) [85, 86]. Randomizing molecule–target pairs (also known as y-scrambling) can further help to detect data leakage if model performance does not drop close to chance [34, 111].

To evaluate if an ML model can truly generalize and extrapolate beyond the compounds present in the training dataset, out-of-sample tests are essential. Random splits mainly probe interpolation and can overestimate performance when similar scaffolds, substrates, or condition combinations appear in both training and test data. In contrast, scaffold-based, class-based, or time-split validation better mimics prospective application and provides a more realistic estimate of extrapolative power. True out-of-sample evaluation, however, can be challenging for molecular and reaction datasets. Molecular similarity depends on the chosen representation and on the question being asked [112]. Scaffolds and reaction components are often unevenly distributed, and excluding different elements such as substrates, catalysts, or coupling partners can lead to markedly different performance. For this reason, multiple out-of-sample strategies are typically required to obtain a reliable picture of a model's generalization behavior [34, 105].

2.4.4 | Experimental Validation Under Practical Constraints

Experimental evaluation of model predictions requires significant time and resources. It is therefore rarely feasible to test all proposed reactions or compounds, and some degree of selection is unavoidable. This makes the sampling strategy central to a meaningful assessment of predictive performance. Restricting validation to chemically intuitive or experimentally straightforward predictions might primarily confirm the preformed hypothesis about model performance. A rigorous evaluation of reaction-level models should therefore include structurally diverse substrates, new catalyst classes, or condition sets obtained through unbiased and carefully motivated sampling strategies. Incorporating a subset of lower-confidence predictions can further delineate the model's limits.

3 | Predictive Molecular Modeling

In this section, we aim to give an overview of successful applications of ML models to predict molecular properties, chemical reaction outcomes, reaction conditions or to guide catalyst design. While we are not providing an extensive summary of all previous studies in these areas, our goal is to highlight possible applicability domains, to convey different concepts for integrating ML models into chemical research and to highlight the opportunities as well as common pitfalls and remaining challenges.

3.1 | Prediction of Molecular Properties Relevant to Synthetic Chemistry

The prediction of molecular properties in silico is one of the most common tasks of ML applications in chemistry. After QM methods have dominated this area for a long time, ML models now offer faster, cheaper, and easier-to-access alternatives for quickly estimating molecular properties. Before highlighting several successful examples of property prediction models in synthetic chemistry, we will first give an overview of selected user-friendly frameworks that enable synthetic chemists to rapidly implement their own ML property prediction pipelines.

3.1.1 | Training Your Own Property Prediction Model is Not Hard (Anymore)

While just a few years ago, training predictive models required deep expertise and familiarity with specialized ML software repositories, high-level chemistry-oriented frameworks nowadays enable anyone with basic Python skills to train models with minimal effort. However, even when the training process itself is highly streamlined and well-supported, the user must still carefully navigate data pitfalls and ensure meaningful evaluation (see Sections 2.3 and 2.4).

In the following, we will highlight some frequently used, well-maintained, well-documented and beginner-friendly software packages (Table 2). Notably, the choice of one of these high-level (or underlying low-level) frameworks depends on the scale of the available data, the specific molecular representation required, and the desired balance between customization and "out-of-the-box" performance. For those practitioners seeking a comprehensive, unified ecosystem, ChemML and DeepChem serve as early and longstanding "all-in-one" platforms that wrap a vast array of chemical ML tasks into a single application programming interface (API) [113, 114]. As a more high-performing yet practical end-to-end alternative, Chemprop provides facile access to directed message-passing neural networks (D-MPNNs) for learning representations and predicting properties directly from molecular structures [12, 115]. If raw computational speed is the priority, fastprop offers an agile alternative that operates on cheap molecular descriptors rather than molecular graphs [116]. In addition to these frameworks allowing model training from scratch, pretrained foundation models (see Section 2.2) such as ChemBERTa [117], Chemeleon [63], or Molformer [64] allow users to leverage patterns already distilled from millions of molecules to jumpstart their own predictions. Moreover, in very-low-data regimes, frameworks like ROBERT not only handle model training but can also calculate QM-descriptors as input features on the fly [53, 54].

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 | Overview of selected, chemistry-specific high-level software frameworks for molecular prediction tasks. In addition, lower-level, general Python frameworks can always be used for prediction tasks, however, requiring deeper ML knowledge.[a]

Framework	Key features	References
ChemML	Early end-to-end framework featuring a variety of modular tools for data mining, molecular featurization and property prediction.	[113]
DeepChem	Swiss Army knife that provides standardized molecular featurizers, dataset splitters, and a massive library of different, pre-implemented model architectures.	[114]
Chemprop (v2)	High "out-of-the-box" performance graph-based model that utilizes directed message-passing neural networks (D-MPNN) to learn property-specific features directly from molecular structures.	[12, 115]
Fastprop	Lightweight framework focusing on speed by relying on Mordred descriptors with simple but optimized feed-forward neural networks.	[116]
MolBERT, ChemBERTa, and Molformer	Large-scale "chemical language" foundation models that used self-supervised transformer architectures and learned universal molecular representations from billions of SMILES strings.	[64, 117, 118]
Chemeleon	D-MPNN-based foundation model, pretrained to reconstruct thousands of classical descriptors, enables downstream prediction tasks directly from molecular graphs.	[63]
ROBERT	End-to-end framework especially designed for low-data scenarios, enables ML training, and generation of QM-descriptors on the fly.	[53, 54]

[a]Most common low-level libraries for ML and deep learning include scikit-learn [119], TensorFlow [120], PyTorch [121], PyTorch Geometric [122], JAX [123], and many more.

Besides the pure ML backend, most of the presented frameworks also include several different featurizers and data splitters. Therefore, testing the influence of the molecular representation and probing the generalizability of trained ML models, for example, by employing out-of-sample splits, should always be considered in addition to standard random splitting (see Section 2.4).

3.1.2 | Prediction of Reactivity-Related Properties

In synthetic chemistry, relevant properties can broadly be divided into intrinsic molecular properties associated with reactivity (Figure 4A), and experimental observables, for example, properties that enable the characterization or analytical detection of compounds (Figure 4B). Within the first category, reactivity-related properties are traditionally derived from QM methods and can help chemists to estimate or rationalize reactivity. However, obtaining them requires extensive computational effort and resources, rendering ML a cheaper and faster alternative [124].

Focusing on radical reactivity, Paton and coworkers, among others [125, 126], addressed the prediction of homolytic bond dissociation enthalpies (BDEs), which are a key quantity for estimating the thermodynamic feasibility of elementary steps in radical mechanisms. For ML model training, they compiled a dataset of ~290 000 DFT-calculated BDEs, spanning small neutral organic molecules containing C, N, O, S, P atoms as well as halogens and hydrogen. Notably, their custom GNN model ALFABET simultaneously predicts all BDEs for a given molecular graph, achieving an MAE of approx. 0.6 kcal/mol versus DFT. Notably, this accuracy enabled the accurate prediction of H-atom abstraction sites or fragmentation pathways in real reaction case studies. Furthermore, a user-friendly web app allows experimental chemists without programming skills to rapidly estimate all BDEs in a provided molecular structure within milliseconds, without requiring any additional input or QM calculation [127, 128].

Focusing on photochemistry, our group recently presented the EnTdecker platform for the identification of novel substrates in energy transfer (EnT) photocatalysis. To ensure applicability across a broad chemical space, the training data spanned ~30 000 small organic molecules with up to 25 heavy atoms, all randomly sampled from the large ZINC database. For this library, adiabatic triplet energies and corresponding spin distributions, both representing key quantities for estimating triplet reactivity, were calculated by DFT and used as targets for ML model training. Notably, GNNs trained on molecular graphs using the Chemprop library achieved the best predictive MAE of approx. 2 kcal/mol for the adiabatic triplet energy. In addition, a transformer model trained on SMILES and binned Mulliken spin values predicted the correct sites with high spin densities. Besides random- and out-of-sample splits, the models were also validated on previously published literature and experimental luminescence quenching data [10]. Combined with an easily accessible web interface, the platform enables chemists to rapidly estimate a molecule's reactivity in triplet-energy-transfer photocatalysis. Moreover, in a subsequent study, the EnTdecker platform was used for high-throughput virtual screening to identify novel substrates for EnT-catalyzed chemical reactions (see Section 4.1) [129].

Furthermore, considerable effort has focused on training ML models to predict sets of DFT-derived molecular descriptors rather than application-specific properties, as discussed above. When trained for specific compound classes, these models enable the rapid in silico estimation of physically meaningful DFT-based

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

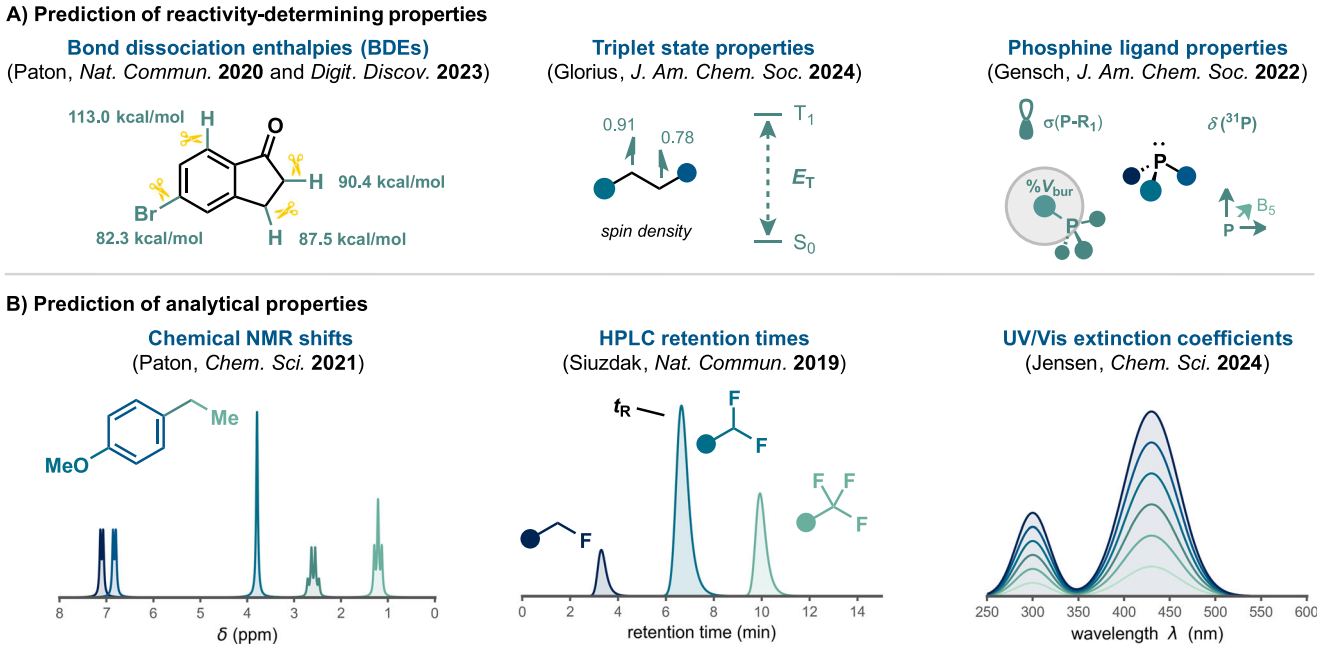


FIGURE 4 | Examples of successfully adressed property prediction problems relevant to synthetic chemistry. (A) Prediction of directly reactivity-related properties such as bond dissociation enthalpies, triplet state properties, or phosphine ligand properties. (B) Prediction of analytical properties assisting the identification and quantification of compounds in reaction assays (e.g., NMR shifts, HPLC retention times, and extinction coefficients).

descriptors [130–132]. A notable example by Gensch, Sigman, Aspuru-Guzik, and coworkers is the kraken virtual database of monophosphine ligands. In this work, an array of comprehensive QM-descriptors relevant for transition metal catalysis was calculated for a training set containing 1558 unique phosphine ligands in their free and metal-ligated state. Specifically, the chosen descriptors span electronic parameters such as Fukui functions, ^{31}P-NMR tensors, and frontier molecular orbitals, as well as steric parameters such as cone angles, buried volumes or pyramidalization extents. A variety of different ML models (RF, GB, GNN, and GP) were all successfully trained on molecular fingerprints or graphs to predict these descriptors for an extended virtual library of 300 000 new phosphines. Such a library subsequently enables virtual screening of a broad ligand space with applications in reaction development and catalyst design (see Section 3.4) [133]. Notably, the on-the-fly prediction of QM-descriptors for downstream applications has also been increasingly incorporated into end-to-end frameworks such as ASKCOS [11, 58] or ROBERT [53].

3.1.3 | Prediction of Experimental Properties in Analytical Chemistry

Moving from intrinsic molecular properties to experimental observables, another common use case for property prediction models is the prediction of molecular characteristics relevant to analytical chemistry (Figure 4B). The estimation of these properties is typically motivated by the absence of authentic standard samples required for the identification or quantification of analytes. Nowadays, ML models have been trained for most analytical techniques relevant to synthetic chemistry, helping chemists to rapidly obtain in silico data that can be compared to their own, experimental results.

At the core of a chemist's toolbox, several studies have addressed the accurate prediction of ^{1}H, ^{13}C, and, to a smaller extent, also ^{19}F or 2D spectra based on a molecular structure as an input [134–136]. As a state-of-the-art example, the CASCADE platform and its 2.0 successor provide ^{13}C and ^{1}H chemical shifts based on the molecular structure as input. While their first publication leveraged transfer learning from a large dataset of ~200 000 QM-simulated NMR shifts to experimental chemical shifts of only ~5000 molecules, the 2.0 follow-up is exclusively based on experimental chemical shifts of ~20 000 molecules, including also less frequent main group elements such as B, Si, Se, Br, and I. In both studies, 3D graphs were chosen as molecular representations that facilitate capturing stereochemical and conformational influences. A GNN trained on these 3D graphs finally achieves practically useful chemical shift MAEs of 0.73 ppm for ^{13}C shifts and 0.16 ppm for ^{1}H shifts. To enhance transparency, the model incorporates uncertainty-aware metrics using deep-kernel learning, providing confidence estimates that correlate strongly with predictive error. For further validation, the model was applied in challenging structure elucidation tasks, such as identifying the correct regioisomers or revising historical misassignments in literature [137, 138]. More recently, with the release of larger experimental NMR datasets [139], the performance of new ML models is expected to increase while simultaneously broadening the covered chemical space [140, 141].

For the analysis of complex mixtures rather than pure compounds, high-performance liquid chromatography (HPLC) is typically employed for the separation of compounds, but the identification often represents a challenge in case of non-ionizable analytes, isobaric interferences, or even in the absence of MS detectors. Therefore, the prediction of retention times (t_R) for a given chromatographic setup has been addressed in multiple recent studies [142–146]. In the context of asymmetric catalysis

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

or synthesis, this task is of particular interest for in silico estimation of chromatographic enantioseparation. As a representative example, Mo and coworkers compiled a dataset of 25 847 chiral molecules curated from 644 asymmetric-catalysis papers and trained a GNN on 3D molecular graphs to predict enantiomer retention times (R^2 = 0.70 and MAE = 3.40 min). By encoding chirality in the 3D graph representation, the model distinguishes enantiomers, and, despite the relatively high MAE in t_R, a separation probability can still be derived for given conditions [147]. Such models, however, both in achiral and chiral chromatography, are typically trained for a defined set of conditions, and generalization across broader chromatographic systems with different stationary phases, gradients and temperatures remains an area of ongoing research [148, 149].

In addition to identifying unknown analytes, quantifying them remains a central challenge for chromatographic analytical techniques. While relative peak areas of various detector signals, such as UV/vis absorption or charged aerosol detectors (CAD), are often used as indirect measures of quantity, authentic product samples are usually required for obtaining accurate, calibrated quantities. Along this line, Jensen and coworkers trained an ML model aiming to predict molar extinction coefficients ε of small organic molecules. With about 40 000 datapoints derived from a proprietary dataset, their D-MPNN model trained using Chemprop achieved an MAE of ca. 0.2 on log(ε). In an experimental evaluation, this translates to an acceptable MAE of 8.3% for yield estimation when using the predicted ε values to calculate yields from HPLC–UV/vis analyses of chemical reactions [88].

Last, the prediction of entire spectra of all kinds (UV/vis, IR, EPR, EI–MS, ESI–MS/MS, and …) for direct comparison with experimentally obtained results represents another common application for ML models. Given the vast scope of these methods, also beyond synthetic chemistry, we refer the reader to separate reviews on this topic [150–154].

3.2 | Reaction Performance Prediction—Yield and Selectivity

Chemical reactions lie at the heart of synthetic chemistry, enabling chemists to create new molecules with desired structures and properties. Over the past years, extensive efforts have focused on predicting chemical reaction outcomes. Here, we focus on the specific challenges associated with predicting reaction yields and selectivities. Notably, such models are not only relevant for evaluating reactions directly but also serve as enabling tools for related tasks, including condition selection (Section 3.3), catalyst design (Section 3.4), reaction optimization (Section 4.2), scope modeling (Section 4.3), and mechanistic analyses (Section 4.4).

3.2.1 | Product Yields

Yield prediction is particularly important in organic synthesis because expected yields influence library design and construction in exploratory medicinal chemistry and steer route selection in multistep pharmaceutical development. In academic settings, yield remains the primary, and often singular, metric used to optimize synthetic methods (see Section 4.2).

Predicting yields is inherently challenging because many interdependent variables influence the outcome. In an idealized view, the thermodynamics and kinetics of an presumed reaction are the exclusive factors that determine reaction yields, and ML models can capture these effects when clear (quantitative) structure–reactivity relationships exist. In practice, however, real experimental systems extend well beyond the simplified model of a single, exclusively progressing reaction. The overall reaction outcome can be affected by competing pathways, solubility and aggregation phenomena, product decomposition, and catalyst deactivation, all of which contribute additional thermodynamic and kinetic complexities that are absent from the idealized picture. As a result, the structures of reaction composition-yield hypersurfaces remain poorly understood [155]. While Grzybowski and coworkers have recently shown that yield distributions often vary smoothly with continuous variables such as concentration or temperature, abrupt product switchovers can occur due to equilibrium shifts or small structural changes [156]. This phenomenon, more broadly known as reactivity cliffs, introduces an additional level of complexity that many models struggle to capture, especially when dealing with categorical variables [157–159]. Capturing these dynamic effects, like product degradation or catalyst inhibition, from static descriptors alone is difficult because a single endpoint yield collapses the entire kinetic and thermodynamic history of the network into one number. Time-resolved data, such as concentration trajectories of substrates, intermediates and products over the course of a reaction, expose the competing pathways, decomposition rates and equilibrium shifts that a final yield obscures (see Section 4.4).

Further variability arises from experimental errors. Contamination of reagent batches, fluctuations in temperature or moisture, material losses during purification, and product decomposition can cause deviations for a series of experiments or between experimental runs. As discussed in Section 2.3, variability across independent studies can be even larger. This is mainly caused by systematic differences in experimental and analytical workflows, leading to results being reported as crude yield, isolated yield, or conversion. This heterogeneity further complicates the integration and alignment of datasets for model development [155].

Consequently, most successful yield-prediction approaches have utilized HTE data collected under standardized experimental and analytical protocols. However, analysis of HTE datasets and critical evaluation of the derived ML models have shown that HTE data quality and homogeneity are not the only factors benefiting model performance. Recently, Maryasin and coworkers showed that a simple statistical model, which only requires knowledge of which reaction components were combined in each well, can predict yields from HTE datasets and even capture reactivity trends between substrates [86]. For three HTE-derived datasets of Buchwald–Hartwig couplings (Table 1, entry 1), Suzuki–Miyaura couplings (Table 1, entry 3) and deoxyfluorinations [160] they treated each reactant as a label rather than encoding its molecular structure. The model assigns a fitted numerical contribution to

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

each individual reactant (catalyst, base, substrate, etc.), including interaction terms that account for how specific pairs of reactants impact reaction outcomes when combined. The additive effects of the reactant contributions and their interaction terms are then considered to make a prediction for a given reaction. Despite its simplicity and full interpretability, this model performed comparably to RFs using DFT features [79, 160] and a pretrained BERT model [161] on random splits across all three datasets. These results confirm that the identities of all reactants are sufficient to predict reactions within combinatorial HTE datasets, provided no generalization to unseen compounds is required. This aspect must be carefully considered when evaluating yield models that rely on HTE data (see Section 2.4).

Focusing on assessing a model's generalizability, evaluating out-of-sample performance is crucial to develop models that truly learn from chemically relevant features. Further claims of superiority of more complex ML models or complex featurization schemes need careful justification if simpler, interpretable models perform similarly. Isayev and coworkers benchmarked yield-prediction models on a literature dataset of 41 239 amide couplings curated from Reaxys [157]. They compared the models trained on this literature data to models trained on Doyle and Dreher's HTE dataset of Buchwald–Hartwig couplings (Table 1, entry 1). On the heterogeneous literature set, performance remained low across different featurization strategies and model classes (including linear models, RF, SVM, GNNs, and transformer-based models), and neither 3D nor QM-descriptors consistently improved accuracy. Error analysis pointed to reactivity cliffs and inconsistent yield reporting as key limitations, indicating that data quality and heterogeneity, not model complexity, are the main bottlenecks. On standardized HTE data, both RFs and SVMs performed only slightly worse than a pretrained BERT model, and notably outperformed GNNs.

Despite these general trends and data limitations, it should be noted that more tailored ML architectures have been shown to improve reaction yield prediction performance. Qi and coworkers developed RXNGraphormer, a deep learning model that represents each reaction as a molecular graph and learns individual structural features using GNNs, while a transformer module captures interactions between reaction components [162]. The authors pretrained the model on more than 13 million real and artificially perturbed reactions to recognize chemically meaningful bond changes, and then fine-tuned it to predict yields by adding a regression layer. RXNGraphormer's yield prediction accuracy was evaluated on two HTE datasets for Buchwald–Hartwig and Suzuki–Miyaura couplings, as well as a literature-derived dataset of nickel-catalyzed C—O couplings collected by Vuilleumier and coworkers (Table 1, entry 4). Because the model was pretrained on a large reaction corpus and utilizes information about bond changes during the reaction, it matched or outperformed established approaches in out-of-sample tests on the HTE benchmark datasets. With an average MAE of 21.76% and R^2 of 0.309, the performance stayed moderate when individual publications were used as a held-out test set but displayed a significant improvement over the values obtained using an RF with DFT descriptors in the original publication (MAE = 28.8% and $R^2 = -0.01$). Notably, the architecture also showed strong performance on related prediction tasks, such as enantioselectivity and forward reaction prediction.

Despite the promises of such pretraining or transfer learning approaches, recent studies agree that data amount, diversity, and quality of data are the main limitations for yield prediction performance [155, 157]. Addressing the issue of limited diversity, Denmark and coworkers improved yield prediction for Buchwald–Hartwig couplings by developing data acquisition strategies that enforce diverse reactant pairings and more uniform yield distributions [163]. The authors selected 50 amines and 50 (hetero)aryl bromides to span a broad range of electronic and steric properties, organized them into descriptor-based clusters, and sampled representative reactant subdomains. Because early data were dominated by zero yields, a deep feed-forward neural network classifier was trained to distinguish between zero and nonzero outcomes and to prioritize productive conditions, thereby shifting the dataset toward a more informative yield distribution while retaining low-yield cases. Neural networks trained on engineered structural descriptors calculated for this dataset were then deployed in an active-learning workflow to iteratively extend the model's applicability domain toward new reactant subspaces. By reducing a nominal 450 000-member reaction space to about 3300 strategically selected experiments, this systematic strategy enabled reliable recommendations of high-yielding conditions for out-of-sample couplings, achieving more than 85% reaction yield in 10 out of 13 test cases. This study reinforces the value of datasets that are designed for chemical diversity. Combined with active learning-based data acquisition, this strategy can help to minimize the number of experiments needed to train accurate predictive models.

The extent to which larger-scale data collection can improve yield prediction models has recently been evaluated by a team at molecule.one [164]. Most HTE datasets cover narrow condition screens and limited product space, which restricts generalization. In contrast, the molecule.one team reported a microliter-scale HTE campaign of about 200 000 reactions across amide, Suzuki–Miyaura, and Buchwald–Hartwig couplings, spanning more than 1000 substrates and roughly 30 000 products, designed to maximize substrate diversity while keeping conditions limited. Reaction outcome predictions were performed with a transformer-based model to predict reaction success, defined as yield above a low-yield threshold, as a binary outcome. For Suzuki–Miyaura couplings, performance followed smooth scaling laws over more than two orders of magnitude in training size, with no clear saturation. This observation is consistent with the scaling behavior reported by Coley and coworkers for generative chemistry models (see Section 2.3) [65]. The models generalized to unseen substrates and were transferable to an external amide HTE dataset, underscoring the importance of scale, diversity, and standardized data for robust outcome prediction.

In summary, these studies imply that progress toward out-of-sample yield prediction depends on the systematic generation of diverse, well-curated datasets. Future advances will likely arise from the combination of standardized large-scale data collection, thoughtful experimental design, and models that integrate chemically meaningful representations of reactions.

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3.2.2 | Enantioselectivity

The enantioselective synthesis of chiral molecules is particularly important for the development of safe and targeted drugs, thus rendering asymmetric catalysis a major research focus for decades. Despite its significance in these applications, developing asymmetric transformations remains challenging and largely empirical. While DFT has found widespread success in the post hoc rationalization of selectivity, computational a priori prediction of a reaction's enantioselectivity remains difficult [165]. This is not surprising as selectivity arises from very small free energy differences between competing transition states. In many catalytic systems, the underlying mechanisms are not fully established and can even vary with changes in substrate, ligand, or catalyst. In addition, selectivity is highly sensitive to conformational preferences and three-dimensional interactions, so accurate theoretical treatment of steric and non-covalent effects is essential yet challenging. In this context, ML methods can learn selectivity-determining patterns directly from data, even when the detailed mechanism is unclear or when multiple pathways contribute [166]. One hurdle in developing ML models for enantioselectivity prediction is the limited data available. Datasets for asymmetric transformations are typically small, often containing fewer than 50 datapoints collected under identical conditions as part of catalyst, ligand and scope explorations. This data sparsity has led to a focus on information-rich featurization strategies for individual reaction components or transition states [167]. At the same time, robust descriptors of catalyst–substrate interactions and the reliable transfer of models between related reaction classes remain persistent challenges.

In synthetic chemistry, enantioselectivity is typically reported as enantiomeric excess (*ee*), defined as the difference between the mole fractions of the two enantiomers in the product mixture. For modeling purposes, however, $\Delta\Delta G^{\ddagger}$ between the two diastereomeric transition states of the enantio-determining pathways is more appropriate and can be derived from *ee* values (Equation 1).

$$\Delta\Delta\, G^{\ddagger} = RT \ln\left(\frac{100 + ee}{100 - ee}\right) \tag{1}$$

In a Curtin–Hammett regime, $\Delta\Delta G^{\ddagger}$ is the direct cause of enantioselectivity, whereas *ee* is an experimental observable that reflects this underlying energy difference. Therefore, $\Delta\Delta G^{\ddagger}$ has a clear physical significance for modeling selectivities. Furthermore, the relationship between $\Delta\Delta G^{\ddagger}$ and *ee* is nonlinear (Equation 1 and Figure 5), reflected by the fact that a $\Delta\Delta G^{\ddagger}$ value of 1.0 kcal/mol leads to ~90% *ee*, but a further 1 kcal/mol is required to achieve ~99% *ee*. This nonlinear relationship generally renders *ee*-distributions more skewed, especially in the context of publication bias. $\Delta\Delta G^{\ddagger}$ further accounts for temperature, allowing the model to capture temperature effects on selectivity. Last, $\Delta\Delta G^{\ddagger}$ is unbounded, avoiding physically impossible extrapolations [166, 168].

When selecting descriptors and models for enantioselectivity prediction, two major strategies can be identified. One approach embeds detailed mechanistic insight, such as transition state geometries and related descriptors, directly into the model [169, 170]. The other relies on general molecular representations or large collections of descriptors that presumably have an influence and allows the algorithm to learn correlations with enantioselectivity with minimal prior assumptions about the enantioinduction pathway. In practice, a full mechanistic understanding is often unavailable, even for established reactions, and transition-state calculations and descriptor generation can be computationally demanding. Still, incorporating stronger physical priors into descriptor design (Figure 6) and model architecture has repeatedly improved extrapolation to unseen substrates and reduced the amount of data required for robust predictions [169, 171].

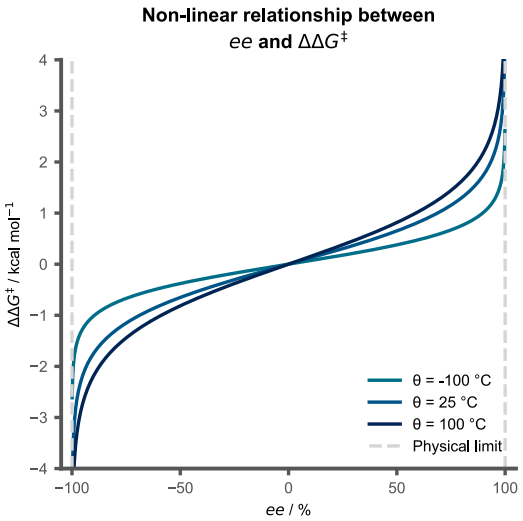


FIGURE 5 | Relationship between *ee* and $\Delta\Delta G^{\ddagger}$ at −100°C, 25°C, and 100°C. The dashed lines mark the physical limits for *ee*.

For both mechanism-agnostic and informed scenarios, MLR models based on QM-descriptors, as pioneered by Sigman and coworkers, have been successfully applied across various catalyst classes and transformations. The development of MLR models typically adhears to the following steps: (i) First, chemically meaningful physical-organic descriptors are selected for either the catalyst, the substrate, or both. These are usually steric and electronic parameters derived from DFT-optimized ground-state structures, which can be informed by a mechanistic hypothesis. If mechanistic understanding is limited, a broad range of descriptors can be selected initially. (ii) A representative dataset is assembled, ensuring structural diversity and a broad range of reaction outcomes, including both high- and low-performing examples, and divided into a training and test set. (iii) The descriptor values are normalized and analyzed for intercorrelation to remove redundant parameters and avoid multicollinearity, which can obscure interpretation. (iv) A multivariate model is built using least-squares regression with systematic feature selection, while limiting the number of descriptors relative to the dataset size to prevent overfitting. Ultimately, this workflow allows for the identification of meaningful descriptors even without mechanistic insights. Notably, the approach can be regarded as an extension of linear free energy relationships (LFERs), however, relying on more tailored descriptors and multivariate instead of univariate models [172].

The described MLR approach has been successfully applied for a wide range of asymmetric transformations by using descriptors calculated separately for catalysts or substrates [171, 173, 174].

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

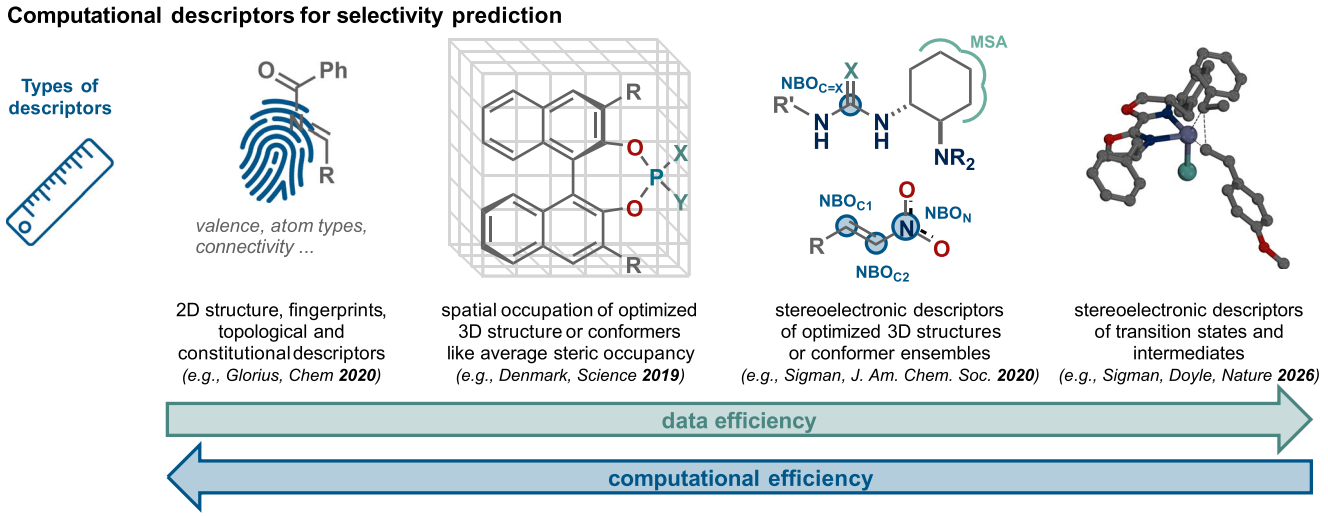


FIGURE 6 | Overview of descriptor classes used in selectivity prediction, ranging from simple 2D constitutional features (e.g., atom types and connectivity) over 3D spatial representations and quantum-mechanical stereoelectronic descriptors of optimized geometries and conformers, to transition state-derived descriptors.

Applying this approach to mechanistically complex reactions or achieving transferability between reaction classes, however, remains difficult as the enantio-determining step can shift with substrate or ligand identity and datasets are often sparse. To overcome this bottleneck, Sigman, Doyle, and coworkers introduced a transition-state-informed parametrization strategy in which steric and electronic descriptors are extracted from low-cost, GFN2-xTB-optimized structures resembling key intermediates and competing transition states, rather than from free ligands or simple surrogate complexes [175]. By capturing catalyst–substrate interactions directly within these transition state representations, MLR models trained on sparse Ni-catalyzed C(sp^3)–H activation and cross-electrophile coupling datasets were able to transfer predictive power across distinct reaction classes and suggest changes in the enantioinduction pathway. Through repeated, stratified nested *k*-fold cross-validation, the initial set of 1251 descriptors was reduced to the most predictive four-descriptor combinations. After cross-validation, an ensemble of the five best-performing models was assembled and used for prediction. With this feature-selection and model-construction pipeline, the authors showed that a model trained on C(sp^2)–C(sp^3) couplings could predict enantioselectivity for an out-of-sample C(sp^3)–C(sp^3) cross-electrophile coupling involving unseen ligand classes. The training set comprised 46 asymmetric reductive arylations and alkenylations of *N*-sulfonyl styrenyl aziridines or benzylic chlorides, using bisoxazoline, bi-oxazoline, and bi-imidazole ligands. Applied to alkylative aziridine ring opening, a reaction that proceeds with optimal yield and enantiomeric excess with pyridine–imidazoline ligands, the model achieved an MAE of 0.18 kcal mol^{-1} and an R^2 of 0.59. The model, therefore, transfered information learned on the C(sp^2)–C(sp^3) coupling to a mechanistically complex C(sp^3)–C(sp^3) case, despite the absence of pyridine–imidazoline ligands from the training data. Low-cost transition-state-based descriptors thus support model generalizability without requiring full DFT-level transition-state searches.

Essentially mechanism-agnostic enantioselectivity prediction models have also been implemented using other ML models like SVMs [102], RFs [176], GB [41], and deep neural networks. As an intermediate example, Hong, Ackermann, and coworkers developed a transition-state knowledge–guided workflow to predict enantioselectivity in pallada-electrocatalyzed C—H activations of biaryl aldehydes with olefins [177]. In this study, the authors generated "TS-like" distorted geometries for the C—H activation and alkene insertion steps and encoded site-specific physical organic descriptors that capture reaction-center properties, steric parameters of the chiral transient directing group, and condition effects. After feature selection, these descriptors were used to train a tree-based regression model for $\Delta\Delta G^{\ddagger}$ prediction. The dataset included 127 distinct C—H olefination reactions involving variations of 12 biaryls, 21 olefins, 20 transient directing groups, 6 solvents and 13 electrolytes. This knowledge-based encoding enabled accurate interpolation and extrapolation, outperforming generic molecular descriptors on out-of-sample substrates, achieving R^2 values of 0.822 and 0.910 on hold-out test data and external experimental validations with unseen biaryl aldehydes, olefins, and transient directing groups.

Shifting toward deep learning architectures, Zhang, Xing, and coworkers developed ChemAHNet, a chemistry-informed framework for predicting both enantioselectivity and absolute configuration in asymmetric hydrogenation [178]. The model operates directly on SMILES and uses three modules: a moiety identification module to extract functional group features around the C═C bond, a reaction component integration module based on multi-head attention to merge substrate, ligand and catalyst information, and a molecular interaction module that models substrate–catalyst interactions to predict interaction modes and $\Delta\Delta G^{\ddagger}$ values. Using this architecture and a curated literature dataset of 9478 asymmetric hydrogenation reactions of olefins, encompassing 1766 substrates and 1480 catalysts across diverse conditions, the authors achieved accuracies close to 90% when predicting the absolute configuration of the major enantiomer. They further demonstrated the model's applicability to $\Delta\Delta G^{\ddagger}$ prediction for other asymmetric catalysis problems, including organocatalysis and photoredox catalysis. Focusing on generalizability, Hong and coworkers reported the augmentation of GNNs by QM-derived descriptors [57]. By attaching atom-level steric

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and electronic parameters to the nodes of molecular graphs, they obtained a highly generalizable enantioselectivity prediction model that outperformed other state-of-the-art models on out-of-sample splits of Denmark's asymmetric thiol addition dataset (Table 1, entry 2) [102].

Collectively, these studies show that both mechanism-informed and largely mechanism-agnostic ML approaches can achieve accurate $\Delta\Delta G^{\ddagger}$ prediction across diverse asymmetric transformations. Mechanism-aware descriptors and transition-state-inspired representations improve extrapolation and reduce data demands, yet they require careful design and domain expertise. At the same time, purely data-driven architectures can reach high predictive performance when supported by appropriate encoding of catalyst, substrate, and interaction features. These models, however, require large amounts of data that are not available in early reaction optimization campaigns.

3.2.3 | Site- and Regioselectivity

Assuming reactivity under a Curtin–Hammett regime, site- and regioselectivities can be analyzed in a manner analogous to enantioselectivity, tracing their origins to $\Delta\Delta G^{\ddagger}$ values of competing, isomeric transition states. Consequently, both the previously discussed MLR-based approach and several deep learning models have been extensively employed to predict $\Delta\Delta G^{\ddagger}$ values that determine site selectivities. In contrast to enantioselectivity prediction, site- and regioselectivity may, however, require the evaluation of multiple competing transition states and sites with chemically distinct environments. This complexity has motivated the use of regression models that predict probabilities for competing sites, as well as property prediction models that identify the most favorable reaction site. As Kabeshov, Jorner, and coworkers recently reviewed computational tools for predicting site and regioselectivity [179], we focus here on general concepts using selected examples and refer the reader to their work for a comprehensive overview.

Classical MLR models have been applied, among others, to site selectivity in reactions such as C—H oxidation [180, 181] and amination [182]. In a notable extension beyond standard static QM-descriptors, Sigman, Davies, and coworkers introduced spatial descriptors that quantify the shape and accessibility of the catalyst's reactive-site cavity [183]. These descriptors enabled multivariate models that relate steric confinement and electronic properties of dirhodium catalysts to site selectivity in carbenoid C—H insertion reactions. The aforementioned approaches based on $\Delta\Delta G^{\ddagger}$ values, derived from the ratio of two isomers, are inherently limited to pairwise comparisons between competing sites. To overcome this limitation, Davies and Sigman later developed a site-by-site prediction strategy for dirhodium-catalyzed C(sp^3)–H functionalization of silyl ethers [184]. Using logistic regression, their model evaluates all potential C—H sites within a substrate based on site-specific steric and electronic descriptors and predicts the preferred site directly. A conceptually different approach, while still based on activation barrier prediction, was introduced by Hong and coworkers in two contributions on hydrogen atom transfer [185] and radical C—H functionalization reactivity [186]. For both classes of transformations, they generated a dataset of activation free energies $\Delta G^{\ddagger}$ for all relevant substrate sites and radicals using DFT calculations. GB and RF models were then trained to predict these barriers from reactant-based descriptors and the predicted activation energies were subsequently used to determine site selectivity across competing positions.

In contrast to $\Delta\Delta G^{\ddagger}$-based approaches that learn activation free-energy differences between competing sites, regression models have also been trained to predict selectivities directly. GNNs learning site selectivity from 2D molecular structures or from graph representations augmented with QM-derived descriptors have been the most prominent model choice for this task. Jensen and coworkers used Weisfeiler–Lehman networks with low-level atomic descriptors to assign likelihood scores for aromatic C—H functionalization to each atom. Models were trained and evaluated on 80/10/10 training/validation/testing pseudo-time splits of about 58 000 brominations, chlorinations, nitrations, and sulfonylations from Reaxys, achieving top-1 accuracies between 50.0% and 89.5% depending on the reaction class [187]. In later work, they extended this concept by combining learned reaction representations with on-the-fly-predicted quantum-mechanical descriptors in a hybrid QM–GNN model [188]. This model retained direct ranking of reactive sites while incorporating physically meaningful features, increasing top-1 accuracy to above 90%. Related GNN-based models that incorporate QM-descriptors have also been applied to metal-catalyzed cross-couplings [189] and Ru-catalyzed C—H functionalization [190]. For a high-throughput dataset of 956 iridium-catalyzed C—H borylation reactions, Schneider, Martin, and coworkers found that including 3D structural information improved performance, whereas DFT-derived atomic partial charges did not provide a significant advantage over simple atom-level features [191].

Site-resolved regression strategies are not limited to GNN-based models. Milo, Reisman, and coworkers combined a random forest model with an active learning strategy to predict site selectivity in oxidation reactions with dimethyldioxirane and methyl(trifluoromethyl)dioxirane [192]. For a target molecule whose site selectivity needs to be predicted, an acquisition function prioritizes candidate molecules for experimental testing. This function weighs both the model uncertainty and the similarity between each candidate and the target. The resulting workflow generates predictive models from fewer experiments and is applicable to complex substrates in a total synthesis context.

Finally, site-specific molecular properties have been used as proxies for selectivity that can be predicted with more general ML models. Examples include pK_a values for aldol reactions [193, 194], C—H hydricities for hydride-transfer reactions [195], BDEs (see Section 3.1) for oxidative metabolism [127], and NMR chemical shifts for electrophilic aromatic substitution [138]. While their conceptual simplicity simplifies data acquisition and training, these proxy approaches have important limitations. Jensen and coworkers highlighted this in their study of aldol reactions using pK_a values predicted by the pKalculator tool [193]. While the predicted pK_a identifies the thermodynamically favored site, it can fail under kinetic control. Analogous model limitations were observed by Grzybowski and coworkers when considering pK_a-based predictions for reactions with pre-coordinating bases [194]. Generally, property-based models can be trained on larger and more accessible datasets than reaction datasets. Their predictions,

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

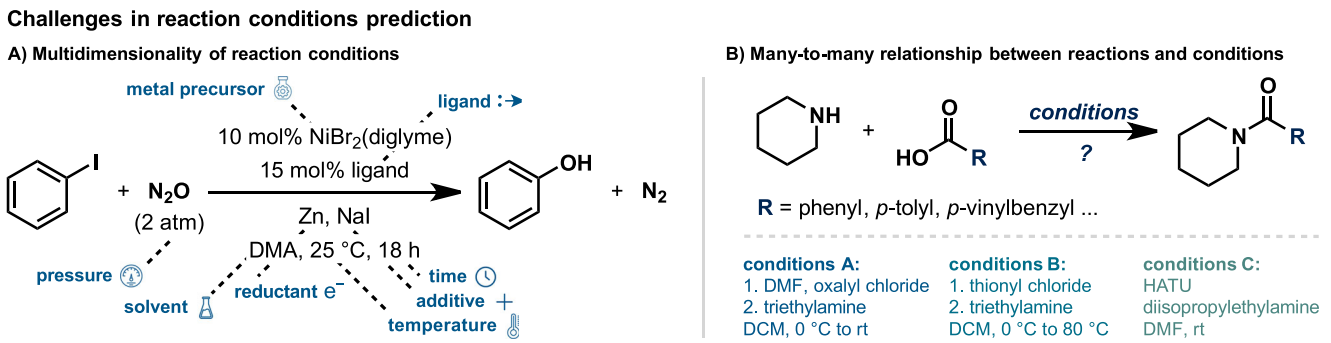


FIGURE 7 | Challenges in reaction conditions prediction. (A) Reaction conditions are multidimensional. A single transformation requires the specification of many interdependent components, including the metal precursor, ligand, reductant, additive, solvent, as well as continuous variables such as pressure, temperature, and time. (B) A many-to-many relationship links transformations and conditions. The same amide bond formation can be achieved by several distinct condition sets (A–C), which at the same time are appropriate for several substrate combinations or transformations.

however, can become unreliable when selectivity is influenced by additional factors not captured by the chosen property.

3.3 | Prediction of Reaction Conditions

Finding suitable reaction conditions for a given transformation is central to synthetic chemistry. From a data science perspective, this task is complex as has been outlined previously [196]. Depending on the reaction, several components, such as a catalyst, base, ligand, or additives, must be predicted (Figure 7A). Although assigning these roles is essential for dataset curation, database ontologies do not always define them unambiguously. In addition, categorical variables such as reagent identity and continuous variables such as temperature or reaction time must often be harmonized through preprocessing (e.g., binning), if a single model is used.

Condition prediction is further complicated by the many-to-many relationship between transformations and conditions (Figure 7B). Multiple distinct condition sets can enable the same reaction, yet typically only one or a few are reported, leading to substantial data sparsity. This intrinsic limitation not only hampers model training but also complicates evaluation. Most datasets link each reaction to a single condition set. A prediction that deviates from this set may still be chemically valid but will be counted as incorrect in a direct comparison. A rigorous evaluation would therefore require experimental validation. Although automated platforms have enabled in-lab validation of solvent predictions, for example, in work by Laino, Schwaller, and coworkers [197], such studies are rarely feasible at scale due to cost and time constraints (see Section 2.4) [196].

As a practical alternative, many studies evaluate top-*k* predictions rather than only the single most likely output. This approach reflects how chemists might screen a shortlist of plausible options, and it partially compensates for incomplete reporting. In addition, clustering of reagents, catalysts, or solvents can account for functional similarity, based on the assumption that similar reagents, catalysts or solvents might lead to similar reactivity. During evaluation, predictions can be accepted if the suggested component falls within the same cluster as the recorded compound, acknowledging that closely related reagents may be interchangeable [198].

Despite these intrinsic challenges, different strategies for condition prediction have been developed. Depending on the context, this might require choosing suitable conditions without a specific reaction in mind (e.g., in complex molecular synthesis) or selecting the optimal combination of catalysts, ligands or reagents from a set of established options (e.g., in library synthesis). Additionally, some applications might require identifying a complete set of reaction conditions, including reagents, catalyst, solvent, temperature, pressure and time, while in other scenarios, certain elements might already be fixed and only a subset of these variables must be predicted. In this regard, a distinction is often made between local and global predictions of reaction conditions [196].

3.3.1 | Local Reaction Condition Models

Local prediction models are typically used to select optimal conditions from a predefined set of combinations or components for a given transformation. An early example by Varnek and coworkers focused on hydrogenative deprotections and closely followed the way chemists reason by structural analogy [199]. They extracted over 70 000 reactions from Reaxys in which a protecting group is either cleaved or retained. From these precedents, they built a similarity-based recommender system that compares the condensed graph of reaction (CGR) [200] fingerprint of a query transformation against those of successful and unsuccessful cases. The system suggests catalysts by finding the nearest neighbors of the query in this reaction space. This conceptually simple approach achieved balanced accuracies between 85% and 95% in leave-one-out cross-validation for protected alcohols, phenols and amines.

Condition prediction can also be framed as a virtual screening problem. A yield regression model trained on substrates and conditions can be utilized to predict all condition combinations within a defined chemical space, with the goal of subsequently ranking them by desired outcome. This strategy has been widely used to identify complete condition sets or optimal ligands and catalysts for specific reactions. As a representative example, Doyle and coworkers designed a diverse HTE dataset of 3904 Chan–Lam couplings and encoded each reaction using DFT-derived sulfonamide descriptors, experimentally derived boronic acid speciation features, and one-hot encoded condition variables [201]. They compared several supervised regressors and selected

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

a multilayer perceptron to predict yields across all screened bases, solvents, catalysts, and boronic acids, assessing performance with rank-based metrics such as Spearman's ρ and average percentile ranking to prioritize high-yielding conditions rather than exact yields. The model was validated using random splits, leave-one-sulfonamide-out cross-validation, and external substrates, and successfully identified improved conditions for unseen substrates while reducing experimental trial-and-error.

However, identifying a single optimal reaction condition may not be necessary in local prediction settings, as multiple reagents or catalysts may be readily available for testing. Ranking conditions offers a practical advantage over predicting a single optimal condition because it prioritizes the best options from a predefined list, directly addressing the experimental goal of testing a small set of promising conditions. Applying this framing of the condition-prediction problem, Zimmerman, Cernak, and coworkers trained a label-ranking RF to learn which substrates share the same best condition and to output a full ranking of discrete reaction conditions [202]. Label-ranking was shown to outperform classical regression and classification models on both fully combinatorial and incomplete literature datasets featuring up to 18 distinct reaction conditions, by learning how conditions compare across substrates and leveraging these aggregated comparisons. Along the same line, Grzybowski and coworkers reported two related recommender systems that frame catalyst selection as a multi-class ranking problem over a predefined library of known catalysts. In an earlier study on enantioselective magnesium catalysis, they encoded a dataset of 1538 manually curated reactions using molecular fingerprints and reaction-core descriptors and trained a multilayer perceptron to rank one-hot encoded catalysts, achieving about 80% top-1 accuracy under cross-validation and validating selected predictions experimentally [203]. In their subsequent photocatalyst recommender, they collected a dataset of >36 000 literature reactions and replaced fixed fingerprints with a RoBERTa transformer pretrained on large reaction corpora, using concatenated reaction SMILES and atom-mapped reaction cores as input; this model ranked 31 discrete photocatalysts and improved top-1 cross-validation accuracy to about 90% [204]. In both cases, evaluation combined cross-validation metrics with prospective experimental validation on recently published reactions, demonstrating that the top-ranked suggestions consistently yielded productive conditions and often identified less obvious catalysts competitive with or superior to commonly screened "go-to" choices.

3.3.2 | Global Reaction Condition Models

Global reaction condition prediction models are typically trained on large amounts of literature data to recommend conditions for a wide range of reaction types. Beyond similarity-based methods global conditions prediction is largely driven by deep learning models such as feed-forward neural networks, GNNs, and transformers. Intrinsic limitations of many global approaches stem from the omission of variables that are rarely relevant or inconsistently reported, including pressure and light intensity. Their performance can also decline at the reaction-class level because they are trained on broad and chemically diverse datasets rather than on focused, class-specific data [205]. Despite these limitations, global models are important for autonomous synthesis and computer-aided synthesis planning [196]. Building on early work by Jensen and coworkers [206], subsequent studies have focused on improving deep learning architectures to enhance predictive accuracy and generality [207, 208].

Two major drawbacks of most global condition prediction models are the need for predefined and potentially ambiguous reagent roles such as "solvent", "base", or "catalyst" and the omission of reactant and reagent amounts. Sun, Jensen, Coley, and coworkers address these key limitations by treating all non-reactant species uniformly as "agents", which also allows the model to handle reactions with variable numbers of components [209]. Their quantitative recommendation of reaction conditions (QUARC) model uses a D-MPNN based on *Chemprop* to encode the reaction graph, which is combined with agent representations and passed to task-specific neural networks. The model predicts reaction conditions in four stages: agent identity, temperature, reactant equivalence ratios, and agent amounts, with agent sets generated autoregressively and quantitative parameters predicted as binned classification tasks. Training was performed on a version of the Pistachio dataset (Table 1, entry 5), which contained ~15 million (4.96 million unique) reactions extracted from patents. Across the different tasks, the model showed modest improvements over chemistry-relevant baselines based on reaction popularity or nearest-neighbor retrieval, showing improved accuracy particularly for reactions with diverse or sensitive conditions. By enabling varying numbers of agents and including reactants and agent equivalents QUARC can generate more flexible sets of conditions that require less added context. QUARC can be accessed via the web interface of the ASKCOS software suite, making it highly accessible for classically trained synthetic chemists [11].

Common to all condition prediction models discussed so far is that they only suggest a list of reaction components and selected numerical values such as equivalents or temperature. Performing a chemical reaction, however, requires a detailed procedure that also specifies the order of addition, time, and workup. Large language models (see Section 5) offer capabilities to bridge this gap. Utilizing this potential Batista, Newhouse and coworkers introduced MOSAIC, which builds on the open Llama model and was fine-tuned on millions of reaction protocols from the Pistachio database (Table 1, entry 5) to generate executable experimental procedures [210]. To balance global coverage with local specialization, the authors first embedded reactions into a learned reaction fingerprint space and then divided this space into many regions of chemically similar transformations, training a dedicated expert model on the protocols within each region. For a new query, MOSAIC identifies the most relevant expert models to produce a complete human-readable procedure. Across diverse reaction classes, this framework achieved a 71% overall experimental success rate and enabled the synthesis of numerous previously unreported compounds, illustrating how literature-scale training can be combined with focused domain expertise in a single system. This example highlights that recent advances have elevated reaction condition prediction from simple component matching toward data-driven ranking and probabilistic modeling across broad chemical spaces. Most importantly, the integration of LLMs enables the generation of complete, executable procedures, moving the field closer to autonomous and in-lab applications.

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3.4 | Data-Driven Catalyst and Ligand Design

Catalyst and ligand design are central to modern organic synthesis. By tuning a catalytic systems' activation barriers and selectivity, chemists can construct complex molecules more efficiently, sustainably, and at scale. Yet, developing new catalysts and ligands has historically been a tedious trial-and-error process. ML, combined with targeted experiments and QM featurization of catalysts, substrates, and transition states, offers a more systematic approach. Quantitative predictive models can guide the design of new catalytic systems in a data-driven manner rather than by intuition alone. This section is organized around conceptually distinct ML-based strategies for ligand and catalyst design, highlighting key principles and representative examples for each approach. More specialized and comprehensive reviews exist for specific areas, such as organocatalysis [211] or individual modeling strategies [166], and are recommended for readers seeking a broader survey of representative studies.

A foundational strategy in ML-guided catalyst design uses MLR models built on physically meaningful QM-descriptors for the virtual screening of candidates. As already discussed in Section 3.2, these models do not require an explicit mechanistic hypothesis and offer a lower-cost alternative to a large number of transition-state DFT calculations. When correlation analysis, feature selection, and validation are performed rigorously, MLR models can predict selectivity or yield for new catalyst structures and extrapolate beyond the best values observed in the training data [212]. Candidates predicted to outperform known systems can then be synthesized and evaluated experimentally. This workflow has enabled the design of bipyridine [213] oxazoline [214], and amino acid-based ligands [215], phosphines [216, 217], and phosphoric acid catalysts [218], among others. Analogous strategies have used other featurization methods and models, like a hybrid molecular field analysis (MFA) and an SVM-based strategy employed by Denmark and coworkers [102, 219]. An advantage of combining these chemically interpretable descriptors with transparent models is that the learned relationships can inform rational structural redesign. A different study that implemented a feature-informed optimization approach for MFA featurized catalysts was reported by Kanai and coworkers. For the optimization of chiral phosphoramidite ligands for iridium/boron hybrid catalysis, the authors mapped feature importances back by projecting the regression coefficients onto the original 3D grid surrounding the Ir-catalysts. This spatial information was then subsequently used to guide the redesign of structures [220].

Complementing and facilitating the application of MLR and other descriptor-based models in catalyst design, several curated databases of transition metal ligands and organocatalysts now provide reusable, standardized descriptor sets for virtual screening (Figure 8A). Representative examples include the ligand knowledge base (LKB) for carbon- and phosphorus–donor ligands [221–223], the Kraken discovery platform for monodentate organophosphorus(III) ligands [133], the BDL-Cu-2023 ligand database for chiral bidentate ligands [224], the OSCAR repository for organocatalysts [225], and NHC-cracker for *N*-heterocyclic carbenes [131]. All five resources follow a modular construction principle. A defined core scaffold is combined with systematically varied substituents to span steric and electronic properties. The libraries capture these properties using QM-descriptors calculated for standardized coordination environments or for free ligands and catalysts and have already supported multiple modeling studies. More importantly, they establish reusable and comparable descriptor frameworks that lower the entry barrier for data-driven catalyst design.

As previously outlined, more challenging prediction and, therefore, also catalyst design tasks might require the inclusion of more mechanistic insights. As an example for such a scenario, Schreiner and coworkers set out to identify improved catalysts for the enantioselective Corey–Bakshi–Shibata (CBS) reduction of butanone, a transformation that requires the challenging differentiation between a methyl and ethyl group in the substrate. They represented each catalyst–substrate pair as a single key-interaction graph by virtually connecting the boron of the oxazaborolidine catalyst with the carbonyl oxygen of the ketone. A GNN was then trained on this combined representation to predict $\Delta\Delta G^{\ddagger}$ values based on a small, high-quality dataset of 100 reactions. Subsequently, the model was used to evaluate a small set of candidates for the targeted reaction. Experimental validation of a catalyst predicted to afford higher selectivities increased the enantioselectivity of the butanone reduction to 80% *ee* and thus surpassed previously reported CBS catalysts [226].

A conceptually distinct strategy for accelerated catalyst identification based on unsupervised ML was pioneered by Schoenebeck and coworkers (Figure 8B). In contrast to supervised models, no training on labeled datasets is required. Instead, catalysts or ligands are embedded in a chemical space defined by physically meaningful descriptors, typically derived from QM calculations or curated parameter databases. An unsupervised clustering algorithm then partitions this space according to intrinsic similarity patterns [227]. The major advantage of this strategy is that a small number of experimental reference points is sufficient to guide the search. As known active examples are located within the descriptor space, new candidates are selected based on proximity to these references in the multidimensional chemical space. Because the algorithm detects structure–property relationships without fitting to experimental outputs, the experimental burden is low. While the approach therefore enables identifying promising catalysts with minimal initial data, the descriptors used to construct the chemical space need to be based on detailed mechanistic hypotheses and accurate DFT methods to provide a chemically meaningful embedding. Following the successful application of this strategy for the identification of dinuclear palladium- [227] and nickel catalysts [228], Schoenebeck and coworkers recently targeted CO_2 insertion at Ni(I) centers. The central design hypothesis was that ligands capable of stabilizing monomeric Ni(I)–Ph species would enable room-temperature carboxylations [229]. A problem-specific bidentate phosphine ligand library was constructed in silico by modifying established backbones covering a range of bite angles with different substituents. DFT calculations were performed on the corresponding Ni(0), Ni(I)–Ph, and Ni(II)–Ph complexes. From these structures, steric and electronic descriptors were extracted to describe how each ligand influences metal speciation. After grouping ligands by unsupervised clustering (Figure 8B), ligands located in the same cluster as the positive reference *tert*-butyl-xantphos were

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

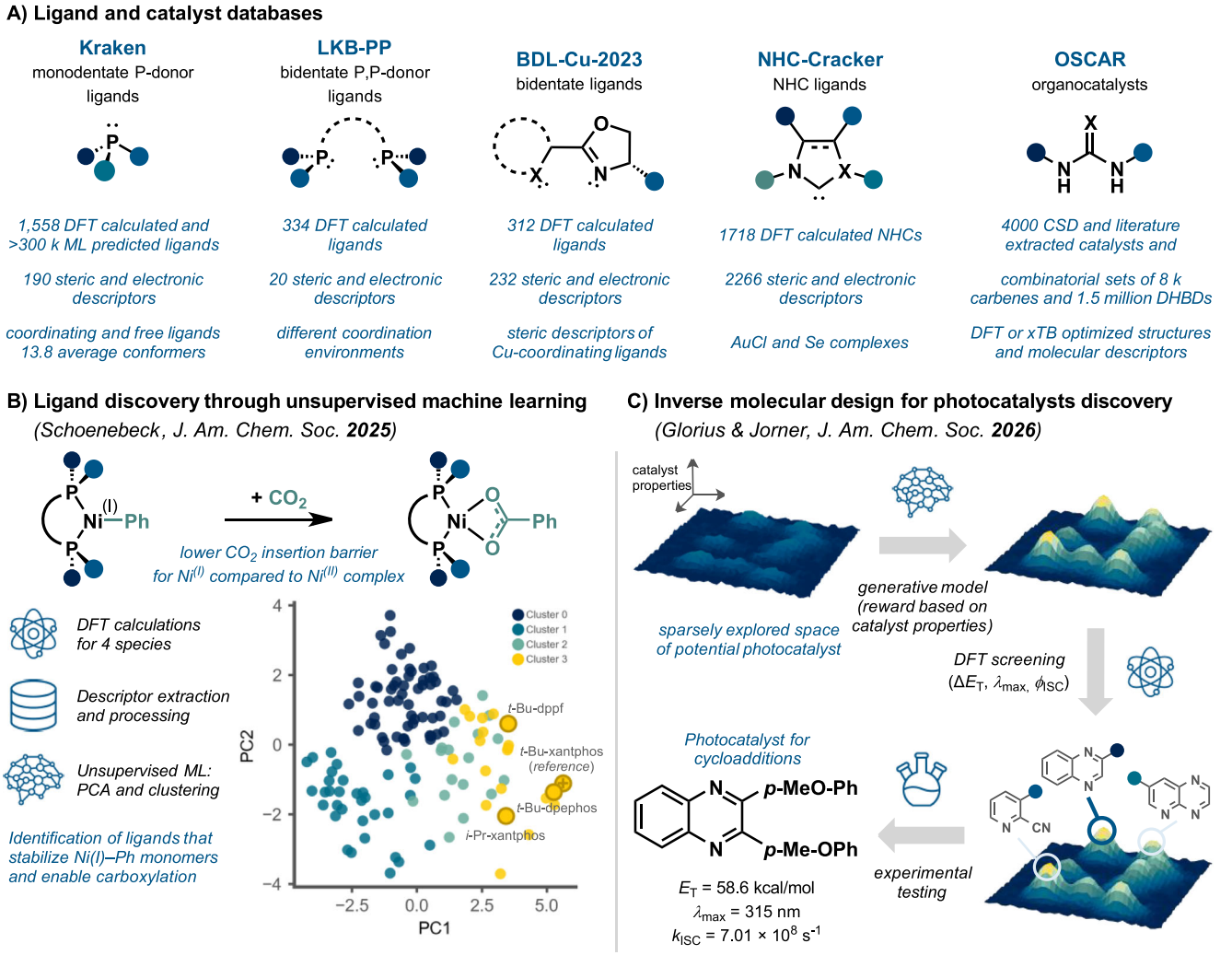


FIGURE 8 | Data-driven strategies for catalyst and ligand design. (A) Representative curated ligand and catalyst databases. (B) Unsupervised clustering for ligand discovery from minimal experimental input, as demonstrated by Schoenebeck and coworkers for Ni-catalyzed CO_2 insertion. (C) Generative modeling enables inverse catalyst design by exploring continuous chemical space and optimizing target properties, as illustrated by Glorius and coworkers for photocatalyst discovery.

predicted to favor Ni(I)–Ph speciation. Representative ligands, including *i*-Pr-xantphos and *t*-Bu-dpephos, were synthesized and tested experimentally, where the observed formation of Ni(I)–Ph complexes and their reactivity toward CO_2 confirmed the initial hypothesis.

Last, a fundamentally different approach toward ML-driven catalyst development is generative modeling [230]. Generative AI for catalyst design combines a generative model and a property predictor to explore chemical space in an inverse manner. Typically, a variational auto-encoder, flow model, or related architecture learns a continuous latent space from string or graph representations of catalysts, while a supervised surrogate model predicts a target property such as binding energy [231], yield, or selectivity [232]. Optimization is then performed directly in this latent space to generate new structures with desired properties [233].

Along this line, our group has recently developed a generative framework for the discovery of photocatalysts (Figure 8C) [234]. Building on the concept of inverse molecular design, we developed a two-step strategy that combines initial global generative exploration with a local refinement of photochemical properties to discover organic energy-transfer photocatalysts. In the global phase, a reinforcement learning model was trained to generate molecules as SMILES strings [235, 236]. Candidate structures were scored using a multi-objective reward function based on predicted triplet energy, absorption wavelength, and calculated proxies for intersystem crossing and structural rigidity. By iteratively updating the generative policy to maximize this score, the model navigated a more continuous property space rather than predefined structural libraries and identified candidates in underexplored regions. Promising scaffolds from this global search were then subjected to a focused local exploration. A diverse subset was selected and validated using TD-DFT calculations to refine photophysical properties. Following this step, bicyclic pyrazine scaffolds were further expanded into a small, synthetically accessible library, from which candidates were synthesized in one step and experimentally validated in multiple energy-transfer reactions.

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Focusing on transition-metal catalysis, Balcells and coworkers developed a generative framework for the inverse design of ligands and complexes based on a variational autoencoder [237]. Ligands were encoded as SMILES strings with explicit metal–ligand bonding and embedded into a latent space learned from a DFT-annotated ligand library. In an initial unconditional phase, the model generates diverse mono- and bidentate ligands, which are assembled into homoleptic complexes and labeled with properties such as the HOMO–LUMO gap and the metal-centered charge. A conditional model then links latent representations to these properties, enabling gradient-based optimization to generate ligands that match user-defined targets, for example, maximizing stability or tuning electron density at the metal center.

These and related generative approaches enable a largely autonomous exploration of catalyst spaces. In contrast to library-based screening, they are not restricted to predefined scaffolds, fragments, or chemist-curated databases. Instead, these models navigate property space directly and can uncover structures that would not emerge from conventional enumeration strategies. However, this flexibility comes at a cost. Each generated candidate requires on-the-fly property prediction or QM-based evaluation, which increases computational demand. In addition, generative models may propose structures without clear synthetic precedent, creating challenges for practical implementation unless synthetic accessibility is explicitly controlled during the design process [238, 239].

Summarized, each of the presented strategies is most effective in a specific design context. Descriptor-based MLR models are well-suited for moderate datasets when interpretable structure–property relationships are needed. Curated databases support efficient screening within established ligand classes, while mechanistically informed supervised models address complex selectivity problems by encoding key interactions. When only a few experimental references are available, unsupervised clustering can guide exploration within a chemically meaningful descriptor space. In contrast, generative modeling is most powerful for venturing beyond known scaffolds into underexplored regions.

4 | Data-Driven Development of New Synthetic Transformations

Developing novel chemical reactions remains a central challenge of synthetic (organic) chemistry. New synthetic methodologies often not only compete with alternative synthesis routes on sustainability and economic grounds but also provide access to unprecedented core structures that were previously unattainable, rendering research in this area indispensable for downstream applications such as drug development. However, developing new chemical reactions as synthetic tools is a labor-, time-, and resource-intensive process, which typically begins with the discovery of new reactivity, followed by optimization efforts, mechanistic investigations, and a scope evaluation to demonstrate practical utility [240]. In this section, we will highlight how data-driven tools can accelerate and enhance all of these stages, aiming to improve the quality of the methodology itself, the corresponding report/publication and ultimately leading to a facilitated adoption by synthetic practitioners.

4.1 | Accelerated Reaction Discovery

Discoveries in synthetic chemistry describe the generation of new chemical knowledge that changes how chemists, or predictive models of chemistry, understand and apply chemistry. It goes beyond incremental optimization or routine variation. Instead, it introduces a new concept or extends an existing concept outside its expected domain of applicability. This definition exposes an intrinsic paradox when ML is used for immediate discovery. Most ML models show a marked decline in performance outside their applicability domain. Yet discovery, by definition, requires moving beyond the knowledge encoded in the training data. With this limitation in mind, we highlight how ML has nonetheless been used to support and accelerate discovery [241, 242].

Most discoveries in synthetic chemistry concern new molecules, new reactivities or new transformations. In this review, we will focus on the latter two cases. We will distinguish them based on whether new reactions emerged from a broad exploration of reaction space without a predefined target or whether a specific product motif or transformation was defined in advance.

Cronin and coworkers have developed a series of autonomous platforms aimed at discovering new reactivity with minimal human intervention (Figure 9A). A common feature of these systems is the unbiased exploration of reagent combinations by a robotic platform equipped with online analytics. Reaction mixtures were evaluated by NMR and IR spectroscopy, and the outcomes of each experiment informed the selection of subsequent experiments. In all cases, the workflow was iterative: experiments generated data, models were updated, and the updated models guided the next round of experiments toward regions of higher expected reactivity [243, 244]. In one of these studies, reactivity assessment was framed as a classification problem. An SVM, trained on a small set of manually labeled reactions, classified mixtures as reactive or nonreactive based on inline NMR and IR data. Chemical space was represented using simple one-hot encodings of reagent identity. A linear discriminant analysis (LDA) model was then trained on the accumulated reactivity labels and iteratively retrained as new data were acquired. This model ranked untested reagent combinations by their predicted reactivity probability, thereby steering exploration toward productive regions of chemical space. Based on a pool of 18 reactive, structurally diverse molecules, new multicomponent reactions were discovered using the platform. While these approaches feature a high degree of autonomy, their impact depends strongly on the composition of the search space. If productive substrate combinations are underrepresented, the likelihood of identifying useful reactivity remains limited [245].

An alternative algorithmic, rather than ML-driven, approach toward the discovery of novel reactivities is the systematic construction and analysis of reaction networks. These networks are typically generated by iteratively applying expert encoded transforms to a set of substrates and intermediate products. While previous studies filtered such networks using thermodynamic criteria or molecular complexity measures to prioritize feasible and potentially novel pathways [246–249], Coley and coworkers introduced a conceptually different strategy that exploits the generalization limits of ML models as a novelty filter (Figure 9B) [250]. Based on hundreds of expert-encoded reaction templates,

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

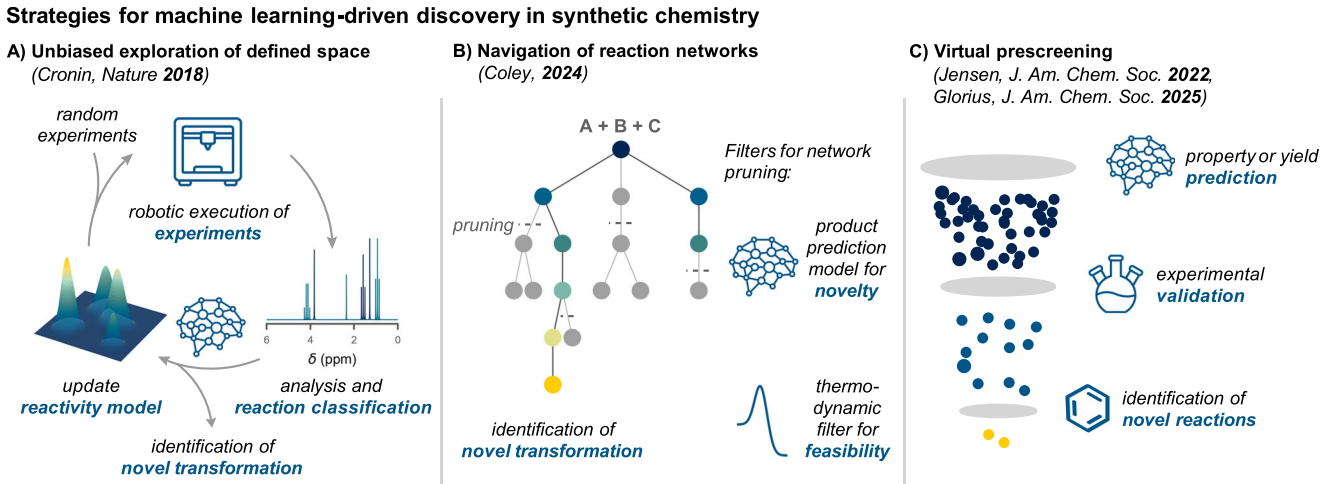


FIGURE 9 | ML strategies for reaction discovery in synthetic chemistry. (A) Autonomous platforms developed by Cronin and coworkers iteratively explore predefined reaction spaces through robotic experimentation, online analytics, and model updates to identify new reactivity without predefined targets. (B) Coley and coworkers construct and prune reaction networks using thermodynamic and ML-based novelty filters to prioritize feasible yet previously unpredicted transformations. (C) Virtual prescreening approaches by Jensen and Glorius and coworkers use ML models to predict reactivity or key properties across large substrate libraries, enriching promising candidates prior to experimental validation.

they constructed a mechanistic network functioning as a "virtual flask" to simulate the reaction pathways of reactants. To ensure feasibility, they applied a thermodynamic filter that removed pathways with large ΔG differences. They further introduced a novelty filter, defining a product as novel if it could not be predicted by state-of-the-art product prediction models. After the platform was validated by rediscovering known multicomponent reactions, the authors enumerated triplets for a random subset of 39 compounds from an in-house inventory (covering amines, acids, and various common nucleophiles) to demonstrate the pipeline's utility in the discovery of novel reactions. Adapting reaction conditions proposed by the previously described recommender from Jensen and coworkers [206], experimental screening confirmed two previously unreported transformations: an anhydride ring-opening and a Mannich-like aminomethylation. An advantage of this approach is that initial screening is fully virtual, minimizing costs and experimental effort until promising reactivities are identified.

With the goal of targeting a specific class of transformations (Figure 9C), Jensen and coworkers introduced a more focused strategy toward discovery using ML not to navigate a small chemical space or evaluate novelty, but to prescreen large numbers of structurally diverse substrates for electrochemical competence [251]. Their study focused on convergent paired electrolysis with 1,4-dicyanobenzene under fixed reaction conditions. Rather than exploring random reagent combinations, they used ML to make informed pre-selections that boost success rates. The workflow integrated three steps. They first constructed a quantum-chemistry-informed molecular representation based on DFT-derived atom features relevant to electrochemistry. Next, they generated a labeled dataset by automated droplet electrochemistry, testing 141 substrates and classifying them as competent or incompetent. Only 60 of these 141 substrates (42%) showed conversion, and just 27% when excluding known reactive motifs. A supervised classifier trained on this dataset was then applied to 38 865 virtual candidates, predicting 824 as reactive. Twenty were selected for validation, of which 16 (80%) converted the electrochemical partner, substantially increasing the hit rate from 42% observed during training data generation and yielding several previously unreported transformations.

This idea of accelerating the discovery of transformations involving specific substrates and reactivity modes through virtual prescreening was also employed by our group for the discovery of novel EnT-catalyzed reactions. In this study, we targeted the intermolecular dearomative cycloaddition of monocyclic (hetero)arenes, whose high triplet energies typically prevent visible-light EnT activation. A data-driven three-layer screening strategy was implemented. First, a virtual library of mono-, di-, and trisubstituted heteroarenes was evaluated in silico using the ML model EnTdecker, which predicts triplet energies and spin densities from molecular structures (see Section 3.1). This mapping revealed substitution patterns and functional groups that lower triplet energies into a catalytically accessible range and maintain spin localization on the heteroarene core. Second, luminescence quenching experiments validated productive catalyst–substrate interaction. Finally, selected candidates were subjected to reaction-based screening with radical acceptors. This ML-guided prescreening narrowed the experimental search space and enabled the discovery of EnT-catalyzed dearomative cycloadditions of thiophenes, oxazoles, and thiazoles, unlocking triplet-state reactivity in a substrate class previously considered inaccessible [129].

ML has proven valuable for accelerating reaction discovery, but its role differs between open exploration and targeted campaigns. In broad searches, it can guide iterative navigation of predefined chemical spaces. However, in more focused studies, it primarily serves as a prescreening tool that enriches promising candidates within clearly defined boundary conditions, because most models still struggle to generalize beyond their training domain or encode meaningful chemical understanding. Nevertheless, successful discovery workflows can, however, use ML to support almost

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

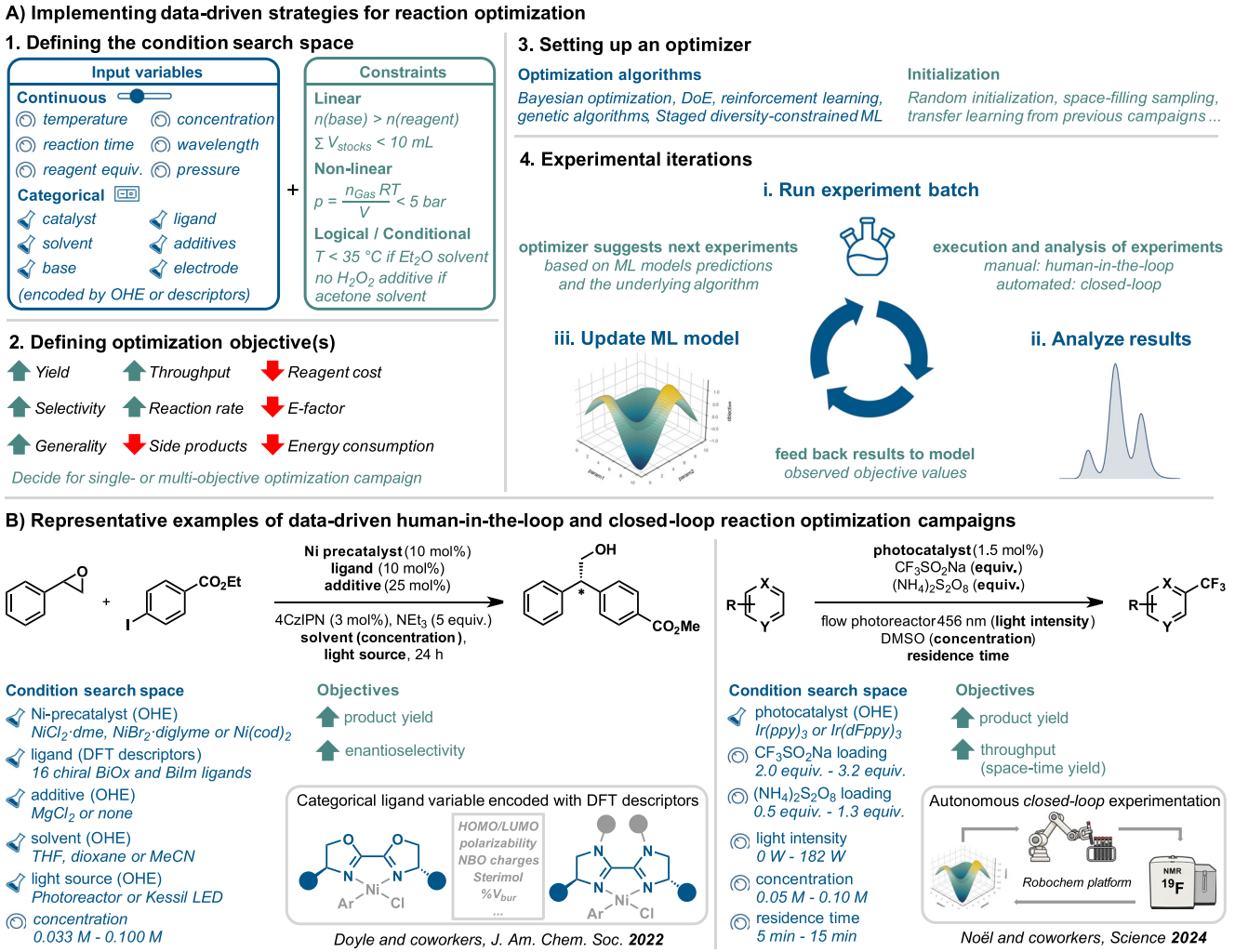


FIGURE 10 | Overview of data-driven strategies for reaction optimization. (A) Workflow steps to consider when applying data-driven reaction optimization in a real-world scenario. (B) Representative previous studies that used Bayesian optimization (BO) in multi-objective optimization campaigns for cross-electrophile couplings (left) and for self-optimizing flow reactions in a fully autonomous and closed-loop manner.

every step of the scientific process, including hypothesis generation, virtual screening, active selection of relevant experiments and data analysis [252].

4.2 | Data-Driven Strategies for Reaction Optimization

The initial discovery of a novel reactivity is just the inception of the discovery process. Often, the desired products are obtained in low yields and under suboptimal conditions. Reaction optimization, therefore, aims to turn this initial finding into a practical and broadly applicable synthetic method. To do so, reaction parameters are systematically varied. These can be discrete variables, such as the solvent, catalyst, additive, or addition order, or continuous variables, such as temperature, reaction time, or irradiation wavelength. The process continues until the defined goals (objectives) of the optimization campaign are met. In data science terms, the full set of possible parameter combinations and their constraints constitutes the (condition) search space of the optimization campaign (Figure 10A, Step 1). Furthermore, optimization campaigns can be divided into single-objective and multi-objective optimization campaigns. In academic settings, a maximum product yield is often considered the classical single objective of reaction optimization studies. However, attempts to simultaneously maximize selectivity (e.g., chemo-, regio-, diastereo-, or enantioselectivity) of a reaction already face the practitioner with a multi-objective problem. When diving deeper, also into industrial settings, economic or sustainability objectives may become relevant—such as minimizing reaction time or temperature, or preventing the use of hazardous compounds. In these multi-objective settings, by definition, no single optimal set of parameters can be identified. Instead, the aim is to identify the set of so-called Pareto-optimal parameters that trade off among the individual objectives.

In most research labs, optimization campaigns are still commonly performed using a one-factor-at-a-time (OFAT) approach. In this strategy, only one parameter (e.g., the catalyst) is varied, while all others (such as solvent, additives, temperature, or reaction time) are kept constant. Although straightforward, this method rarely identifies the true global optimum and often requires many experiments. A similar issue arises in multi-objective settings, when individual objectives are

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

optimized sequentially, overlooking their potential interdependence [253].

Considering these challenges, data-driven optimization strategies have proven capable of reaching better optima more quickly and with fewer experiments. The key distinction from the traditional human-led OFAT approach is that experimentalists often pursue incremental yield improvements with each new trial, whereas algorithms aim to construct a multidimensional surrogate model that captures the quantitative relationship between input parameters and outcomes. While such complexity is difficult for humans to grasp across many variables, these models make it possible to mathematically pinpoint the global optimum rather than settling for a local one.

By now, data-driven optimization strategies are standard tools for solving optimization problems across many fields. While we focus on the adoption of those strategies in synthetic labs in this review, we refer to an excellent in-depth review for a broader background on optimization strategies and specific applications [253].

4.2.1 | Implementing Data-Driven Optimization Strategies in a Real-World Scenario

4.2.1.1 | Defining a Realistic Search Space. The first, but perhaps most important, step in any optimization campaign is to clearly define the search space. This means identifying all adjustable parameters and specifying their type (continuous or discrete) and allowed values. For continuous variables, it is advised to set practical limits and increments (e.g., temperature from 0°C to 100°C in 10°C steps). For categorical variables, possible options, such as different catalysts or solvents, should be listed. For chemical compounds, adding suitable descriptors (see Section 2.2) can improve optimization performance. A frequent mistake is starting with a search space that is too broad, yet having no initial results indicating the presence of promising conditions within it. Although exploring vast chemical spaces is attractive, the number of experiments needed to find an optimum increases exponentially with dimensionality. It is therefore advised to begin with a manageable space aligned with experimental capacities. Expanding boundaries or adding parameters in later campaigns, after obtaining early results and insights into the importance of each dimension, is often more effective than spreading efforts too thinly across an unmanageable space. Beyond defining parameters, practical constraints must also be considered. High temperatures might exceed solvents' boiling points, stoichiometric ratios may need to remain fixed (e.g., excess base), and safety considerations are essential, for example, are potentially dangerous mixtures part of the search space?

4.2.1.2 | Defining Objectives. Before launching an optimization campaign, the objectives must be clearly defined. While yield is the most ubiquitous target in academic methodology development, the scope of data-driven optimization strategies extends far beyond it. Modern campaigns increasingly address multi-objective problems that reflect the complexity of real-world chemical synthesis, such as simultaneously maximizing selectivity (chemo-, regio-, or enantioselectivity) while minimizing catalyst loading, reaction time, or ecological and economic footprints.

More recently, reaction generality has also gained focus as a central optimization objective. While a universally accepted metric for generality remains missing, recent studies shift the focus from optimizing a single model substrate to evaluating average performance across a diverse set of substrates [254–259]. For a more detailed discussion of reaction generality, we refer to an excellent recent review [260].

4.2.1.3 | Choosing and Setting Up an Optimization Algorithm. With the search space and objectives in place, the next step is choosing an optimization strategy. While Design of Experiments (DoE) [261–263] remains a powerful "non-ML" tool for robust process characterization, Bayesian optimization (BO) is widely regarded as the gold standard for efficiently navigating complex, nonlinear chemical landscapes with limited data [264, 265]. Notably, BO algorithms come in many flavors, and a variety of surrogate model architectures, kernels and acquisition functions can be chosen and/or expert designed. As these choices are beyond the scope of this review, the interested reader is referred to specialized reviews about BO for detailed information [266–270]. Fortunately for synthetic chemists and practitioners, a variety of well-maintained and documented software packages for chemical reaction optimization using BO have been developed by both academic and industrial communities (Table 3). These frameworks are easy to use with little or even without Python knowledge, while offering both advanced customization for experts and accessible entry points for beginners.

Furthermore, in addition to pure and adapted BO-based algorithms [278, 283], bandit optimization [284, 285] or genetic algorithms [286] have recently been employed in combination with HTE techniques to optimize chemical reactions with a particular focus on general reaction conditions [260]. Also, Hong and coworkers have shown that algorithmic approaches such as staged diversity-constrained ML can outperform BO, especially in high-dimensional settings [287]. Last, LLMs combined with specialized prompting strategies have recently also shown promising performance as "optimizers". The key difference in this approach is that LLMs do not explicitly build a surrogate model of the investigated reaction but rather leverage their pretrained knowledge about reaction conditions to warm-start optimization campaigns or suggest next experimental conditions in a conversational interface (see Section 5.2 for details).

4.2.1.4 | Optimization Cycles. Once the search space is constructed, the objectives are defined and the optimization algorithm is initialized, the experimental phase begins. For the first set of reactions, the optimizer suggests initial experiments either randomly or based on sampling techniques that broadly cover the search space. Alternatively, the campaign can be warm-started using prior experimental results from the same or a related campaign (transfer learning [288]). Once the chemist performs and analyzes these experiments, the results are returned to the model. In a typical BO loop, a surrogate ML model predicts all untested conditions of the search space, and an acquisition function selects the next experiment, balancing exploitation of promising regions with exploration of uncertain ones. Consequently, the algorithm will occasionally suggest experimental conditions that appear counterintuitive or even illogical to the expert chemist. It is vital to resist the urge to discard these suggestions, especially the unexpected or low-performing results, which are information-rich

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 | Non-exhaustive overview of selected, community-driven software end-to-end frameworks for chemical reaction optimization using Bayesian optimization algorithms. Not including other, non-Bayesian frameworks.[a]

Framework	Key features	References
Phoenics/Chimera	Early contribution, BO over continuous variables, Chimera added multi-objective support.	[271–273]
Gryffin	BO over continuous and categorical variables.	[273, 274]
EDBO/EDBO+	BO over continuous and categorical variables, evaluated experimentally. EDBO+ update added multi-objective support and an interactive web app.	[275, 276]
Atlas	Broad general purpose BO framework, aiming also beyond reaction optimization toward self-driving laboratories.	[277]
BoTier	BO for hierarchically structured (tiered) objectives.	[14]
CurryBO	BO over curried functions, specialized framework for optimizing toward general reaction conditions.	[278]
BayBE	Aiming at industrial and "real-world" applications, including transfer-learning strategies.	[15]
BoFire	Aiming at industrial and "real-world" applications, broad constraint handling, serialization options and API-ready.	[279]

[a]Honegumi provides a graphical interface to set up an advanced BO campaigns. Code is automatically generated based on more low-level Python frameworks [280]. Labmate.ML is a non-Bayesian alternative specifically designed for reaction optimization in low-data settings [281]. The Summit framework is not directly focused on application, but rather on benchmarking different models in the context of reaction optimization [282].

and essential for refining the model's decision boundary, rather than failures of the algorithm. In this experimental phase, the chemist's task is to ensure that the experiments are carefully executed and analyzed, so that the fed-back data is of high quality. This iterative cycle continues until the objectives are met. Many platforms also allow dynamic adjustment of the search space, enabling contraction (e.g., lowering a temperature limit) or expansion (e.g., adding ligands) as new insights emerge. Notably, this type of iterative campaign is termed human-in-the-loop optimization. With advances in laboratory automation, the loop can also be closed by employing hardware and software capable of setting up, running and analyzing reactions automatically. These closed-loop approaches minimize the need for human intervention and are at the core of the self-driving lab concept (see Section 5.3), in which scientists are primarily responsible for planning and overseeing the campaign, while experimental goals are achieved autonomously [289, 290].

4.2.1.5 | Representative Examples. As an example for a multi-objective human-in-the-loop optimization campaign, Doyle and coworkers employed their EDBO+ platform to simultaneously maximize yield and enantiomeric excess in an enantioselective Ni/photoredox-catalyzed cross-electrophile coupling. In this study, the search space consisted of 1728 configurations, including both categorical (3 Ni-precatalysts, 16 ligands with DFT featurization, 2 additives, 3 solvents, 2 light sources) and continuous (reaction concentration) variables (Figure 10B). The algorithm identified Pareto-optimal conditions (80% yield, 91% *ee*) in just 24 experiments, significantly surpassing a previous OFAT optimization campaign that required approximately 500 trials to achieve comparable results [276]. In a pioneering example of closed-loop optimization, Noël and coworkers introduced RoboChem, an autonomous platform for optimizing chemical reactions in flow. The authors combined a Python-based BO backend with a modular programmable hardware architecture. Specifically, a continuous-flow photoreactor was coupled to an inline NMR spectrometer, enabling automated analysis of crude ^{19}F-NMR yields. After feeding the results back to the BO algorithm, a programmatically instrumented setup (liquid handler, pumps, light sources, etc.) can autonomously execute the next suggested experiment. As a case study, the authors reported the multi-objective maximization of both reaction yield and throughput in a photocatalytic (hetero-)aryl trifluoromethylation (Figure 10B). In contrast to Doyle's work, the investigated search space included several continuous (concentrations of substrate, CF_3-agent and oxidant, residence time in the photoreactor and light intensity) and only one categorical variable (photocatalyst, one-hot-encoded) [291].

4.2.1.6 | Troubleshooting: When Data-Driven Optimization Campaigns Do Not Yield Any Improvements. Despite having a data-driven optimization campaign set up and running, synthetic chemists may encounter situations in which no improvements in reaction performance are achieved across multiple optimization iterations. While problems with model parameters are technically possible [292], they are unlikely when high-level optimization frameworks (Table 3) with established presets are used. Most often, failure is caused by more tangible reasons. First, experimental noise can prevent surrogate models from effectively learning meaningful input–output relationships. This can be easily diagnosed by evaluating the model's cross-validation metrics over successive iterations; if the training error remains persistently high or the cross-validation performance flatlines early on, it strongly indicates that the algorithm is struggling to separate true chemical effects from experimental variance. To mitigate, the reproducibility of the experimental setup and sensitivity to operational parameters [293, 294] should be carefully evaluated. Furthermore, the robustness of the chosen analytical procedure for readout of the target variables should be critically assessed. For example, crude reaction yields can be more accurate than isolated yields as a readout, but spectra or chromatograms should always be checked for convoluted peaks and correct integration [295, 296]. Second, poorly chosen descriptors can prevent the surrogate model from learning meaningful

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

relationships, especially if high-dimensional features such as long 2048-bit ECFPs or latent language model representations are used without adapting technical parameters [297]. When no meaningful and suitable physical descriptors are available at all, even one-hot encodings can represent an alternative that might enable efficient local optimization by treating each entry of that search space dimension as a distinct, independent option, even though the algorithm will lose the ability to extrapolate to novel chemical entities. Last, some of the most common issues in data-driven optimization campaigns stem from poorly defined search spaces, which can arise for various reasons. If the search space becomes too large, ML models will eventually run into the curse of dimensionality. In such cases, the model itself is not necessarily failing. Rather, it simply requires an impractically large amount of data before it can capture meaningful input–output relationships [298, 299]. This case is typically characterized by a landscape dominated by maximum predictive uncertainty, which causes the acquisition function to always favor exploration over exploitation, continuously sampling distant, uncorrelated regions without ever converging. To mitigate this issue, practitioners must strategically downsize the search space. This entails either removing low-impact variables or tightening the boundaries of existing parameters to focus experimental resources on the most promising regions. However, even with adequately sized search spaces, campaigns can still struggle to yield improvements. In these scenarios, the underlying cause is often associated with non-smooth and complex reactivity landscapes, typically involving reactivity cliffs that are inherently challenging to capture with smooth models. Besides the elimination of those cliffs from the search space, a common mitigation strategy is either switching the surrogate model to tree-based models, moving to ranking-based optimization methodologies [300] or using non-model algorithms such as particle swarm optimization, which are inherently well-suited for rough reactivity landscapes [301]. With all of these considerations in mind, last and most disappointingly, a possible reason for the failure of data-driven optimization campaigns can also be chemistry itself: if there are simply no high-yielding conditions present in the investigated search space, neither a human nor the best algorithm will ever achieve the desired optimization goals.

4.3 | Advancing the Evaluation of Reaction Scopes Using Data Science

Once a chemical reaction is optimized, the next key step is to evaluate its scope to clearly show both the strengths and limitations of the method. However, growing pressure to present results positively, satisfy reviewers, and secure publication has led to an increase in average scope size. Despite the larger number of examples reported, selection and reporting biases skew the results toward successful substrates, often overlooking or even hiding the method's limitations (Figure 11A) [302–306]. In contrast, we argue that substrate scope quality should not be judged by the number of entries or their average performance, but rather by the information content it provides to the reader, allowing to estimate the performance of unseen substrates. Notably, this goal aligns with ML models that predict reactivity across different substrates. This opens the door to designing scope studies that a) provide higher information content and b) maximize the predictive power of models that are trained on the scope data and can be subsequently used for extrapolation (Figure 11B). Achieving both requires selecting the most informative experiments, including "negative" ones [67, 70], raising the question of how to reduce bias and select substrates for an ideal scope study.

Broadly, two main strategies have emerged, differing in whether substrates are defined upfront or selected dynamically during scope evaluation. All approaches, however, require careful construction of a candidate space, which consists of substrates that could realistically be tested. The assembly of that space starts with clearly defining the reactive substructure. Considering olefins as an example, this could also include distinguishing between unactivated olefins and activated olefins such as styrenes. Suitable candidates are then gathered from commercial compound libraries, such as vendor catalogs. Before launching a data-driven scope study, these lists must be carefully filtered based on practical factors (e.g., price and availability) and scientific constraints. Incompatible functional groups, for instance, should be excluded based on intuition or prior knowledge [307, 308], requiring some understanding of the reaction even before the study begins (Figure 11C).

4.3.1 | Unbiased Substrate Selection by Prior Chemical Space Analysis

Aligning with the first category, early contributions by researchers at Merck established the concept of "informer libraries"—curated collections of small molecules (e.g., aryl boronates) that are selected via principal component analysis (PCA) to broadly cover the physicochemical space of marketed drugs [309]. This approach enabled a direct, standardized comparison of synthetic methods that used these informers, which, however, are not readily available and provide only limited insight into the underlying reactivity or mechanism.

Shifting toward more reactivity-related properties, academic groups leveraged molecular descriptors to guide selection. As a pioneering example, Sigman and coworkers used physical descriptors, such as Sterimol parameters and IR frequencies, in combination with DoE principles to identify diverse and representative subsets of ketones, which were then used to obtain a reactivity dataset for multilinear enantioselectivity models [310]. Extending this concept, Doyle and coworkers developed a comprehensive, general workflow integrating automated DFT calculations [51] to derive descriptors and generate a scope candidate space before performing unsupervised sampling techniques. Specifically, they employed Uniform Manifold Approximation and Projection (UMAP) and clustering algorithms to select aryl bromide scope substrates based on computed electronic and steric features, resulting in a substrate scope with maximized chemical space coverage and ultimately enabling the training of predictive regression models [311]. This workflow has become a workhorse, extensively used in many derivative works across different substrates [169, 177, 254, 312–318].

However, when dealing with novel candidate spaces, many expensive calculations still need to be performed for meaningful featurization, before even starting the actual scope campaign. Additionally, this workflow neglects the influence of (remote) structural features, such as functional groups. To address these limitations and democratize unbiased evaluation, a recent work

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

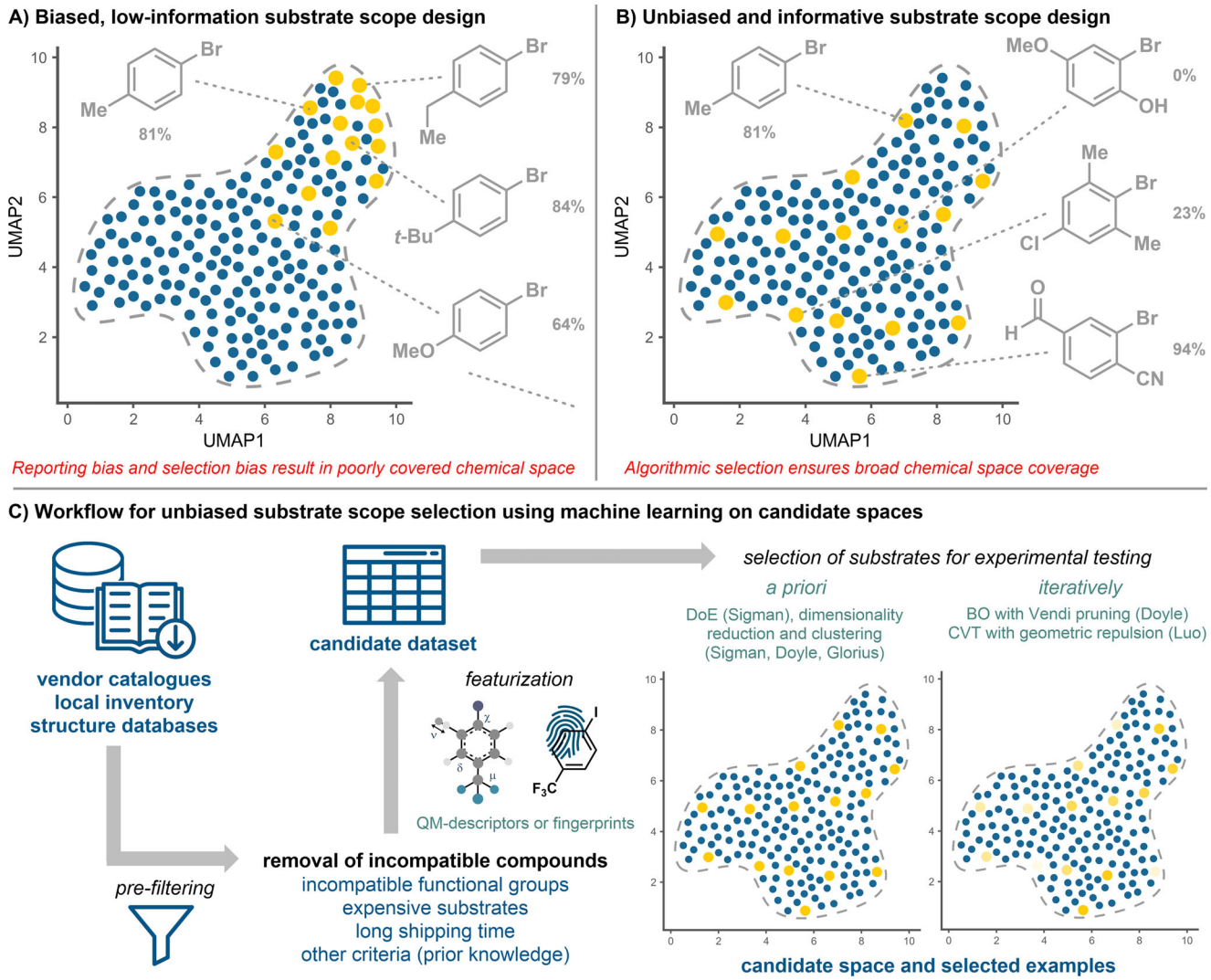


FIGURE 11 | ML guided substrate scope studies. (A) Conventional, human-led substrate scope selection often possesses strong biases, leading to low information contents. (B) Unbiased substrate scope selection leads to broader chemical space coverage, often including more negative results and diverse compounds, enriching the informativeness of the study. (C) Data-driven strategies/workflow for unbiased substrate scope selection starting from large compound libraries.

by our group utilizes fingerprint representations to encode molecular structures. Potential substrates are then projected onto a universal map of "drug chemical space" generated by applying UMAP to the DrugBank database, followed by a clustering algorithm. While this approach does not capture physically grounded and interpretable structure–reactivity relationships [319], it maximizes the overall structural diversity of the scope while also mitigating human biases. In addition, the absence of expensive molecular descriptors allows this workflow to be easily applied by nonexperts via a corresponding web app [304].

Taken together, all these methods rely on an effective mapping of the entire accessible chemical space of substrate candidates [320], that needs to be performed before sampling a fixed amount of substrates. While this approach offers the opportunity to facilitate and drastically accelerate a scope study employing parallelization and HTE, it lacks the flexibility to adapt or extend the scope study on the fly, for example, when scope experiments yield novel insights about the reaction.

4.3.2 | Iterative Exploration of a Substrate Scope

More recently, data-driven approaches have emerged that view scope evaluation as a dynamic, iterative process. In this paradigm, the outcome of an experiment informs the selection of the next, allowing the chemist to navigate chemical space much like it is done in traditional scope studies. However, rather than simply evaluating reactions on a few hand-picked substrates, these algorithms can be used to iteratively guide the campaign toward greater substrate diversity and boundary detection. In this regard, active learning has emerged as a promising approach for building a general and predictive model of the reaction using a minimal number of experimental data points. In contrast to optimization algorithms such as pure BO, which maximize an objective, the active learning algorithm queries datapoints to iteratively reduce its internal ML model's uncertainty until it achieves high predictive accuracy, enabling the identification of both well-performing substrates and reactivity boundaries [321, 322]. In a recent study by Novartis, this approach significantly outperformed random

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

selection of substrates from the candidate space, ultimately leading to a predictive scope model for a Ni/photoredox cross-electrophile coupling (XEC) [323]. Along the same line, Doyle and coworkers have integrated diversity metrics directly into the selection loop of a BO algorithm. By utilizing a Vendi score pruning step, the adapted BO algorithm penalizes structural similarity, effectively deprioritizing highly similar substrates in the candidate spaces and forcing the BO to optimize for high-yielding, but still diverse substrates. Based on the selected and experimentally evaluated datapoints, the authors propose a scope score—a composite metric of yield and diversity—that offers a more objective standard for success than average yield alone [324]. While both methodologies still require the manual elimination of incompatible substrates in the candidate space a priori, a recent study by Luo and colleagues introduces an innovative human-in-the-loop approach in which incompatible substrates can be flagged by chemists on the fly. Notably, the employed Centroidal Voronoi Tessellation (CVT) algorithm combined with a dynamic geometric repulsion mechanism transforms negative experimental feedback into structural constraints, actively steering the sampling trajectory toward unexplored areas of the substrate space. Therefore, this iterative approach aims to reconstruct the reaction's scope topology and create a comprehensive ScopeMap rather than a list of performance results [325].

4.3.3 | Ultra-Large Reaction Scopes, Combinatorial Datasets, and General Reactivity Models

Last, with the emerging use of HTE in synthetic chemistry, the collection of ultra-large reaction datasets facilitates large-scale scope studies [84, 326–328]. In particular, while typically only one reactant is varied in traditional scope studies and co-reactants are kept constant, HTE-based scope studies enable the creation of combinatorial datasets. By considering orthogonal co-reactants, reactivity cross-relationships and interaction effects can be effectively uncovered and also modeled (see Sections 2.3 and 3.2) [79, 160, 329].

Extending the concept of considering more influencing factors even further, for example, by including different reaction conditions [330], boundaries between general reaction optimization and actual scope campaigns vanish. In particular, the objective of obtaining a reaction scope model for a single (optimized) reaction condition shifts toward a general reactivity model capable of predicting yields and selectivities across a very large substrate and condition space, which can be considered the holy grail of data-driven synthetic chemistry. As the complexity of these modeling tasks scales exponentially with the number of factors considered, even traditional HTE scales might not be sufficient. Just like in single-substrate scope studies and discussed for the development of yield prediction models, careful sampling of substrates (and conditions) is crucial to cover a meaningful reaction space (see Section 3.3) [163, 331].

4.4 | Mechanistic Interrogation Assisted by Data Science

With optimization and scope studies completed, a novel chemical reaction is generally considered ready for synthetic applications. However, investigating, analyzing and understanding the underlying reaction mechanism holds immense scientific value, as it allows chemists to gain fundamental knowledge that applies to the wider molecular world. In mechanistic studies, data science has emerged as a tool complementing traditional mechanistic experiments such as spectroscopy, mass spectrometry and isotope experiments, among many more [240]. However, it is important to consider that data science, by definition, investigates correlations within (potentially highly complex) data. In contrast, the central goal of mechanistic studies is to uncover causality, physics-grounded reasons for molecular behavior. Therefore, the interpretation of results from data-scientific techniques must be carried out with extreme caution, and the tools should be regarded as assistance to deal with large datasets in mechanistic studies. The following section outlines common conceptual approaches of how ML tools might assist mechanistic studies. For a more detailed discussion of this topic, the reader is referred to additional comprehensive reviews [172, 332–334].

4.4.1 | Mechanistic Knowledge Extracted From Reactivity Models

A fundamental approach to extracting interpretable trends from reactivity datasets is the utilization of physically informed features in combination with simple ML architectures such as multilinear regression or tree-based models. Subsequent extraction of feature importance reveals direct correlations between the chosen descriptors and observed reactivity [45]. Pioneered and widely employed by the Sigman group, this approach effectively expands upon the historic concept of linear free energy relationships (LFERs) [335], which investigates the (linear) dependence of experimentally observed reactivities and standardized parameters such as the widely used Hammett parameters σ [336]. Conceptually aligning with a more digital era, the replacement of standardized, fixed LFER parameters with molecular descriptors derived from semiempirical or QM calculations allows for tailoring parameters to the investigated reaction. Furthermore, multilinear models replacing the unilinear ones increased the complexity of the models, while still ensuring interpretability through the coefficients of the features [172, 212, 337].

This approach can be easily implemented by following the workflow for building simple selectivity or yield prediction models (see Section 3.2), followed by extracting feature influences through either MLR coefficients, read-out of decision tree nodes or feature importance analyses, for example, by SHapley Additive exPlanations (SHAP) [338], for example, when dealing with deep learning models (Figure 12A) [339]. Note that while this workflow is highly related to building predictive reactivity models, the important features are typically not known a priori and are derived from a model that is fitted to experimental data.

Having the most important parameters extracted from the model, mechanistic insights can be derived by analyzing their influence within the context of a proposed mechanistic pathway. For example, a positive correlation coefficient of a ligand's Sterimol [340] B_1 descriptor with $\Delta\Delta G^{\ddagger}$ for enantioselectivity can align with a mechanistic hypothesis where steric repulsion is the underlying reason for enantio-induction at a metal complex.

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

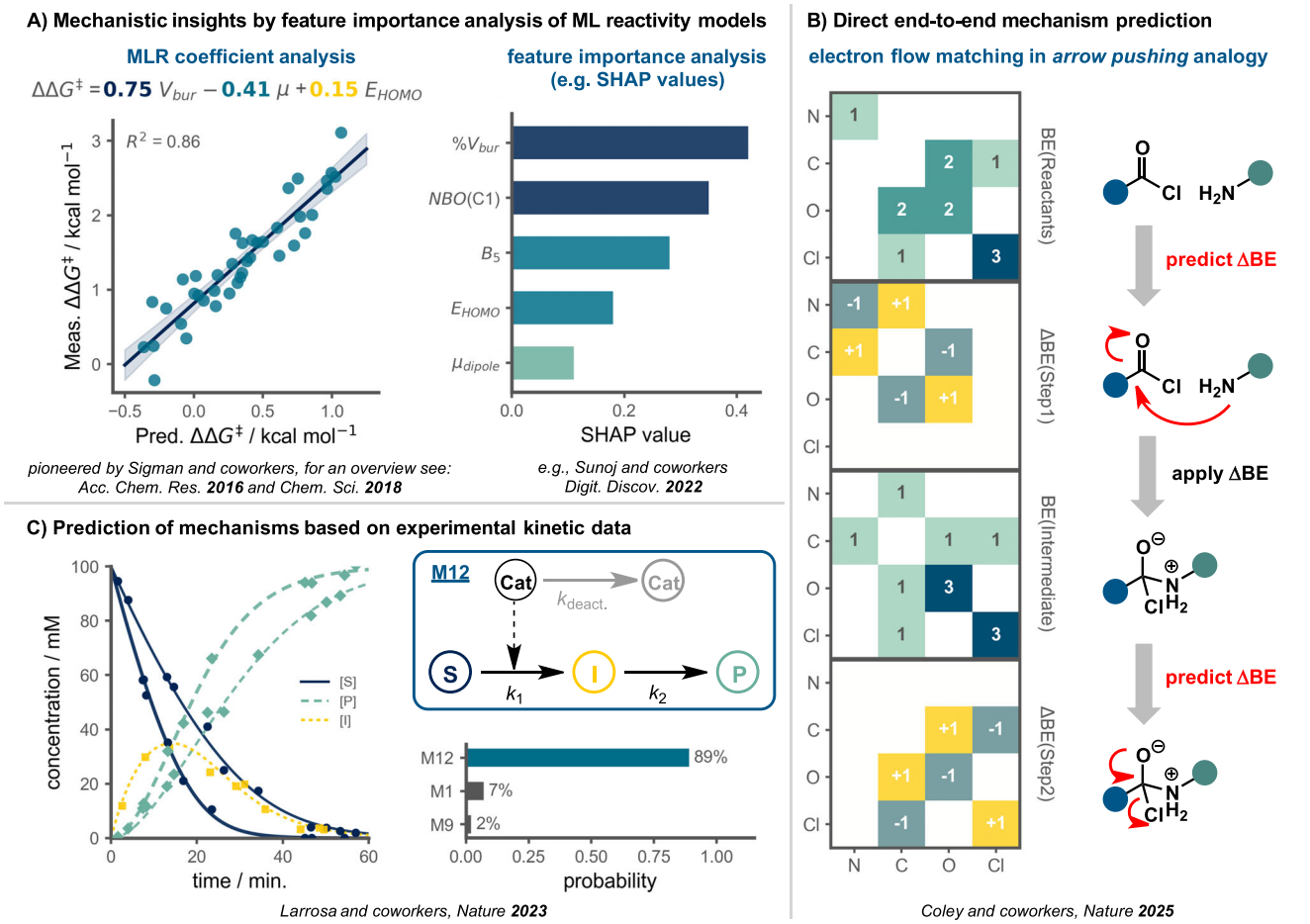


FIGURE 12 | Different approaches of applying ML for mechanistic studies. (A) Extracting and interpreting feature influences of MLR models (left) or black-box models (using SHAP, right) in the context of mechanistic proposals. (B) Step-by-step generative modeling of "arrow pushing" mechanisms by predicting the electron flow and intermediates along a reaction coordinate. (C) Direct classification of mechanisms by classifying experimental kinetic profiles.

Notably, the number of input parameters in these simple models can vary and should be carefully chosen to balance meaningful relationships and overfitting. While in many applications, three to five input variables are investigated, drastic effects have been previously uncovered using this workflow, even when considering only a single parameter. In an illustrative example, Doyle and coworkers studied reactivity cliffs in Pd and Ni-catalyzed cross-coupling reactions. For this purpose, the authors built univariate classification models based on experimentally observed reactivity and DFT-based ligand descriptors. By univariate analysis, the authors identified that the minimal possible buried volume $\%V_{bur}$(min) of a conformer ensemble of the ligated metal complex is highly predictive for binary reactivity classification. Based on these data-driven insights, subsequent studies revealed that $\%V_{bur}$(min) determines the ligation state of the involved catalytic species, ultimately giving a causal mechanistic explanation for reactivity [341].

The same workflow can disentangle overlapping electronic and steric effects in C—H functionalization and even reassign the selectivity-determining step. To probe the site selectivity of TBADT-catalyzed C(sp^3)–H Giese additions to olefins, Tekle-Smith and coworkers represented each C—H position in a literature-curated dataset of 39 substrates and 232 total sites by DFT descriptors of both elementary steps (H-atom abstraction and Giese-type addition) and trained an interpretable logistic regression model on these data. Feature-importance analysis revealed that radical descriptors, SOMO energy and radical philicity were more informative than hydrogen polarity descriptors, indicating that selectivity is governed less by the hydrogen abstraction step than by the subsequent radical addition. These data-driven findings were then validated by mechanistic experiments. Isotopic labeling established that TBADT samples a broad set of electron-rich C—H bonds during abstraction, whereas kinetic studies showed that the rate of the subsequent radical addition scales with the radical's SOMO energy. Selectivity is thus expressed predominantly at the carbon–carbon bond-forming step under frontier-orbital control, demonstrating that decatungstate HAT-photocatalysis follows the long-standing radical sampling principles of Giese-type additions [342].

4.4.2 | Acceleration of Quantum-Chemical Studies of Reaction Coordinates

In addition to experimental mechanistic studies, computational chemistry remains a central part of analyzing reaction mechanisms and coordinates in silico. Even though modern DFT and

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

semiempirical methods have already reduced the computational cost associated with this task significantly, the computational analysis of multiple reaction pathways or even reaction networks remains time- and resource-intensive. In this context, ML promises advances in mainly two areas:

First, through general ML interatomic potential (MLIP) foundation models like Meta's universal models for atoms (UMA) family [343, 344], geometry optimizations and energy evaluations can be drastically accelerated. Contrary to previous MLIPs [345, 346], the UMA models were trained on a massive dataset of ~100 million DFT calculations with varying spin states and charges, rendering it an almost universally applicable model for small molecules while showing DFT accuracy. The MLIP-enabled acceleration of electronic structure "calculations", in turn, increases calculation throughput and enables faster identification and analysis of substrates, products, intermediates, and transition states [347, 348].

Second, specialized generative models have emerged as tools for directly generating transition states or even reaction coordinates de novo, requiring only the substrate and/or product structures as inputs [349–353]. However, applicability domains of the employed ML models must be carefully monitored [349, 354]. Also, slight technical changes of a model (e.g., altered hyperparameters, new training datasets, different initializations, etc.) render their output somewhat indeterministic. Therefore, whenever dealing with a mechanistic study, structures and stationary points obtained by ML models should always be validated and confirmed by traditional QM methods.

4.4.3 | Generative Modeling of Entire Reaction Mechanisms

By combining data from all mechanistic experiments and computational studies, the ultimate goal of mechanistic studies is to propose a full reaction pathway, including all intermediates and the flow of electrons, as typically depicted in "arrow-pushing" schemes. While these schemes are often oversimplified, they strictly ensure the conservation of mass and electrons, a central and fundamental law of chemical reactions, which is, however, not at all guaranteed by many product-prediction models. Recent strategies have addressed these shortcomings by reframing reaction prediction as an electron-redistribution problem. For example, a recent study by Coley and coworkers introduced FlowER, a graph transformer that employs flow matching to predict (changes in) bond-electron (BE) matrices (Figure 12B). Trained on 1.4 million elementary steps across polar and radical reaction classes, it matches or exceeds state-of-the-art accuracy while ensuring atom and electron conservation. Furthermore, minimal fine-tuning also enables extension to unseen elementary steps [355].

By grounding predictions in the movement of electron pairs, this approach, as well as related flow-matching- [356], text- [357], or graph-based [358–361] approaches resolve the hallucination errors of direct product prediction models while mirroring the intuitive arrow-pushing approach. However, it is important to note that these models are also strictly limited by the (often simple and hand-curated) data they were trained on [362]. Therefore, carefully monitoring the applicability domain is important. For example, models trained solely on polar mechanisms are not expected to produce meaningful predictions for radical- or transition-metal-involving mechanisms.

4.4.4 | ML-Assisted Analysis of Analytical and Kinetic Data

In addition to direct mechanistic modeling, ML models have been employed for the analysis and interpretation of experimental analytical data, which can in turn reveal mechanistic insights. As an example, Liu and coworkers trained deep learning classifiers on simulated cyclic voltammograms (CVs). Following training, they used this model to determine the mechanisms of observed electrochemical events in real measurements, for example, pure electron-transfer events, disproportionation events or chemically coupled electron transfers [363, 364]. In a related study, Leonard and coworkers employed convolutional neural networks (CNNs), also directly on CV curves, to drastically speed up the estimation of standard electrochemical rate constants (k^0) and transfer coefficients (α) directly from the experimental data [365].

Furthermore, Larrosa and coworkers reported that ML models can guide the elucidation of a reaction mechanism solely based on kinetic data (Figure 12C). Specifically, the authors trained deep long short-term memory (LSTM) networks on 5 million simulated kinetic profiles (time–concentration series) generated by numerically solving sets of ordinary differential equations for 20 commonly encountered catalytic mechanisms. The training set included simple core Michaelis–Menten cycles as well as complex pathways involving catalyst activation, deactivation, and hybrid-catalytic steps. The model subsequently categorized unseen time-concentration series into their underlying mechanistic blueprint, achieving a categorical accuracy of 92.6%. The system maintains high accuracy even with substantial experimental noise (up to 5% Gaussian error) and has been successfully applied to real-world experimental data for reactions like ring-closing metathesis and C—H aminations [366]. Therefore, by directly operating on the time-concentration curves, the challenging and error-prone determination of initial rates and rate laws can be circumvented, significantly speeding up the kinetic analysis [367, 368].

In a complementary study, ML has been coupled to the high-throughput generation of kinetic data itself. Gomes and coworkers recently reported on their PRISM platform to overcome the low throughput and poor reproducibility of conventional rate measurements. Using an inexpensive, 3D-printed setup that tracks pH-induced indicator-dye color changes with cameras on 96-well plates, they assembled a standardized dataset of 1162 second-order rate constants for the amide coupling of acyl chlorides and amines, spanning more than four orders of magnitude and cross-validated against ^{19}F-NMR. Reactants and intermediates were represented through feature vector embeddings extracted from the pretrained AIMNet2 foundation model. The compiled dataset was then used to train two models: a classifier distinguishing experimentally measurable from immeasurably fast rates, and a graph neural network predicting the quantitative rate constant. The classifier reached 96% to 99% accuracy on out-of-distribution reactant combinations, while the regression model predicted rates for unseen amines with a test

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

MAE of 0.52 log(k) units, demonstrating that cheaply generated, internally consistent kinetic datasets can directly train predictive rate models [94].

5 | Synthetic Chemistry in the Era of Large Language Models

The classical ML modeling techniques presented so far are typically constrained to narrowly defined tasks, such as the ones discussed in Sections 3 and 4. Consequently, they excel at making predictions strictly within their specific training domains. Challenging these traditional model types, rapidly emerging LLMs are deep learning architectures trained on vast amounts of text, including books, scientific literature, and websites. Accompanying this enormous scale in training data is a fundamental shift from specialized predictions to universal text processing, encompassing tasks such as summarization, translation, and generative text creation. For this reason, LLMs are also considered foundational models for language, capable of performing a wide variety of natural-language tasks. Since the public release of conversational interfaces such as OpenAI's ChatGPT 3.5, this kind of AI has experienced unprecedented growth and adoption, becoming a ubiquitous tool in both everyday life and scientific research [369].

5.1 | Introduction to Large Language Models and AI Agents

To contextualize their application within the synthetic chemistry community, several basic concepts are briefly introduced below. At their core, LLMs generate text by taking an input and repeatedly predicting the most statistically probable next word or sequence. To achieve this, text is broken down into tokens, which serve as basic vocabulary units for words, subwords, and punctuation. Tokens also serve as the standard computational and financial billing unit for model usage. Crucially, this means LLMs do not "think" or understand chemistry in the classical human sense; rather, they rely on learned probability distributions to determine which tokens are most likely to follow in a given context. Recent advancements have also introduced "thinking" or reasoning capabilities: Before delivering a final response, these models generate an internal "chain-of-thought"—a self-dialogue that allows the LLM to analyze the given task step-by-step, draw intermediate conclusions, and self-correct. This is particularly valuable for complex logical, mathematical, or programmatic problems (as encountered in chemical settings as well), though it is important to note that this reasoning is fundamentally still based on statistical text prediction. Furthermore, since LLMs are general-purpose models for many everyday use cases, obtaining useful, chemistry-specific outputs often requires prompt engineering. This technique involves systematically designing the input text to guide the model toward a desired format or logical structure, avoiding the need to fine-tune the model on task-related or domain-specific text corpora. For an in-depth discussion of the underlying neural architectures, we refer readers to dedicated technical reviews [370–377]. Last, focusing on applicability, one must primarily distinguish between two types of LLM distribution strategies. Open weights models provide public access to their underlying parameters via platforms like Hugging Face. Given adequate computational hardware, researchers can deploy and fine-tune these architectures locally. Conversely, frontier models represent the most advanced systems developed by major technology companies. These remain strictly proprietary and are accessible solely through commercial APIs or conversational interfaces.

A significant advancement in AI is the transition of LLMs from static conversational models to autonomous agents. Standard LLMs are constrained by their static knowledge embedded within their training data. If asked a factual question about an event that occurred after their training cutoff, they can only guess based on historical probabilities. In contrast, agentic systems overcome this limitation by interacting with virtual or physical environments, reasoning through potential actions, and executing decisions iteratively [372]. This autonomy is achieved by embedding the LLM within an orchestration framework comprising planning, memory, tools, and acquired skills (Figure 13). Planning allows the model to break down a complex, high-level goal into a logical sequence of actionable steps. Memory modules retain past observations and failed attempts, facilitating iterative self-correction and the accumulation of procedural memory, which manifests as learned skills over time. Skills represent the agent's refined capacity to consistently execute specific procedural workflows or accurately utilize external modules without explicit step-by-step guidance. Tools are specialized external modules that extend the model's capabilities beyond text generation. Precisely, tools are defined as domain-specific scientific packages, remote APIs, and laboratory automation interfaces that accept structured inputs to execute targeted operations and return actionable data [378].

The integration of such functionalities is exemplified by ecosystems such as ToolUniverse, which provides a standardized interaction protocol for AI models to autonomously identify, execute, and compose over 600 scientific tools into complex research workflows [379]. In synthetic organic chemistry, chemical agents leverage these ecosystems to autonomously execute analytical scripts, query structural databases, and direct physical robotic hardware for experimentation [372].

Having established these foundational concepts, we will examine selected examples of pure LLMs (Section 5.2) and agentic AI (Section 5.3) systems applied to synthetic chemistry. Rather than providing an exhaustive catalog of this rapidly moving field—for which we refer the reader to dedicated specialized reviews [370–377]—we will highlight representative examples across different domains of synthetic chemistry, evaluating their specific methodologies and practical utility [370–377].

5.2 | Application Examples of Large Language Models

As foundational models for language, LLMs are well-suited to extracting experimental data from scientific literature into standardized formats. Chemical research is published at increasing volumes, yet standardized and deposited data remain scarce (see Section 2.3). This data, however, is needed to train predictive models and to build the knowledge bases that self-driving laboratories rely on [380–386]. The vast majority of historical knowledge of organic synthesis remains locked in the PDF files

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

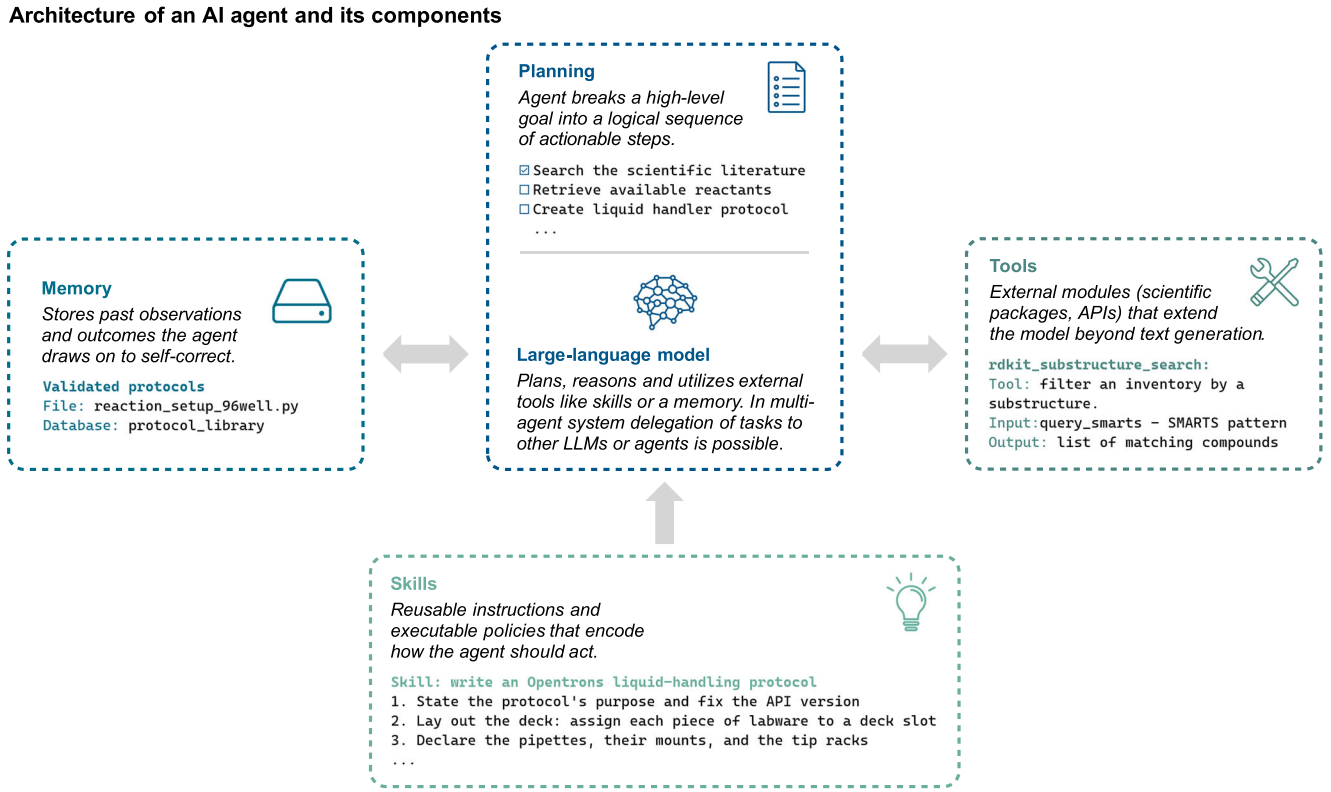


FIGURE 13 | Anatomy of an LLM-based agent for autonomous chemistry. A central large-language model plans and reasons over a high-level goal, decomposing it into actionable steps. It can draw on memory (past observations and artifacts), invoke external tools (scientific packages and APIs), and follow skills (reusable instructions encoding how to act) to carry out tasks.

of various journals [380]. The information is fragmented across unstructured prose, optimization tables, complex figures, and supplementary information. Traditional rule-based algorithms struggle to digitize these documents because the relevant data is scattered across both visual elements and their associated textual contexts (e.g., footnotes and captions) within highly diverse table and graphical layouts.

To address this challenge, Aspuru-Guzik and coworkers introduced MERMaid (multimodal aid for reaction mining). The pipeline processes diverse scientific literature, including main text and unstructured supplementary materials, including papers dating back to 1949, specifically targeting graphical elements alongside their associated textual descriptors [387]. Ultimately, MERMaid translates this visual and textual data into structured, machine-readable formats, converting molecules into SMILES strings and reaction conditions into standardized JSON dictionaries. MERMaid achieves this through a modular pipeline that pairs local deep-learning tools with proprietary vision-language models (VLMs). Unlike traditional text-only LLMs, VLMs are multimodal models natively trained on both text and images, allowing them to perform visual reasoning. However, several limitations remain for practical deployment. Applying the pipeline to entirely new chemical domains requires researchers to manually define novel parameter keys via JSON configuration files and explicitly program structural updates to the underlying static knowledge graph schema. Because the pipeline relies on commercial, closed-weight frontier models via APIs, users face recurring token costs and lack control over backend model updates or deprecations that could alter reproducibility. Additionally, molecular structure recognition and translation into SMILES are currently offloaded to a dedicated external toolkit (RxnScribe) and are restricted to basic chemical structures, struggling with complex Markush structures [388].

As established in Section 4.2, reaction optimization campaigns increasingly rely on data science and ML. However, sophisticated optimizers, including BO, struggle within sparse, irregular, and high-dimensional search spaces. A fundamental limitation of standard BO is its lack of inherent domain knowledge. Without expert-selected molecular descriptors—which require deep chemical knowledge about the system and computational effort to generate—BO must rely on basic one-hot encoding for categorical variables. For instance, lacking a descriptor that explicitly defines the pK_a value of various bases, the surrogate model cannot extrapolate base strength and must blindly evaluate each base individually [389, 390]. Furthermore, in the absence of initial hits, BO algorithms typically default to random or Latin Hypercube sampling during early phases. This results in slow initialization ("cold starts") and frequent entrapment in local minima. These issues are pronounced when optimizing complex, non-convex landscapes, such as multicomponent mixtures, where most parameter combinations yield no product (see Section 4.2; troubleshooting).

To address this challenge, Cooper and coworkers introduced a hybrid optimization framework that integrates knowledge-based insights from LLMs to contextualize BO [391]. The Bayesian

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Optimization Research Assistant (BORA) was evaluated on a 10-dimensional photocatalytic hydrogen evolution problem characterized by a massive search space (10 dimensions, 98 million combinations) and sparse high-yielding conditions. Performance metrics show that the hybrid optimization framework significantly outperforms standard BO, specifically during early-stage exploration where initial hits are rare. Using reasoning and hypothesis-driven exploration, the LLM recognizes that an entire class of additives is detrimental, therefore restricting the sampling space, an efficiency standard that algorithms cannot achieve without explicit prior categorization. BORA dictates the search trajectory via an adaptive heuristic policy linked to gaussian process uncertainty. The system transitions between three action modes: standard optimization when uncertainty is low, language model filtering of algorithmic suggestions, and pure language model hypothesis generation when high uncertainty signals optimization stagnation. While the global boundaries of the discrete search space remain fixed by an initial user prompt defining parameter limits and objectives, the active search space is continuously adapted by the model. This adaptability relies on in-context learning, a mechanism where the model updates its operational logic and chemical hypotheses directly from the accumulated sequence of experimental results in the prompt, entirely without requiring structural retraining of its neural network weights [389, 392]. Consequently, the system yields structured experimental proposals accompanied by clear rationales and confidence scores to escape performance plateaus.

In a related study, Gomes and coworkers demonstrated a fully LLM-guided optimization (LLM-GO) pipeline that eliminates mathematical optimization frameworks entirely. Rather than merely replacing the gaussian process surrogate, the algorithm operates strictly through prompt-based in-context learning. In each iterative round, the full categorical search space and all historical experimental results are fed into the LLM, relying solely on pretrained knowledge to propose subsequent reaction conditions. To evaluate this approach, the researchers benchmarked the pipeline across six fully enumerated experimental datasets containing 768 to 5684 reactions, addressing chemical problems such as Chan–Lam, Suzuki–Miyaura, and Buchwald–Hartwig cross-couplings [393].

Performance metrics from both the Cooper and Gomes groups reveal that LLM-driven approaches consistently match or outperform standard BO in single-objective, high-dimensional categorical optimizations [394]. Because foundation models contain embedded chemical understanding, such as recognizing specific base and ligand compatibilities, the LLM effectively warm starts the optimization. This mechanism bypasses unpromising chemical spaces that standard BO would waste resources exploring blindly. Gomes and coworkers quantified this advantage within these cross-coupling datasets, revealing that LLMs maintain a systematically higher exploratory diversity than BO. This diversity proves highly advantageous when optimal conditions are extremely scarce and represent less than 5% of the total search space.

However, critical limitations remain. The stochastic nature of LLMs occasionally leads to outlier campaigns where the model fixates on unprofitable hypotheses or loses experimental memory, resulting in the suggestion of duplicate experiments. Furthermore, LLMs struggle to appropriately weight conflicting replicate measurements and navigate explicit multi-objective trade-offs compared to the exact mathematical uncertainty quantification of BO. A profound limitation arises from the reliance on pretrained knowledge, which heavily biases exploration toward historical literature data. Performance significantly decreases when the search space contains unknown compounds, such as custom-synthesized catalysts or proprietary ligands. In these sparse or novel domains, the LLM cannot draw upon existing literature associations [393, 394].

Therefore, when navigating highly dimensional categorical problems comprising established literature components, such as screening discrete commercial catalysts, ligands, and solvents, LLMs offer a distinct advantage by circumventing the need for complex manual descriptor engineering. Conversely, traditional BO remains the strictly superior choice for optimizing continuous parameters like temperature and concentration, navigating multi-objective trade-offs, or exploring novel reactions and custom compounds where no pretrained knowledge is available, and exact mathematical uncertainty quantification is explicitly required.

5.3 | Application Examples of Agentic AI Systems

The preceding examples demonstrated that language models are employed primarily as text processors and in-context predictors, guided by advanced prompt engineering. However, the greatest potential to accelerate chemical research lies in deploying agentic models. By equipping language models with external tools, memory architectures, and the ability to autonomously plan and execute multistep processes, passive models are transformed into active scientific agents. The immediate practical application of this concept for the experimental chemist involves deploying tool-augmented systems to interact directly with external software, APIs, and physical laboratory hardware, prior to exploring more abstract, multi-agent computational frameworks [372]. A central challenge in this field is that laboratory environments are highly heterogeneous. Experimental protocols must be mapped onto a diverse array of instruments, such as liquid handlers, flow reactors, and analytical devices, many of which lack standardized programmatic interfaces. Furthermore, successful physical execution relies heavily on implicit, experiential knowledge, including proper labware placement, phase separation techniques, or handling heterogeneous mixtures, which is inherently difficult to codify into traditional algorithms [289, 395]. Several pioneering platforms have recently demonstrated the viability of agent-driven synthesis by integrating language models with external toolkits to directly orchestrate laboratory hardware and translate high-level natural-language goals into structured robotic execution commands [370, 396].

White, Schwaller, and coworkers introduced ChemCrow, an agent augmented with 18 expert-designed tools bridging language model reasoning with chemical informatics. The framework plans and executes the synthesis of target molecules, directly addressing the gap between digital planning and physical robotic execution. By querying specialized tools like chemical property calculators, patent checkers, and safety screeners, the agent drastically mitigates chemical hallucinations. The system

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

autonomously planned and executed the synthesis of an insect repellent and three organocatalysts on a cloud-connected robotic platform [397].

Concurrently, Gomes and coworkers introduced Coscientist, a multi-module system that controls liquid-handling robots. The system automates complex synthesis and optimization workflows, validated on palladium-catalyzed Suzuki–Miyaura and Sonogashira cross-couplings. Optimization capabilities were benchmarked using previously generated reaction condition spaces, specifically a flow-based Suzuki–Miyaura coupling dataset and a carbon-nitrogen cross-coupling dataset. Coscientist uses tools such as web search, document retrieval, and isolated code execution to control a pipetting robot. The agent is also capable of correcting hardware execution errors dynamically by consulting technical instrument documentation [398].

While foundational agent architectures like ChemCrow and Coscientist operate as monolithic, single-agent systems routing queries to multiple external computational or hardware tools, a distinct architectural shift has emerged toward multi-agent systems. In a multi-agent framework, a complex scientific objective is decentralized across an ensemble of discrete, specialized agents that possess distinct behavioral profiles, role assignments, or domain boundaries. These independent agents communicate dynamically through predefined collaboration pipelines, group chats, or consensus protocols to simulate a distributed research team. By distributing cognitive subtasks—such as safety auditing, experimental design, and analytical interpretation—across localized specialists, multi-agent frameworks minimize context window saturation and mitigate single-point-of-failure hallucinations inherent to monolithic structures [372]. Along these lines, Mo and coworkers recently introduced LLM-RDF, a pioneering multi-agent framework utilizing six specialized, interacting agents to manage the entire reaction development lifecycle autonomously from initial literature scouting to product purification. High-throughput screening data was autonomously generated via integrated robotic platforms coupled with simultaneous GC/MS analysis. The system further transitions from initial literature scouting to real-time spectral analysis (extracting yields from GC–MS and NMR data) and provides scale-up purification instructions. This framework lowers the barrier to deploying built-in Bayesian optimization and high-throughput workflows by translating natural language into executable hardware instructions without requiring the chemist to possess advanced coding skills [396].

These systems represent a leap in laboratory automation and an unprecedented level of autonomy in synthetic chemistry. Nevertheless, these agentic frameworks are frequently evaluated using classic, well-documented reactions (e.g., Suzuki–Miyaura couplings or the synthesis of paracetamol). It is therefore yet to be demonstrated whether the systems actually demonstrate independent chemical reasoning or merely benefit from data leakage by reproducing known procedures memorized during training. Furthermore, while these agents succeed in execution on standardized robotics, they struggle with fundamental mechanistic interpretation. For example, while automated systems can accurately fit kinetic models to experimental time-course data, they lack the deep domain knowledge required to independently infer underlying catalytic cycles or complex inhibition mechanisms without human oversight. Optimization modules also frequently rely on existing mapped datasets rather than generating original kinetic insights de novo [396]. Crucially, as a general note, empowering nonexperts with autonomous synthesis capabilities introduces safety and dual-use risks, including the potential for generating hazardous or restricted compounds. Recent agent architectures actively mitigate these risks by incorporating strict programmatic guardrails, such as cross-referencing proposed targets against established chemical weapons precursor databases and hazardous materials lists before execution [399]. Despite these automated safety checks, fully autonomous and unsupervised physical execution presents liabilities. Therefore, chemical agents should function as collaborative systems requiring explicit human-in-the-loop approval before any physical action is triggered.

Another real-world use case of multi-agentic workflows has recently been shown in the area of in silico research. As synthetic chemists frequently struggle with the steep entry barriers of computational chemistry workflows, agentic LLMs hold promise for making these tools more accessible. The in silico analysis of molecular properties and mechanistic pathways via DFT or molecular dynamics requires highly specialized expertise. Formulating input scripts, selecting appropriate methods and basis sets, converting file formats, and scheduling high-performance computing jobs constitute an iterative loop where researchers must continuously adjust parameters to resolve failed runs or convergence issues [400].

To navigate these computational challenges autonomously, Aspuru-Guzik and coworkers introduced El Agente Q, a multi-agent framework that dynamically translates natural language prompts into executable quantum chemistry workflows. Rather than merely generating text, the system directly interfaces with established computational packages such as ORCA [401], xTB [47, 48], and RDKit. The framework employs a hierarchical network of specialized agents: a top-level computational chemistry agent delegates tasks to mid-level modules tailored for geometry manipulation, quantum calculation execution, and file handling.

The system was evaluated across a diverse dataset of university-level computational chemistry exercises and complex case studies. These tasks required the agents to compute the relative thermodynamic stabilities of carbocations, determine cycloalkane ring strain, calculate the pK_a of carboxylic acids, and predict vertical excitation energies. For advanced applications, the agents successfully modeled the effects of implicit and explicit solvation on the infrared spectrum of alanine and executed a multistep structural optimization workflow for a multi-ligand lanthanoid complex. A critical performance feature of this hierarchical structure is the capacity for automated, in situ error recovery. If a calculation fails due to an imaginary frequency or a convergence error, the lower-level execution agent parses the output file, identifies the specific failure, and autonomously adjusts the geometry or keywords before resubmitting the job [402]. To cover a broader spectrum of scientific domains, this agentic architecture has since been expanded into a comprehensive family of specialized agents, including El Agente Sólido (solid-state materials) [403], El Agente Estructural (multimodal 3D geometric manipulations) [404], El Agente Cuántico (quantum computing) [405], as well as Gráfico [406], Quntur [407], and Forjador [408].

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

While this democratization of computational tools is highly advantageous, it introduces novel adoption risks. Nonexperts may execute complex quantum-mechanical calculations without a fundamental understanding of the underlying approximations. If the model selects an inappropriate level of theory for a specific molecular system, the calculation may execute flawlessly while yielding physically meaningless results. This scenario represents a higher-order hallucination where the computational syntax is correct but the chemical physics are fundamentally flawed. Additionally, operating these continuous, autonomous loops using proprietary frontier models consumes substantial computational resources and generates significant API costs.

For the experimental organic chemist, platforms like El Agente Q serve as highly capable virtual collaborators for rapidly computing thermodynamic variables, predicting spectroscopic signatures, and validating mechanistic hypotheses prior to laboratory synthesis. However, practitioners must treat the output with rigorous skepticism. Experimentalists must systematically review the system's exported action traces, which are provided as transparent Python scripts, to ensure the chosen basis sets, functional combinations, and solvation models are chemically appropriate for the target system. Furthermore, journals and reviewers must require the disclosure of these autonomous action traces alongside traditional computational supplementary information to ensure reproducibility and scientific integrity.

5.4 | Critical Considerations on Large Language Model Usage

Despite their impressive capabilities, there are profound drawbacks and critical limitations to the use of LLMs and AI agents in synthetic chemistry. Most significant is the generation of grammatically flawless and highly convincing text despite it containing factually incorrect, harmful, or chemically illogical information [370]. Because LLMs predict the statistical likelihood of words rather than verifying factual truth, these so-called hallucinations remain a persistent risk. Recent theoretical work even suggests that such hallucinations are mathematically inevitable in current LLM architectures [409]. Therefore, deploying these models in synthetic chemical environments poses significant safety and reliability risks. An agent itself might reduce the risks of operating in a laboratory by acting as a planning-stage guardrail, screening protocols for chemical hazards, reaction risks and dual-use concerns before any physical execution occurs [399]. Such pre-execution screening narrows the operational risk that an autonomous platform carries into the laboratory, yet it does not eliminate the hazards introduced once an agent controls physical apparatus. Rigorous safety checks and human supervision, therefore, remain necessary, and physical safeguards such as restricting the agent's action space to validated safe operations may be required before autonomous laboratories can run without direct oversight. It is imperative that LLM outputs are carefully monitored, such as via a human-in-the-loop approach, and that rigorous safeguards are established [410]. Furthermore, because LLMs are statistical predictors, identical inputs may yield varying outputs. The continuous updating or deprecation of proprietary models contributes to this reproducibility issue. This shift in technology severely impedes other researchers' ability to reproduce scientific findings, as published prompts may yield different results on newer models [389].

Beyond reproducibility, genuine chemical discovery is limited by the information contained within their training data: recent comprehensive evaluations, such as the ChemBench framework introduced by Jablonka and coworkers, demonstrate that model performance correlates strongly with proximity to training data rather than true reasoning about molecular complexity or topology. This limitation is compounded by their propensity for extreme overconfidence; models frequently provide factually incorrect answers with absolute certainty and lack the internal mechanisms required to estimate their own limitations [411].

Finally, while frontier models currently offer the most advanced capabilities, their training and inference are computationally expensive and energy-intensive, raising valid concerns regarding their environmental footprint. While comparable open-weight models exist, their local deployment necessitates substantial hardware investments. This dependence on expensive compute infrastructure or costly API tokens risks rising inequality in research capabilities in well-resourced and under-resourced research institutions [394].

6 | Conclusion and Outlook

The transition of ML and data science from theoretical frameworks to laboratory practice marks a critical era in synthetic chemistry, as evidenced by pioneering examples of integrating these tools across chemistry. ML models are already at an advanced level in specific domains, particularly for predicting molecular properties such as bond dissociation enthalpies, triplet energies, and NMR shifts. In these areas, predictive accuracies rival traditional QM methods or even experimental measurements at a fraction of the cost. Furthermore, data-driven optimization strategies, spearheaded by Bayesian optimization, offer truly useful frameworks for the efficient navigation of complex chemical landscapes. Their application promises achieving better, more robust and more general reaction optimization results with fewer experiments than traditional human-led one-factor-at-a-time approaches—given the algorithms are applied carefully with expert-guided boundaries. Furthermore, powerful LLMs and agentic workflows have more recently emerged to increasingly streamline a variety of complex tasks such as literature mining, code generation, data analysis, and even the orchestration of autonomous laboratory hardware.

However, we argue that significant technical and cultural challenges remain that the community must address to fully realize the potential of ML in synthetic chemistry. Most importantly, because all AI-related technologies are powered by data, there is a pressing need to improve how we as a community generate and store experimental results. Given the high resource costs of synthetic experiments, every outcome, especially negative or low-yielding results, must be treated as a valuable datapoint. Along this line, the most informative data often also lies beyond endpoint yields. Other targets, such as reaction kinetics, robustness, and generality, can capture aspects of reactivity that a single yield number cannot and considering these dimensions may sometimes prove more useful than sheer yield dataset

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

size for building predictive, mechanistically grounded models. Regardless of the target, all generated datasets should be shared according to FAIR principles. Only with this awareness, present or future ML-technologies can leverage the true informational value of generated data. These considerations are especially important for reaction data, which is often still stored and reported in unstandardized formats today. Furthermore, addressing systemic selection and reporting biases currently skewing literature datasets will require a shift in publication culture, where editors and reviewers acknowledge the disclosure of "failed" experiments, in addition to well-performing examples, for their high informational content.

The field is moving toward increasingly complex foundation models and multi-agent systems. Even highly complex tasks that are nowadays not satisfactorily solved, such as predicting reaction outcomes or reaction conditions for a broad chemical space, will likely be successfully addressed with increasing training data in the future. This is motivated by drastic scaling effects that have already been observed for many AI applications both in everyday life (e.g., chatbots) and in chemistry (e.g., general MLIPs or ultra-large HTE datasets). Despite these high expectations, ensuring the reliability, decidability and physical grounding of these models remains essential, especially for increasingly untransparent and less interpretable deep learning models. Even when these models excel at identifying complex correlations and drastically increase productivity, the investigation of chemical causality and mechanistic understanding remains a core scientific desire, with data science serving as an assistant for data analysis rather than an end-to-end prediction tool.

Looking ahead, the future synthetic organic laboratory will likely be defined by a seamless synergy between human expertise and autonomous systems. We envision an ecosystem where experimental chemists focus on high-level hypothesis generation and creative problem-solving, while digital systems facilitate routine jobs like optimization campaigns, real time analytical interpretation, and the navigation of large chemical spaces. As with previous waves of transformative technology, the skillset expected of synthetic chemists, therefore, must also evolve, and digital literacy will likely become a core competency in chemical education. Moreover, recent advancements in laboratory automation also highlight that some, though not all, experimental tasks in synthetic chemistry can be executed in a fully autonomous, closed-loop fashion. Yet these systems are mostly useful for possibly complex, routine tasks and repeated procedures. The expensive equipment and demanding setup render self-driving labs not yet efficient for many core synthetic chemistry operations, or tasks that dynamically evolve in a research project.

Last, we argue that emerging data-driven tools will not replace chemists. Indeed, ML models, including LLMs, by definition, rely on statistics and correlations and can often appear as a "black box". Therefore, the causal, chemical interpretation of modeling results and the derivation of actionable insights remain with the chemist. However, on that mission, data science tools can empower them to back up hypotheses with rigorous data, leading to higher success rates, improved sustainability, and a more systematic exploration of the vast chemical space. As we embrace this digital transformation, the question for synthetic practitioners is no longer whether to adopt these technologies at all, but in which areas they can benefit the most to solve the most pressing challenges of synthetic chemistry.

Author Contributions

Niklas Hölter: conceptualization, investigation, writing – original draft, validation, visualization, writing – review and editing. **Felix Katzenburg:** conceptualization, investigation, writing – original draft, validation, visualization, and writing – review and editing. **Florian Boser:** conceptualization, investigation, writing – original draft, validation, visualization, writing – review and editing. **Frank Glorius:** conceptualization, validation, writing – review and editing, supervision, resources, funding acquisition.

Acknowledgments

The authors acknowledge generous financial support by the Deutsche Forschungsgemeinschaft (SPP2363—Utilization and Development of Machine Learning for Molecular Applications—Molecular Machine Learning) and the European Research Council (advanced grant agreement no. 101098156, HighEnT.). N.H. further acknowledges PhD fellowship funding by the German Academic Scholarship Foundation (Studienstiftung des deutschen Volkes). The authors warmly thank Manh Hung Tran, Leon Schlosser, Dr. Ajnabiul Hoque, Dr. Debanjan Rana, Carla Hümpel, Nils Rendel, and Rubal Sharma (University of Münster) for insightful feedback. Different artists at flaticon.com are acknowledged for various icons used in the figures.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

References

1. J. McCarthy, M. L. Minsky, N. Rochester, and C. E. Shannon, "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence (August 31, 1955)," *AI Magazine* 27 (2006): 12–14, https://doi.org/10.1609/aimag.v27i4.1904.

2. R. K. Lindsay, B. G. Buchanan, E. A. Feigenbaum, and J. Lederberg, "Dendral: A Case Study of the First Expert System for Scientific Hypothesis Formation," *Artificial Intelligence* 61 (1993): 209, https://doi.org/10.1016/0004-3702(93)90068-m.

3. J. Lederberg, "How DENDRAL Was Conceived and Born," in Proceedings of ACM Conference on History of Medical Informatics (1987), 5, https://doi.org/10.1145/41526.41528.

4. E. J. Corey and W. T. Wipke, "Computer-Assisted Design of Complex Organic Syntheses," *Science* 166 (1969): 178, https://doi.org/10.1126/science.166.3902.178.

5. E. Callaway, "Chemistry Nobel Goes to Developers of AlphaFold AI That Predicts Protein Structures," *Nature* 634 (2024): 525, https://doi.org/10.1038/d41586-024-03214-7.

6. The Royal Swedish Academy of Sciences, "The Nobel Prize in Chemistry 2024" can be found under https://www.nobelprize.org/prizes/chemistry/2024/press-release/(accessed: 9.10.2024).

7. P. M. Pflüger and F. Glorius, "Molecular Machine Learning: The Future of Synthetic Chemistry?," *Angewandte Chemie International Edition* 59 (2020): 18860, https://doi.org/10.1002/anie.202008366.

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

8. A. F. de Almeida, R. Moreira, and T. Rodrigues, "Synthetic Organic Chemistry Driven by Artificial Intelligence," *Nature Reviews Chemistry* 3 (2019): 589, https://doi.org/10.1038/s41570-019-0124-0.

9. W. L. Williams, L. Zeng, T. Gensch, M. S. Sigman, A. G. Doyle, and E. V. Anslyn, "The Evolution of Data-Driven Modeling in Organic Chemistry," *ACS Central Science* 7 (2021): 1622, https://doi.org/10.1021/acscentsci.1c00535.

10. L. Schlosser, D. Rana, P. M. Pflüger, F. Katzenburg, and F. Glorius, "EnTdecker—A Machine Learning-Based Platform for Guiding Substrate Discovery in Energy Transfer Catalysis," *Journal of the American Chemical Society* 146 (2024): 13266, https://doi.org/10.1021/jacs.4c01352.

11. Z. Tu, S. J. Choure, M. H. Fong, et al., "ASKCOS: Open-Source, Data-Driven Synthesis Planning," *Accounts of Chemical Research* 58 (2025): 1764, https://doi.org/10.1021/acs.accounts.5c00155.

12. E. Heid, K. P. Greenman, Y. Chung, et al., "Chemprop: A Machine Learning Package for Chemical Property Prediction," *Journal of Chemical Information and Modeling* 64 (2024): 9, https://doi.org/10.1021/acs.jcim.3c01250.

13. J. P. Dürholt, T. S. Asche, J. Kleinekorte, et al., "BoFire: Bayesian Optimization Framework Intended for Real Experiments," *Journal of Machine Learning Research* 26 (2025): 1.

14. M. Haddadnia, L. Grashoff, and F. Strieth-Kalthoff, "BoTier: Multi-Objective Bayesian Optimization With Tiered Objective Structures," *Digital Discovery* 4 (2025): 1417, https://doi.org/10.1039/D5DD00039D.

15. M. Fitzner, A. Šošić, A. V. Hopp, et al., "BayBE: A Bayesian Back End for Experimental Planning in the Low-to-No-Data Regime," *Digital Discovery* 4 (2025): 1991, https://doi.org/10.1039/D5DD00050E.

16. M. H. S. Segler, M. Preuss, and M. P. Waller, "Planning Chemical Syntheses With Deep Neural Networks and Symbolic AI," *Nature* 555 (2018): 604, https://doi.org/10.1038/nature25978.

17. C. W. Coley, D. A. Thomas, J. A. M. Lummiss, et al., "A Robotic Platform for Flow Synthesis of Organic Compounds Informed by AI Planning," *Science* 365 (2019): eaax1566, https://doi.org/10.1126/science.aax1566.

18. F. Strieth-Kalthoff, H. Hao, V. Rathore, et al., "Delocalized, Asynchronous, Closed-Loop Discovery of Organic Laser Emitters," *Science* 384 (2024): eadk9227, https://doi.org/10.1126/science.adk9227.

19. B. Burger, P. M. Maffettone, V. V. Gusev, et al., "A Mobile Robotic Chemist," *Nature* 583 (2020): 237, https://doi.org/10.1038/s41586-020-2442-2.

20. V. Z. Zeidler, "Digitalization Paving the Ways for Sustainable Chemistry: Switching on More Green Lights," *Science* 384 (2024): eadq3537, https://doi.org/10.1126/science.adq3537.

21. I. V. Tetko, O. Engkvist, U. Koch, J.-L. Reymond, and H. Chen, "BIGCHEM: Challenges and Opportunities for Big Data Analysis in Chemistry," *Molecular Informatics* 35 (2016): 615, https://doi.org/10.1002/minf.201600073.

22. J. A. Slezak, "Simple Computer Programming in the Physical Chemistry Course," *Journal of Chemical Education* 45 (1968): 767, https://doi.org/10.1021/ed045p767.

23. H. Kim, "Computer Programming in Physical Chemistry Laboratory: Least-Squares Analysis," *Journal of Chemical Education* 47 (1970): 120, https://doi.org/10.1021/ed047p120.

24. D. F. DeTar, "A Computer Program for Making Steady state Calculations: Notes on Effective Programming Techniques," *Journal of Chemical Education* 44 (1967): 193, https://doi.org/10.1021/ed044p193.

25. J. M. Stevens, *Enabling Tools and Techniques for Organic Synthesis. A Practical Guide to Experimentation, Automation, and Computation*, ed. S. G. Newman, (Wiley, 2023), 347–391, https://doi.org/10.1002/9781119855668.

26. A. Ringer McDonald and J. A. Nash, eds., *Teaching Programming Across the Chemistry Curriculum*, (American Chemical Society, 2021), 1–11, https://doi.org/10.1021/bk-2021-1387.

27. M. D. Wilkinson, M. Dumontier, I. J. J. Aalbersberg, et al., "The FAIR Guiding Principles for Scientific Data Management and Stewardship," *Scientific Data* 3 (2016): 160018, https://doi.org/10.1038/sdata.2016.18.

28. S. Berber, M. Brückner, N. Maurer, and J. Huwer, "Artificial Intelligence in Chemistry Research—Implications for Teaching and Learning," *Journal of Chemical Education* 102 (2025): 1445, https://doi.org/10.1021/acs.jchemed.4c01033.

29. A. R. McCluskey, M. Rivera, and A. S. J. S. Mey , "Digital Skills in Chemical Education," *Nature Chemistry* 16 (2024): 1383, https://doi.org/10.1038/s41557-024-01613-x.

30. F. M. Fung, M. Lederbauer, Y. S. Choo, et al., "Chemical Education in Digital Chemistry," *Chemistry* 10 (2024): 3519, https://doi.org/10.1016/j.chempr.2024.10.010.

31. F. Strieth-Kalthoff, F. Sandfort, M. H. S. Segler, and F. Glorius, "Machine Learning the Ropes: Principles, Applications and Directions in Synthetic Chemistry," *Chemical Society Reviews* 49 (2020): 6154, https://doi.org/10.1039/c9cs00786e.

32. S. Park, H. Han, H. Kim, and S. Choi, "Machine Learning Applications for Chemical Reactions," *Chemistry—An Asian journal* 17 (2022): e202200203, https://doi.org/10.1002/asia.202200203.

33. D. S. Wigh, J. M. Goodman, and A. A. Lapkin, "A Review of Molecular Representation in the Age of Machine Learning," *WIREs Computational Molecular Science* 12, no. 5 (2022): e1603, https://doi.org/10.1002/wcms.1603.

34. N. Artrith, K. T. Butler, F.-X. Coudert, et al., "Best Practices in Machine Learning for Chemistry," *Nature Chemistry* 13 (2021): 505, https://doi.org/10.1038/s41557-021-00716-z.

35. K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, "Machine Learning for Molecular and Materials Science," *Nature* 559 (2018): 547, https://doi.org/10.1038/s41586-018-0337-2.

36. L. David, A. Thakkar, R. Mercado, and O. Engkvist, "Molecular Representations in AI-Driven Drug Discovery: A Review and Practical Guide," *Journal of Cheminformatics* 12 (2020): 56, https://doi.org/10.1186/s13321-020-00460-5.

37. D. Rogers and M. Hahn, "Extended-Connectivity Fingerprints," *Journal of Chemical Information and Modeling* 50 (2010): 742, https://doi.org/10.1021/ci100050t.

38. F. Sandfort, F. Strieth-Kalthoff, M. Kühnemund, C. Beecks, and F. Glorius, "A Structure-Based Platform for Predicting Chemical Reactivity," *Chemistry* 6 (2020): 1379, https://doi.org/10.1016/j.chempr.2020.02.017.

39. D. Weininger, "SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules," *Journal of Chemical Information and Computer Sciences* 28 (1988): 31, https://doi.org/10.1021/ci00057a005.

40. M. Krenn, F. Häse, A. Nigam, P. Friederich, and A. Aspuru-Guzik, "Self-Referencing Embedded Strings (SELFIES): A 100% Robust Molecular String Representation," *Machine Learning: Science and Technology* 1 (2020): 45024, https://doi.org/10.1088/2632-2153/aba947.

41. P. M. Pflüger, M. Kühnemund, F. Katzenburg, H. Kuchen, and F. Glorius, "An Evolutionary Algorithm for Interpretable Molecular Representations," *Chemistry* 10 (2024): 1391, https://doi.org/10.1016/j.chempr.2024.02.004.

42. J. L. Durant, B. A. Leland, D. R. Henry, and J. G. Nourse, "Reoptimization of MDL Keys for Use in Drug Discovery," *Journal of Chemical Information and Computer Sciences* 42 (2002): 1273, https://doi.org/10.1021/ci010132r.

43. P. Schwaller, T. Laino, T. Gaudin, et al., "Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction," *ACS Central Science* 5 (2019): 1572, https://doi.org/10.1021/acscentsci.9b00576.

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

44. D. Dalmau, M. S. Sigman, and J. V. Alegre-Requena, "Machine Learning Workflows Beyond Linear Models in Low-Data Regimes," *Chemical Science* 16 (2025): 8555, https://doi.org/10.1039/d5sc00996k.

45. D. M. Lustosa and A. Milo, "Mechanistic Inference From Statistical Models at Different Data-Size Regimes," *ACS Catalysis* 12 (2022): 7886, https://doi.org/10.1021/acscatal.2c01741.

46. H. Shalit Peleg and A. Milo, "Small Data Can Play a Big Role in Chemical Discovery," *Angewandte Chemie International Edition* 62 (2023): e202219070, https://doi.org/10.1002/anie.202219070.

47. C. Bannwarth, S. Ehlert, and S. Grimme, "GFN2-xTB-An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method With Multipole Electrostatics and Density-Dependent Dispersion Contributions," *Journal of Chemical Theory and Computation* 15 (2019): 1652, https://doi.org/10.1021/acs.jctc.8b01176.

48. T. Froitzheim, M. Müller, A. Hansen, and S. Grimme, ChemRxiv preprint (2025), https://doi.org/10.26434/chemrxiv-2025-bjxvt.

49. R. Jacobs, D. Morgan, S. Attarian, et al., "A Practical Guide to Machine Learning Interatomic Potentials—Status and Future," *Current Opinion in Solid State and Materials Science* 35 (2025): 101214, https://doi.org/10.1016/j.cossms.2025.101214.

50. J. V. Alegre-Requena, S. Sowndarya S V, R. Pérez-Soto, T. M. Alturaifi, and R. S. Paton, "AQME: Automated Quantum Mechanical Environments for Researchers and Educators," *WIREs Computational Molecular Science* 13, no. 5 (2023): e1663, https://doi.org/10.1002/wcms.1663.

51. A. M. Żurański, J. Y. Wang, B. J. Shields, and A. G. Doyle, "AutoQChem: An Automated Workflow for the Generation and Storage of DFT Calculations for Organic Molecules," *Reaction Chemistry & Engineering* 7 (2022): 1276, https://doi.org/10.1039/D2RE00030J.

52. I. Migliaro, M. G. S. Weiss, and A. J. Sterling, "ChemRefine: An Open-Source Automated and Interoperable Platform for Machine Learning and Quantum Chemistry Simulations," *Journal of Chemical Theory and Computation* 22, no. 4 (2026): 1736–1747, https://doi.org/10.1021/acs.jctc.5c01881.

53. D. Dalmau and J. V. Alegre-Requena, "ROBERT: Bridging the Gap Between Machine Learning and Chemistry," *WIREs Computational Molecular Science* 14, no. 5 (2024): e1733, https://doi.org/10.1002/wcms.1733.

54. D. Dalmau, L. Jacot-Descombes, A. Kalikadien, et al., "Cost-Effective Quantum-Mechanical Workflows for Molecular Machine Learning," *ACS Catalysis* 16, no. 13 (2026): 12565–12574, https://doi.org/10.1021/acscatal.6c02583.

55. MolSSI—The Molecular Science Software Institute, "Descriptor Libraries Project" can be found under https://descriptor-libraries.molssi.org/(accessed: 24.02.2026).

56. S.-Q. Zhang, L.-C. Xu, S.-W. Li, et al., "Bridging Chemical Knowledge and Machine Learning for Performance Prediction of Organic Synthesis," *Chemistry—A European Journal* 29 (2023): e202202834, https://doi.org/10.1002/chem.202202834.

57. S.-W. Li, L.-C. Xu, C. Zhang, S.-Q. Zhang, and X. Hong, "Reaction Performance Prediction With an Extrapolative and Interpretable Graph Model Based on Chemical Knowledge," *Nature Communications* 14 (2023): 3569, https://doi.org/10.1038/s41467-023-39283-x.

58. S.-C. Li, H. Wu, A. Menon, K. A. Spiekermann, Y.-P. Li, and W. H. Green, "When Do Quantum Mechanical Descriptors Help Graph Neural Networks to Predict Chemical Properties?," *Journal of the American Chemical Society* 146 (2024): 23103, https://doi.org/10.1021/jacs.4c04670.

59. W. Gee, A. G. Doyle, S. Vargas, and A. Alexandrova, "Multi-Level QTAIM-Enriched Graph Neural Networks for Resolving Properties of Transition Metal Complexes," *Digital Discovery* 4 (2025): 3378, https://doi.org/10.1039/d5dd00220f.

60. D. A. Boiko, T. Reschützegger, B. Sanchez-Lengeling, S. M. Blau, and G. Gomes, "Advancing Molecular Machine Learning Representations With Stereoelectronics-Infused Molecular Graphs," *Nature Machine Intelligence* 7 (2025): 771, https://doi.org/10.1038/s42256-025-01031-9.

61. F. Zhuang, Z. Qi, K. Duan, et al., "A Comprehensive Survey on Transfer Learning," *Proceedings of the IEEE* 109 (2021): 43, https://doi.org/10.1109/JPROC.2020.3004555.

62. E. King-Smith, "Transfer Learning for a Foundational Chemistry Model," *Chemical Science* 15 (2024): 5143, https://doi.org/10.1039/d3sc04928k.

63. J. W. Burns, A. Zalte, and W. H. Green, arXiv preprint (2025), https://doi.org/10.48550/arxiv.2506.15792.

64. J. Ross, B. Belgodere, V. Chenthamarakshan, I. Padhi, Y. Mroueh, and P. Das, "Large-Scale Chemical Language Representations Capture Molecular Structure and Properties," *Nature Machine Intelligence* 4 (2022): 1256, https://doi.org/10.1038/s42256-022-00580-7.

65. N. C. Frey, R. Soklaski, S. Axelrod, et al., "Neural Scaling of Deep Chemical Models," *Nature Machine Intelligence* 5 (2023): 1297, https://doi.org/10.1038/s42256-023-00740-3.

66. M. L. Schrader, F. R. Schäfer, F. Schäfers, and F. Glorius, "Bridging the Information Gap in Organic Chemical Reactions," *Nature Chemistry* 16 (2024): 491, https://doi.org/10.1038/s41557-024-01470-8.

67. F. Strieth-Kalthoff, F. Sandfort, M. Kühnemund, F. R. Schäfer, H. Kuchen, and F. Glorius, "Machine Learning for Chemical Reactivity: The Importance of Failed Experiments," *Angewandte Chemie International Edition* 61 (2022): e202204647, https://doi.org/10.1002/anie.202204647.

68. R.-R. Griffiths, P. Schwaller, and A. A. Lee, arXiv preprint (2021), 10.48550/arxiv.2105.02637v1.

69. X. Jia, A. Lynch, Y. Huang, et al., "Anthropogenic Biases in Chemical Reaction Data Hinder Exploratory Inorganic Synthesis," *Nature* 573 (2019): 251, https://doi.org/10.1038/s41586-019-1540-5.

70. W. Beker, R. Roszak, A. Wołos, et al., "Machine Learning May Sometimes Simply Capture Literature Popularity Trends: A Case Study of Heterocyclic Suzuki-Miyaura Coupling," *Journal of the American Chemical Society* 144 (2022): 4819, https://doi.org/10.1021/jacs.1c12005.

71. F. Boser, J. C. Spies, and F. Glorius, "Yield Prediction of Organic Reactions in Biased Data Sets via Positive-Unlabeled Learning," *Journal of the American Chemical Society* 148 (2026): 15066, https://doi.org/10.1021/jacs.6c00127.

72. J. A. Hueffel, Q. P. Bindschaedler, F. Sala, and F. Schoenebeck, "Empowering Reactivity Predictions Through Noise-Based Data Augmentation," *Journal of the American Chemical Society* 147, no. 36 (2025): 33232–33239, https://doi.org/10.1021/jacs.5c11632.

73. D. S. Wigh, J. Arrowsmith, A. Pomberger, K. C. Felton, and A. A. Lapkin, "ORDerly: Data Sets and Benchmarks for Chemical Reaction Data," *Journal of Chemical Information and Modeling* 64 (2024): 3790, https://doi.org/10.1021/acs.jcim.4c00292.

74. T. R. Gimadiev, A. Lin, V. A. Afonina, et al., "Reaction Data Curation I: Chemical Structures and Transformations Standardization," *Molecular Informatics* 40 (2021): e2100119, https://doi.org/10.1002/minf.202100119.

75. J. Schleinitz, M. Langevin, Y. Smail, B. Wehnert, L. Grimaud, and R. Vuilleumier, "Machine Learning Yield Prediction From NiCOlit, a Small-Size Literature Data Set of Nickel Catalyzed C-O Couplings," *Journal of the American Chemical Society* 144 (2022): 14722, https://doi.org/10.1021/jacs.2c05302.

76. D. M. Lowe, "Apollo—University of Cambridge Repository," (2012), https://doi.org/10.17863/CAM.89993.

77. W. Jin, C. W. Coley, R. Barzilay, and T. Jaakkola, arXiv preprint (2017), 10.48550/arXiv.1709.04555.

78. P. Schwaller, T. Gaudin, D. Lányi, C. Bekas, and T. Laino, ""Found in Translation": Predicting Outcomes of Complex Organic Chemistry Reactions Using Neural Sequence-to-Sequence Models," *Chemical Science* 9 (2018): 6091, https://doi.org/10.1039/C8SC02339E.

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

79. D. T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher, and A. G. Doyle, "Predicting Reaction Performance in C-N Cross-Coupling Using Machine Learning," *Science* 360 (2018): 186, https://doi.org/10.1126/science.aar5169.

80. S. K. Ha, D. Kalyani, M. S. West, et al., "Developing Pharmaceutically Relevant Pd-Catalyzed C-N Coupling Reactivity Models Leveraging High-Throughput Experimentation," *Journal of the American Chemical Society* 147 (2025): 19602, https://doi.org/10.1021/jacs.5c00933.

81. J. Götz, E. Richards, I. A. Stepek, et al., "Predicting Three-Component Reaction Outcomes From ∼40,000 Miniaturized Reactant Combinations," *Science Advances* 11 (2025): eadw6047, https://doi.org/10.1126/sciadv.adw6047.

82. J. M. Stevens, J. Li, E. M. Simmons, et al., "Advancing Base Metal Catalysis Through Data Science: Insight and Predictive Models for Ni-Catalyzed Borylation Through Supervised Machine Learning," *Organometallics* 41 (2022): 1847, https://doi.org/10.1021/acs.organomet.2c00089.

83. H. Zhong, Y. Liu, H. Sun, et al., "Towards Global Reaction Feasibility and Robustness Prediction With High Throughput Data and Bayesian Deep Learning," *Nature Communications* 16 (2025): 4522, https://doi.org/10.1038/s41467-025-59812-0.

84. E. King-Smith, S. Berritt, L. Bernier, et al., "Probing the Chemical 'Reactome' With High-Throughput Experimentation Data," *Nature Chemistry* 16 (2024): 633, https://doi.org/10.1038/s41557-023-01393-w.

85. K. V. Chuang and M. J. Keiser, "Comment on ″Predicting Reaction Performance in C-N Cross-Coupling Using Machine Learning," *Science* 362 (2018): eaat8603, https://doi.org/10.1126/science.aat8603.

86. T. Krivobokova, R.-A. Morariu, G. Finocchio, and B. Maryasin, "Modelling and Estimation of Chemical Reaction Yields From High-Throughput Experiments," *Communications Chemistry* 9 (2026): 61, https://doi.org/10.1038/s42004-025-01866-8.

87. J. G. Estrada, D. T. Ahneman, R. P. Sheridan, S. D. Dreher, and A. G. Doyle, "Response to Comment on "Predicting Reaction Performance in C-N Cross-Coupling Using Machine Learning"," *Science* 362 (2018): eaat8763, https://doi.org/10.1126/science.aat8763.

88. M. A. McDonald, B. A. Koscher, R. B. Canty, and K. F. Jensen, "Calibration-Free Reaction Yield Quantification by HPLC With a Machine-Learning Model of Extinction Coefficients," *Chemical Science* 15 (2024): 10092, https://doi.org/10.1039/d4sc01881h.

89. F. Katzenburg, F. Boser, F. R. Schäfer, P. M. Pflüger, and F. Glorius, "Calibration-Free Quantification and Automated Data Analysis for High-Throughput Reaction Screening," *Digital Discovery* 4 (2025): 384, https://doi.org/10.1039/D4DD00347K.

90. M. Saebi, B. Nan, J. E. Herr, et al., "On the Use of Real-World Datasets for Reaction Yield Prediction," *Chemical Science* 14 (2023): 4997, https://doi.org/10.1039/d2sc06041h.

91. P. Neves, K. McClure, J. Verhoeven, et al., "Global Reactivity Models Are Impactful in Industrial Synthesis Applications," *Journal of Cheminformatics* 15 (2023): 20, https://doi.org/10.1186/s13321-023-00685-0.

92. P. Raghavan, B. C. Haas, M. E. Ruos, et al., "Dataset Design for Building Models of Chemical Reactivity," *ACS Central Science* 9 (2023): 2196, https://doi.org/10.1021/acscentsci.3c01163.

93. A. F. Zahrt, J. J. Henle, and S. E. Denmark, "Cautionary Guidelines for Machine Learning Studies With Combinatorial Datasets," *ACS Combinatorial Science* 22 (2020): 586, https://doi.org/10.1021/acscombsci.0c00118.

94. M. Baumer, L. Gallegos, D. Anstine, et al., "Democratizing Reaction Kinetics Through Machine Vision and Learning," (2026) , https://doi.org/10.26434/chemrxiv-2025-4tk40/v2.

95. H. Barrington, T. J. D. McCabe, K. Donnachie, et al., "Parallel and High Throughput Reaction Monitoring With Computer Vision," *Angewandte Chemie International Edition* 64 (2025): e202413395, https://doi.org/10.1002/anie.202413395.

96. J. S. Moore and K. F. Jensen, ""Batch" Kinetics in Flow: Online IR Analysis and Continuous Control," *Angewandte Chemie International Edition* 53 (2014): 470, https://doi.org/10.1002/anie.201306468.

97. K. C. Aroh and K. F. Jensen, "Efficient Kinetic Experiments in Continuous Flow Microreactors," *Reaction Chemistry & Engineering* 3 (2018): 94, https://doi.org/10.1039/C7RE00163K.

98. C. A. Hone, N. Holmes, G. R. Akien, R. A. Bourne, and F. L. Muller, "Rapid Multistep Kinetic Model Generation From Transient Flow Data," *Reaction Chemistry & Engineering* 2 (2017): 103, https://doi.org/10.1039/c6re00109b.

99. D. F. Nippa, A. T. Müller, K. Atz, et al., "Simple User-Friendly Reaction Format," *Molecular Informatics* 44 (2025): e202400361, https://doi.org/10.1002/minf.202400361.

100. S. M. Kearnes, M. R. Maser, M. Wleklinski, et al., "The Open Reaction Database," *Journal of the American Chemical Society* 143 (2021): 18820, https://doi.org/10.1021/jacs.1c09820.

101. P. Tremouilhac, P.-C. Huang, C.-L. Lin, et al., "Chemotion Repository, a Curated Repository for Reaction Information and Analytical Data," *Chemistry—Methods* 1 (2021): 8, https://doi.org/10.1002/cmtd.202000034.

102. A. F. Zahrt, J. J. Henle, B. T. Rose, Y. Wang, W. T. Darrow, and S. E. Denmark, "Prediction of Higher-Selectivity Catalysts by Computer-Driven Workflow and Machine Learning," *Science* 363 (2019): eaau5631, https://doi.org/10.1126/science.aau5631.

103. D. Perera, J. W. Tucker, S. Brahmbhatt, et al., "A Platform for Automated Nanomole-Scale Reaction Screening and Micromole-Scale Synthesis in Flow," *Science* 359 (2018): 429, https://doi.org/10.1126/science.aap9112.

104. "NextMove Software | Pistachio" can be found under https://www.nextmovesoftware.com/pistachio (accessed: 15.06.2026).

105. A. Bender, N. Schneider, M. Segler, W. P. Walters, O. Engkvist, and T. Rodrigues, "Evaluation Guidelines for Machine Learning Tools in the Chemical Sciences," *Nature Reviews Chemistry* 6 (2022): 428, https://doi.org/10.1038/s41570-022-00391-9.

106. T. Hanser, C. Barber, J. F. Marchaland, and S. Werner, "Applicability Domain: Towards a More Formal Definition," *SAR and QSAR in Environmental Research* 27 (2016): 893, https://doi.org/10.1080/1062936X.2016.1250229.

107. T. Frömbgen, E. Surzhikova, J. Dölz, J. Proppe, B. Kirchner, and C. R. Jacob, "Uncertainty Quantification for in Silico Chemistry," *Chemical Reviews* 126, no. 7 (2026): 4189–4236, https://doi.org/10.1021/acs.chemrev.5c00931.

108. E. Hüllermeier and W. Waegeman, "Aleatoric and Epistemic Uncertainty in Machine Learning: An Introduction to Concepts and Methods," *Machine Learning* 110 (2021): 457, https://doi.org/10.1007/s10994-021-05946-3.

109. E. Heid, C. J. McGill, F. H. Vermeire, and W. H. Green, "Characterizing Uncertainty in Machine Learning for Chemistry," *Journal of Chemical Information and Modeling* 63 (2023): 4012, https://doi.org/10.1021/acs.jcim.3c00373.

110. I. O. Betinol, A. Demchenko, and J. P. Reid, "Evaluating Predictive Accuracy in Asymmetric Catalysis: A Machine Learning Perspective on Local Reaction Space," *ACS Catalysis* 15 (2025): 6067, https://doi.org/10.1021/acscatal.5c01051.

111. K. V. Chuang and M. J. Keiser, "Adversarial Controls for Scientific Machine Learning," *ACS Chemical Biology* 13 (2018): 2819, https://doi.org/10.1021/acschembio.8b00881.

112. G. Maggiora, M. Vogt, D. Stumpfe, and J. Bajorath, "Molecular Similarity in Medicinal Chemistry," *Journal of Medicinal Chemistry* 57 (2014): 3186, https://doi.org/10.1021/jm401411z.

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

113. M. Haghighatlari, G. Vishwakarma, D. Altarawy, et al., "ChemML: A Machine Learning and Informatics Program Package for the Analysis, Mining, and Modeling of Chemical and Materials Data," *WIREs Computational Molecular Science* 10, no. 4 (2020): e1458, https://doi.org/10.1002/wcms.1458.

114. B. Ramsundar, P. Eastman, P. Walters, and V. Pande, "Deep Learning for the Life Sciences," (2019), Applying deep learning to genomics, microscopy, drug discovery, and more, O'Reilly, Beijing, Boston, Farnham, Sebastopol, Tokyo.

115. D. E. Graff, N. K. Morgan, J. W. Burns, et al., "Chemprop v2: An Efficient, Modular Machine Learning Package for Chemical Property Prediction," *Journal of Chemical Information and Modeling* 66 (2026): 28, https://doi.org/10.1021/acs.jcim.5c02332.

116. J. W. Burns and W. H. Green, "Generalizable, Fast, and Accurate DeepQSPR With Fastprop," *Journal of Cheminformatics* 17 (2025): 73, https://doi.org/10.1186/s13321-025-01013-4.

117. S. Chithrananda, G. Grand, and B. Ramsundar, arXiv preprint (2020), 10.48550/arxiv.2010.09885.

118. B. Fabian, T. Edlich, H. Gaspar, et al., arXiv preprint (2020), 10.48550/arxiv.2011.13230.

119. F. Pedregosa, G. Varoquaux, A. Gramfort, et al., "Scikit-Learn: Machine Learning in Python," *Journal of Machine Learning Research* 12 (2011): 2825, arXiv:1201.0490.

120. M. Abadi, P. Barham, J. Chen, et al., arXiv preprint (2016), 10.48550/arxiv.1605.08695.

121. A. Paszke, S. Gross, F. Massa, et al., arXiv preprint (2019), 10.48550/arXiv.1912.01703.

122. M. Fey and J. E. Lenssen, arXiv preprint (2019), 10.48550/arxiv.1903.02428.

123. J. Bradbury, R. Frostig, P. Hawkins, et al., "JAX: Composable Transformations of Python+NumPy Programs" (2018), can be found under http://github.com/jax-ml/jax.

124. M. Rupp, R. Ramakrishnan, and O. A. von Lilienfeld, "Machine Learning for Quantum Mechanical Properties of Atoms in Molecules," *Journal of Physical Chemistry Letters* 6 (2015): 3309, https://doi.org/10.1021/acs.jpclett.5b01456.

125. M. Wen, S. M. Blau, E. W. C. Spotte-Smith, S. Dwaraknath, and K. A. Persson, "BonDNet: A Graph Neural Network for the Prediction of Bond Dissociation Energies for Charged Molecules," *Chemical Science* 12 (2020): 1858, https://doi.org/10.1039/d0sc05251e.

126. X. Qu, D. A. R. S. Latino, and J. Aires-de-Sousa, "A Big Data Approach to the Ultra-Fast Prediction of DFT-Calculated Bond Energies," *Journal of Cheminformatics* 5 (2013): 34, https://doi.org/10.1186/1758-2946-5-34.

127. P. C. St John, Y. Guan, Y. Kim, S. Kim, and R. S. Paton, "Prediction of Organic Homolytic Bond Dissociation Enthalpies at Near Chemical Accuracy With Sub-Second Computational Cost," *Nature Communications* 11 (2020): 2328, https://doi.org/10.1038/s41467-020-16201-z.

128. S. V. Shree Sowndarya, Y. Kim, S. Kim, P. C. St. John, and R. S. Paton, "Expansion of Bond Dissociation Prediction With Machine Learning to Medicinally and Environmentally Relevant Chemical Space," *Digital Discovery* 2 (2023): 1900, https://doi.org/10.1039/D3DD00169E.

129. D. Rana, C. Hümpel, R. Laskar, et al., "Accelerated Discovery of Energy Transfer-Catalyzed Dearomative Cycloadditions Through a Data-Driven Three-Layer Screening Strategy," *Journal of the American Chemical Society* 147 (2025): 28359, https://doi.org/10.1021/jacs.5c09249.

130. B. C. Haas, M. A. Hardy, S. Sowndarya S V, et al., "Rapid Prediction of Conformationally-Dependent DFT-Level Descriptors Using Graph Neural Networks for Carboxylic Acids and Alkyl Amines," *Digital Discovery* 4 (2025): 222, https://doi.org/10.1039/d4dd00284a.

131. G. Takasao, B. Maity, S. Dutta, R. Kancherla, M. Rueping, and L. Cavallo, "NHC-Cracker: A Platform for the in Silico Engineering of N-Heterocyclic Carbenes for Diverse Chemical Applications," *ACS Catalysis* 15 (2025): 5915, https://doi.org/10.1021/acscatal.4c06474.

132. R. A. Mansson, A. H. Welsh, N. Fey, and A. G. Orpen, "Statistical Modeling of a Ligand Knowledge Base," *Journal of Chemical Information and Modeling* 46 (2006): 2591, https://doi.org/10.1021/ci600212t.

133. T. Gensch, G. Dos Passos Gomes, P. Friederich, et al., "A Comprehensive Discovery Platform for Organophosphorus Ligands for Catalysis," *Journal of the American Chemical Society* 144 (2022): 1205, https://doi.org/10.1021/jacs.1c09718.

134. E. Jonas and S. Kuhn, "Rapid Prediction of NMR Spectral Properties With Quantified Uncertainty," *Journal of Cheminformatics* 11 (2019): 50, https://doi.org/10.1186/s13321-019-0374-3.

135. Y. Li, W.-S. Huang, L. Zhang, D. Su, H. Xu, and X.-S. Xue, "Prediction of 19F NMR Chemical Shift by Machine Learning," *Artificial Intelligence* 2 (2024): 100043, https://doi.org/10.1016/j.aichem.2024.100043.

136. Y. Li, H. Xu, A. Kumar, et al., "TransPeakNet for Solvent-Aware 2D NMR Prediction via Multi-Task Pre-Training and Unsupervised Learning," *Communications Chemistry* 8 (2025): 51, https://doi.org/10.1038/s42004-025-01455-9.

137. A. Bhadauria, Z. Feng, M. Popescu, and R. Paton, ChemRxiv preprint (2025), 10.26434/chemrxiv-2025-r8m9m.

138. Y. Guan, S. V. Shree Sowndarya, L. C. Gallegos, P. C. St John, and R. S. Paton, "Real-Time Prediction of 1H and 13C Chemical Shifts With DFT Accuracy Using a 3D Graph Neural Network," *Chemical Science* 12 (2021): 12012, https://doi.org/10.1039/d1sc03343c.

139. J.-J. Wang, Y. Jin, C.-Y. Zhi, et al., "NMRexp: A Database of 3.3 Million Experimental NMR Spectra," *Scientific Data* 12 (2025): 1954, https://doi.org/10.1038/s41597-025-06245-5.

140. J. B. Rowlands, R. J. Lewis, P. Howe, and J. M. Goodman, "CSP5: Large-scale Neural Chemical Shift Prediction From 2.5 Million Experimental NMR Spectra," (2026), https://doi.org/10.26434/chemrxiv.15001823/v1.

141. F. Xu, J. Hu, J. Zou, et al., arXiv preprint (2026), https://doi.org/10.48550/arXiv.2602.08752.

142. P. Bonini, T. Kind, H. Tsugawa, D. K. Barupal, and O. Fiehn, "Retip: Retention Time Prediction for Compound Annotation in Untargeted Metabolomics," *Analytical Chemistry* 92 (2020): 7515, https://doi.org/10.1021/acs.analchem.9b05765.

143. X. Domingo-Almenara, C. Guijas, E. Billings, et al., "The METLIN Small Molecule Dataset for Machine Learning-Based Retention Time Prediction," *Nature Communications* 10 (2019): 5811, https://doi.org/10.1038/s41467-019-13680-7.

144. Q. Yang, H. Ji, H. Lu, and Z. Zhang, "Prediction of Liquid Chromatographic Retention Time With Graph Neural Networks to Assist in Small Molecule Identification," *Analytical Chemistry* 93 (2021): 2200, https://doi.org/10.1021/acs.analchem.0c04071.

145. S. A. Dymura, O. O. Viniichuk, K. P. Melnykov, D. S. Radchenko, and O. O. Grygorenko, "Machine Learning-Based Retention Time Prediction Tool for Routine LC-MS Data Analysis," *Journal of Chemical Information and Modeling* 65 (2025): 7415, https://doi.org/10.1021/acs.jcim.5c00514.

146. E. S. Fedorova, D. D. Matyushin, I. V. Plyushchenko, A. N. Stavrianidi, and A. K. Buryak, "Deep Learning for Retention Time Prediction in Reversed-Phase Liquid Chromatography," *Journal of Chromatography A* 1664 (2022): 462792, https://doi.org/10.1016/j.chroma.2021.462792.

147. H. Xu, J. Lin, D. Zhang, and F. Mo, "Retention Time Prediction for Chromatographic Enantioseparation by Quantile Geometry-Enhanced Graph Neural Network," *Nature Communications* 14 (2023): 3095, https://doi.org/10.1038/s41467-023-38853-3.

148. J. Stanstrup, S. Neumann, and U. Vrhovšek, "PredRet: Prediction of Retention Time by Direct Mapping Between Multiple Chromatographic Systems," *Analytical Chemistry* 87 (2015): 9421, https://doi.org/10.1021/acs.analchem.5b02287.

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

149. D. Pasin, C. B. Mollerup, B. S. Rasmussen, K. Linnet, and P. W. Dalsgaard, "Development of a Single Retention Time Prediction Model Integrating Multiple Liquid Chromatography Systems: Application to New Psychoactive Substances," *Analytica Chimica Acta* 1184 (2021): 339035, https://doi.org/10.1016/j.aca.2021.339035.

150. X.-Y. Lu, H.-P. Wu, H. Ma, et al., "Deep Learning-Assisted Spectrum-Structure Correlation: State-of-the-Art and Perspectives," *Analytical Chemistry* 96 (2024): 7959, https://doi.org/10.1021/acs.analchem.4c01639.

151. X. Xue, H. Sun, M. Yang, et al., "Advances in the Application of Artificial Intelligence-Based Spectral Data Interpretation: A Perspective," *Analytical Chemistry* 95 (2023): 13733, https://doi.org/10.1021/acs.analchem.3c02540.

152. S. Feng, M. Huang, Y. Li, et al., "Intelligent Understanding of Spectra: From Structural Elucidation to Property Design," *Chemical Society Reviews* 54 (2025): 8243, https://doi.org/10.1039/d4cs01293c.

153. K. Guo, Y. Shen, G. A. Gonzalez-Montiel, et al., arXiv preprint (2025), 10.48550/arxiv.2502.09897.

154. J. Westermayr and P. Marquetand, "Machine Learning Spectroscopy to Advance Computation and Analysis," *Chemical Science* 16 (2025): 21660, https://doi.org/10.1039/d5sc05628d.

155. V. Voinarovska, M. Kabeshov, D. Dudenko, S. Genheden, and I. V. Tetko, "When Yield Prediction Does Not Yield Prediction: An Overview of the Current Challenges," *Journal of Chemical Information and Modeling* 64 (2024): 42, https://doi.org/10.1021/acs.jcim.3c01524.

156. Y. Jia, R. Frydrych, Y. I. Sobolev, et al., "Robot-Assisted Mapping of Chemical Reaction Hyperspaces and Networks," *Nature* 645 (2025): 922, https://doi.org/10.1038/s41586-025-09490-1.

157. Z. Liu, Y. S. Moroz, and O. Isayev, "The Challenge of Balancing Model Sensitivity and Robustness in Predicting Yields: A Benchmarking Study of Amide Coupling Reactions," *Chemical Science* 14 (2023): 10835, https://doi.org/10.1039/D3SC03902A.

158. D. van Tilborg, A. Alenicheva, and F. Grisoni, "Exposing the Limitations of Molecular Machine Learning With Activity Cliffs," *Journal of Chemical Information and Modeling* 62 (2022): 5938, https://doi.org/10.1021/acs.jcim.2c01073.

159. D. Horvath, G. Marcou, A. Varnek, S. Kayastha, A. de La Vega León, and J. Bajorath, "Prediction of Activity Cliffs Using Condensed Graphs of Reaction Representations, Descriptor Recombination, Support Vector Machine Classification, and Support Vector Regression," *Journal of Chemical Information and Modeling* 56 (2016): 1631, https://doi.org/10.1021/acs.jcim.6b00359.

160. M. K. Nielsen, D. T. Ahneman, O. Riera, and A. G. Doyle, "Deoxyfluorination With Sulfonyl Fluorides: Navigating Reaction Space With Machine Learning," *Journal of the American Chemical Society* 140 (2018): 5004, https://doi.org/10.1021/jacs.8b01523.

161. P. Schwaller, A. C. Vaucher, T. Laino, and J.-L. Reymond, "Prediction of Chemical Reaction Yields Using Deep Learning," *Machine Learning: Science and Technology* 2 (2021): 15016, https://doi.org/10.1088/2632-2153/abc81d.

162. L.-C. Xu, M.-J. Tang, J. An, F. Cao, and Y. Qi, "A Unified Pre-Trained Deep Learning Framework for Cross-Task Reaction Performance Prediction and Synthesis Planning," *Nature Machine Intelligence* 7 (2025): 1561, https://doi.org/10.1038/s42256-025-01098-4.

163. N. I. Rinehart, R. K. Saunthwal, J. Wellauer, et al., "A Machine-Learning Tool to Predict Substrate-Adaptive Conditions for Pd-Catalyzed C-N Couplings," *Science* 381 (2023): 965, https://doi.org/10.1126/science.adg2114.

164. M. Sadowski, L. Sztukiewicz, M. Wyrzykowska, et al., "Scaling High-Throughput Experimentation Unlocks Robust Reaction-Outcome Prediction," "NeurIPS 2025 AI for Science Workshop," (2025).

165. J. M. Brown and R. J. Deeth, "Is Enantioselectivity Predictable in Asymmetric Catalysis?," *Angewandte Chemie International Edition* 48 (2009): 4476, https://doi.org/10.1002/anie.200900697.

166. S. Pinus, J. Genzling, M. Burai-Patrascu, and N. Moitessier, "Computational Methods for Asymmetric Catalysis," *Nature Catalysis* 7 (2024): 1272, https://doi.org/10.1038/s41929-024-01258-6.

167. B. C. Haas, D. Kalyani, and M. S. Sigman, "Applying Statistical Modeling Strategies to Sparse Datasets in Synthetic Chemistry," *Science Advances* 11 (2025): eadt3013, https://doi.org/10.1126/sciadv.adt3013.

168. M. Ruth, T. Gensch, and P. R. Schreiner, "Contrasting Historical and Physical Perspectives in Asymmetric Catalysis: $\Delta\Delta G^{\neq}$ versus Enantiomeric Excess," *Angewandte Chemie International Edition* 63 (2024): e202410308, https://doi.org/10.1002/anie.202410308.

169. B. C. Haas, N.-K. Lim, J. Jermaks, et al., "Enantioselective Sulfonimidamide Acylation via a Cinchona Alkaloid-Catalyzed Desymmetrization: Scope, Data Science, and Mechanistic Investigation," *Journal of the American Chemical Society* 146 (2024): 8536, https://doi.org/10.1021/jacs.4c00374.

170. M. Orlandi, J. A. S. Coelho, M. J. Hilton, F. D. Toste, and M. S. Sigman, "Parametrization of Non-Covalent Interactions for Transition State Interrogation Applied to Asymmetric Catalysis," *Journal of the American Chemical Society* 139 (2017): 6803, https://doi.org/10.1021/jacs.7b02311.

171. J. P. Reid and M. S. Sigman, "Holistic Prediction of Enantioselectivity in Asymmetric Catalysis," *Nature* 571 (2019): 343, https://doi.org/10.1038/s41586-019-1384-z.

172. C. B. Santiago, J.-Y. Guo, and M. S. Sigman, "Predictive and Mechanistic Multivariate Linear Regression Models for Reaction Development," *Chemical Science* 9 (2018): 2398, https://doi.org/10.1039/C7SC04679K.

173. J. Werth and M. S. Sigman, "Connecting and Analyzing Enantioselective Bifunctional Hydrogen Bond Donor Catalysis Using Data Science Tools," *Journal of the American Chemical Society* 142 (2020): 16382, https://doi.org/10.1021/jacs.0c06905.

174. S. Zhao, T. Gensch, B. Murray, Z. L. Niemeyer, M. S. Sigman, and M. R. Biscoe, "Enantiodivergent Pd-Catalyzed C-C Bond Formation Enabled Through Ligand Parameterization," *Science* 362 (2018): 670, https://doi.org/10.1126/science.aat2299.

175. S. Gallarati, E. M. Bucci, A. G. Doyle, and M. S. Sigman, "Transferable Enantioselectivity Models From Sparse Data," *Nature* 651 (2026): 637–646, https://doi.org/10.1038/s41586-026-10239-7.

176. S. Singh, M. Pareek, A. Changotra, et al., "A Unified Machine-Learning Protocol for Asymmetric Catalysis as a Proof of Concept Demonstration Using Asymmetric Hydrogenation," *Proceedings of the National Academy of Sciences of the United States of America* 117 (2020): 1339, https://doi.org/10.1073/pnas.1916392117.

177. L.-C. Xu, J. Frey, X. Hou, et al., "Enantioselectivity Prediction of Pallada-Electrocatalysed C–H Activation Using Transition state Knowledge in Machine Learning," *Nature Synthesis* 2 (2023): 321, https://doi.org/10.1038/s44160-022-00233-y.

178. L. Cheng, P.-L. Shao, J. Lv, et al., "Chemistry-Informed Deep Learning Model for Predicting Stereoselectivity and Absolute Configuration in Asymmetric Hydrogenation," *Nature Computational Science* 6 (2026): 145, https://doi.org/10.1038/s43588-025-00920-8.

179. L. M. Sigmund, M. Assante, M. J. Johansson, P.-O. Norrby, K. Jorner, and M. Kabeshov, "Computational Tools for the Prediction of Site- and Regioselectivity of Organic Reactions," *Chemical Science* 16 (2025): 5383, https://doi.org/10.1039/d5sc00541h.

180. P. E. Gormisky and M. C. White, "Catalyst-Controlled Aliphatic C-H Oxidations With a Predictive Model for Site-Selectivity," *Journal of the American Chemical Society* 135 (2013): 14052, https://doi.org/10.1021/ja407388y.

181. A. J. Bischoff, B. M. Nelson, Z. L. Niemeyer, M. S. Sigman, and M. Movassaghi, "Quantitative Modeling of Bis(pyridine)Silver(I) Permanganate Oxidation of Hydantoin Derivatives: Guidelines for Predicting the Site of Oxidation in Complex Substrates," *Journal of the American Chemical Society* 139 (2017): 15539, https://doi.org/10.1021/jacs.7b09541.

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

182. E. N. Bess, R. J. DeLuca, D. J. Tindall, et al., "Analyzing Site Selectivity in Rh2(esp)2-Catalyzed Intermolecular C-H Amination Reactions," *Journal of the American Chemical Society* 136 (2014): 5783, https://doi.org/10.1021/ja5015508.

183. R. C. Cammarota, W. Liu, J. Bacsa, H. M. L. Davies, and M. S. Sigman, "Mechanistically Guided Workflow for Relating Complex Reactive Site Topologies to Catalyst Performance in C-H Functionalization Reactions," *Journal of the American Chemical Society* 144 (2022): 1881, https://doi.org/10.1021/jacs.1c12198.

184. Y. T. Boni, R. C. Cammarota, K. Liao, M. S. Sigman, and H. M. L. Davies, "Leveraging Regio- and Stereoselective C(sp3)-H Functionalization of Silyl Ethers to Train a Logistic Regression Classification Model for Predicting Site-Selectivity Bias," *Journal of the American Chemical Society* 144 (2022): 15549, https://doi.org/10.1021/jacs.2c04383.

185. L.-C. Yang, X. Li, S.-Q. Zhang, and X. Hong, "Machine Learning Prediction of Hydrogen Atom Transfer Reactivity in Photoredox-Mediated C–H Functionalization," *Organic Chemistry Frontiers* 8 (2021): 6187, https://doi.org/10.1039/D1QO01325D.

186. X. Li, S.-Q. Zhang, L.-C. Xu, and X. Hong, "Predicting Regioselectivity in Radical C-H Functionalization of Heterocycles Through Machine Learning," *Angewandte Chemie International Edition* 59 (2020): 13253, https://doi.org/10.1002/anie.202000959.

187. T. J. Struble, C. W. Coley, and K. F. Jensen, "Multitask Prediction of Site Selectivity in Aromatic C–H Functionalization Reactions," *Reaction Chemistry & Engineering* 5 (2020): 896, https://doi.org/10.1039/D0RE00071J.

188. Y. Guan, C. W. Coley, H. Wu, et al., "Regio-Selectivity Prediction With a Machine-Learned Reaction Representation and on-the-Fly Quantum Mechanical Descriptors," *Chemical Science* 12 (2020): 2198, https://doi.org/10.1039/d0sc04823b.

189. B. Li, Y. Liu, H. Sun, et al., "Regio-MPNN: Predicting Regioselectivity for General Metal-Catalyzed Cross-Coupling Reactions Using a Chemical Knowledge Informed Message Passing Neural Network," *Digital Discovery* 3 (2024): 2019, https://doi.org/10.1039/D4DD00244J.

190. X. Chen, Z.-J. Zhang, X. Hong, and L. Ackermann, "Integrating a Multitask Graph Neural Network With DFT Calculations for Site-Selectivity Prediction of Arenes and Mechanistic Knowledge Generation," *Nature Synthesis* 4 (2025): 877, https://doi.org/10.1038/s44160-025-00770-2.

191. D. F. Nippa, K. Atz, R. Hohler, et al., "Enabling Late-Stage Drug Diversification by High-Throughput Experimentation With Geometric Deep Learning," *Nature Chemistry* 16 (2024): 239, https://doi.org/10.1038/s41557-023-01360-5.

192. J. Schleinitz, A. Carretero-Cerdán, A. Gurajapu, et al., "Designing Target-Specific Data Sets for Regioselectivity Predictions on Complex Substrates," *Journal of the American Chemical Society* 147 (2025): 7476, https://doi.org/10.1021/jacs.4c15902.

193. R. M. Borup, N. Ree, and J. H. Jensen, "pKalculator: A pKa Predictor for C-H Bonds," *Beilstein Journal of Organic Chemistry* 20 (2024): 1614, https://doi.org/10.3762/bjoc.20.144.

194. R. Roszak, W. Beker, K. Molga, and B. A. Grzybowski, "Rapid and Accurate Prediction of pKa Values of C-H Acids Using Graph Convolutional Neural Networks," *Journal of the American Chemical Society* 141 (2019): 17142, https://doi.org/10.1021/jacs.9b05895.

195. R. M. Borup, N. Ree, and J. H. Jensen, "Computational Prediction of C-H Hydricities and Their Use in Predicting the Regioselectivity of Electron Rich C-H Functionalisation Reactions," *Beilstein Journal of Organic Chemistry* 22 (2026): 603–610, https://doi.org/10.3762/bjoc.22.46.

196. M. Ball, D. Horvath, T. Kogej, M. Kabeshov, and A. Varnek, "Predicting Reaction Conditions: A Data-Driven Perspective," *Chemical Science* 16 (2025): 17523, https://doi.org/10.1039/d5sc03045e.

197. O. T. Schilter, C. Baldassari, T. Laino, and P. Schwaller, "Machine Learning Models for Solvent Prediction in Organic Reactions: Bridging the Gap Between Theory and Practical Efficacy," (2025), https://doi.org/10.26434/chemrxiv-2023-hmml5-v2.

198. Z. Wang, K. Lin, J. Pei, and L. Lai, "Reacon: A Template- and Cluster-Based Framework for Reaction Condition Prediction," *Chemical Science* 16 (2025): 854, https://doi.org/10.1039/D4SC05946H.

199. A. I. Lin, T. I. Madzhidov, O. Klimchuk, R. I. Nugmanov, I. S. Antipin, and A. Varnek, "Automatized Assessment of Protective Group Reactivity: A Step Toward Big Reaction Data Analysis," *Journal of Chemical Information and Modeling* 56 (2016): 2140, https://doi.org/10.1021/acs.jcim.6b00319.

200. F. Hoonakker, N. Lachiche, A. Varnek, and A. Wagner, "A Representation To Apply Usual Data Mining Techniques to Chemical Reactions — Illustration on the Rate Constant of Sn 2 Reactions in Water," *International Journal on Artificial Intelligence Tools* 20 (2011): 253, https://doi.org/10.1142/S0218213011000140.

201. S. S. Gandhi, G. Z. Brown, S. Aikonen, et al., "Data Science-Driven Discovery of Optimal Conditions and a Condition-Selection Model for the Chan–Lam Coupling of Primary Sulfonamides," *ACS Catalysis* 15 (2025): 2292, https://doi.org/10.1021/acscatal.4c07972.

202. E. Shim, A. Tewari, T. Cernak, and P. M. Zimmerman, "Recommending Reaction Conditions With Label Ranking," *Chemical Science* 16 (2025): 4109, https://doi.org/10.1039/d4sc06728b.

203. P. Baczewska, M. Kulczykowski, B. Zambroń, et al., "Machine Learning Algorithm Guides Catalyst Choices for Magnesium-Catalyzed Asymmetric Reactions," *Angewandte Chemie International Edition* 63 (2024): e202318487, https://doi.org/10.1002/anie.202318487.

204. F. Millward, M. Kulczykowski, J. Badland-Shaw, et al., "Design and Experimental Validation of a Photocatalyst Recommender Based on a Large Language Model," *Angewandte Chemie International Edition* 65 (2026): e14544, https://doi.org/10.1002/anie.202514544.

205. V. A. Afonina, D. A. Mazitov, A. Nurmukhametova, et al., "Prediction of Optimal Conditions of Hydrogenation Reaction Using the Likelihood Ranking Approach," *International Journal of Molecular Sciences* 23 (2021), https://doi.org/10.3390/ijms23010248.

206. H. Gao, T. J. Struble, C. W. Coley, Y. Wang, W. H. Green, and K. F. Jensen, "Using Machine Learning To Predict Suitable Conditions for Organic Reactions," *ACS Central Science* 4 (2018): 1465, https://doi.org/10.1021/acscentsci.8b00357.

207. M. R. Maser, A. Y. Cui, S. Ryou, T. J. DeLano, Y. Yue, and S. E. Reisman, "Multilabel Classification Models for the Prediction of Cross-Coupling Reaction Conditions," *Journal of Chemical Information and Modeling* 61 (2021): 156, https://doi.org/10.1021/acs.jcim.0c01234.

208. M. Nouman, R. B. Canty, B. A. Koscher, M. A. McDonald, and K. F. Jensen, "General Chemically Intuitive Atom- and Bond-Level DFT Descriptors for Machine Learning Approaches to Reaction Condition Prediction," *Journal of Chemical Information and Modeling* 65 (2025): 6499, https://doi.org/10.1021/acs.jcim.4c02255.

209. X. Sun, J. Liu, B. Mahjour, K. F. Jensen, and C. W. Coley, "Data-Driven Recommendation of Agents, Temperature, and Equivalence Ratios for Organic Synthesis," *Chemical Science* 16 (2025): 18176, https://doi.org/10.1039/D5SC04957A.

210. H. Li, S. Sarkar, W. Lu, et al., "Collective Intelligence for AI-Assisted Chemical Synthesis," *Nature* 651 (2026): 107–115, https://doi.org/10.1038/s41586-026-10131-4.

211. S. P. Schmid, L. Schlosser, F. Glorius, and K. Jorner, "Catalysing (organo-)Catalysis: Trends in the Application of Machine Learning to Enantioselective Organocatalysis," *Beilstein Journal of Organic Chemistry* 20 (2024): 2280, https://doi.org/10.3762/bjoc.20.196.

212. M. S. Sigman, K. C. Harper, E. N. Bess, and A. Milo, "The Development of Multidimensional Analysis Tools for Asymmetric Catalysis and Beyond," *Accounts of Chemical Research* 49 (2016): 1292, https://doi.org/10.1021/acs.accounts.6b00194.

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

213. M. E. Akana, S. Tcyrulnikov, B. D. Akana-Schneider, et al., "Computational Methods Enable the Prediction of Improved Catalysts for Nickel-Catalyzed Cross-Electrophile Coupling," *Journal of the American Chemical Society* 146 (2024): 3043, https://doi.org/10.1021/jacs.3c09554.

214. S. Dana, N. K. Pandit, P. Boos, et al., "Parametrization of κ2-N,O-Oxazoline Preligands for Enantioselective Cobaltaelectro-Catalyzed C-H Activations," *ACS Catalysis* 15 (2025): 4450, https://doi.org/10.1021/acscatal.5c00250.

215. K. C. Harper and M. S. Sigman, "Predicting and Optimizing Asymmetric Catalyst Performance Using the Principles of Experimental Design and Steric Parameters," *Proceedings of the National Academy of Sciences of the United States of America* 108 (2011): 2179, https://doi.org/10.1073/pnas.1013331108.

216. J. F. Goebel, J. Löffler, Z. Zeng, et al., "Computer-Driven Development of Ylide Functionalized Phosphines for Palladium-Catalyzed Hiyama Couplings," *Angewandte Chemie International Edition* 62 (2023): e202216160, https://doi.org/10.1002/anie.202216160.

217. K. Wu and A. G. Doyle, "Parameterization of Phosphine Ligands Demonstrates Enhancement of Nickel Catalysis via Remote Steric Effects," *Nature Chemistry* 9 (2017): 779, https://doi.org/10.1038/nchem.2741.

218. A. Milo, A. J. Neel, F. D. Toste, and M. S. Sigman, "A Data-Intensive Approach to Mechanistic Elucidation Applied to Chiral Anion Catalysis," *Science* 347 (2015): 737, https://doi.org/10.1126/science.1261043.

219. C. L. Olen, A. F. Zahrt, S. W. Reilly, et al., "Chemoinformatic Catalyst Selection Methods for the Optimization of Copper–Bis(oxazoline)-Mediated, Asymmetric, Vinylogous Mukaiyama Aldol Reactions," *ACS Catalysis* 14 (2024): 2642, https://doi.org/10.1021/acscatal.3c05903.

220. H. Chen, S. Yamaguchi, Y. Morita, et al., "Data-Driven Catalyst Optimization for Stereodivergent Asymmetric Synthesis by Iridium/Boron Hybrid Catalysis," *Cell Reports Physical Science* 2 (2021): 100679, https://doi.org/10.1016/j.xcrp.2021.100679.

221. J. Jover, N. Fey, J. N. Harvey, et al., "Expansion of the Ligand Knowledge Base for Monodentate P-Donor Ligands (LKB-P)," *Organometallics* 29 (2010): 6245, https://doi.org/10.1021/om100648v.

222. J. Jover, N. Fey, J. N. Harvey, et al., "Expansion of the Ligand Knowledge Base for Chelating P,P-Donor Ligands (LKB-PP)," *Organometallics* 31 (2012): 5302, https://doi.org/10.1021/om300312t.

223. N. Fey, M. F. Haddow, J. N. Harvey, C. L. McMullin, and A. G. Orpen, "A Ligand Knowledge Base for Carbenes (LKB-C): Maps of Ligand Space," *Dalton Transactions* 38, no. 39 (2009): 8183–8196, https://doi.org/10.1039/B909229C.

224. A. A. Schoepfer, R. Laplaza, M. D. Wodrich, J. Waser, and C. Corminboeuf, "Reaction-Agnostic Featurization of Bidentate Ligands for Bayesian Ridge Regression of Enantioselectivity," *ACS Catalysis* 14 (2024): 9302, https://doi.org/10.1021/acscatal.4c02452.

225. S. Gallarati, P. van Gerwen, R. Laplaza, S. Vela, A. Fabrizio, and C. Corminboeuf, "OSCAR: An Extensive Repository of Chemically and Functionally Diverse Organocatalysts," *Chemical Science* 13 (2022): 13782, https://doi.org/10.1039/D2SC04251G.

226. O. Pereira, M. Ruth, D. Gerbig, R. C. Wende, and P. R. Schreiner, "Leveraging Limited Experimental Data With Machine Learning: Differentiating a Methyl From an Ethyl Group in the Corey-Bakshi-Shibata Reduction," *Journal of the American Chemical Society* 146 (2024): 14576, https://doi.org/10.1021/jacs.4c01286.

227. J. A. Hueffel, T. Sperger, I. Funes-Ardoiz, J. S. Ward, K. Rissanen, and F. Schoenebeck, "Accelerated Dinuclear Palladium Catalyst Identification Through Unsupervised Machine Learning," *Science* 374 (2021): 1134, https://doi.org/10.1126/science.abj0999.

228. T. M. Karl, S. Bouayad-Gervais, J. A. Hueffel, et al., "Machine Learning-Guided Development of Trialkylphosphine Ni(I) Dimers and Applications in Site-Selective Catalysis," *Journal of the American Chemical Society* 145 (2023): 15414, https://doi.org/10.1021/jacs.3c03403.

229. J. A. Hueffel, M. Rigoulet, S. Wellig, et al., "Discovery of Ni(I) Complexes for CO2 Insertion Enabled by a Machine Learning-Computational-Selection Sequence," *Journal of the American Chemical Society* 147 (2025): 26149, https://doi.org/10.1021/jacs.5c00441.

230. P. Tiwary, L. Herron, R. John, S. Lee, D. Sanwal, and R. Wang, "Generative AI for Computational Chemistry: A Roadmap to Predicting Emergent Phenomena," *Proceedings of the National Academy of Sciences of the United States of America* 122 (2025): e2415655121, https://doi.org/10.1073/pnas.2415655121.

231. O. Schilter, A. C. Vaucher, P. Schwaller, and T. Laino, "Designing Catalysts With Deep Generative Models and Computational Data. A Case Study for Suzuki Cross Coupling Reactions," *Digital Discovery* 2 (2023): 728, https://doi.org/10.1039/D2DD00125J.

232. A. Kengkanna, Y. Kikuchi, T. Niwa, and M. Ohue, "Reaction-Conditioned Generative Model for Catalyst Design and Optimization With CatDRX," *Communications Chemistry* 8 (2025): 314, https://doi.org/10.1038/s42004-025-01732-7.

233. R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, et al., "Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules," *ACS Central Science* 4 (2018): 268, https://doi.org/10.1021/acscentsci.7b00572.

234. L. Schlosser, N. H. Rendel, J. Gemen, F. Glorius, and K. Jorner, "Inverse Molecular Design for the Discovery of Organic Energy Transfer Photocatalysts: Bridging Global and Local Chemical Space Exploration," *Journal of the American Chemical Society* 148 (2026): 6451, https://doi.org/10.1021/jacs.5c20087.

235. M. Olivecrona, T. Blaschke, O. Engkvist, and H. Chen, "Molecular De-Novo Design Through Deep Reinforcement Learning," *Journal of Cheminformatics* 9 (2017): 48, https://doi.org/10.1186/s13321-017-0235-x.

236. H. H. Loeffler, J. He, A. Tibo, et al., "Reinvent 4: Modern AI-Driven Generative Molecule Design," *Journal of Cheminformatics* 16 (2024): 20, https://doi.org/10.1186/s13321-024-00812-5.

237. M. Strandgaard, T. Linjordet, H. Kneiding, et al., "A Deep Generative Model for the Inverse Design of Transition Metal Ligands and Complexes," *JACS Au* 5 (2025): 2294, https://doi.org/10.1021/jacsau.5c00242.

238. W. Gao, S. Luo, and C. W. Coley, "Generative AI for Navigating Synthesizable Chemical Space," *Proceedings of the National Academy of Sciences of the United States of America* 122 (2025): e2415665122, https://doi.org/10.1073/pnas.2415665122.

239. S. M. Papidocha, A. Burger, V. Bernales, and A. Aspuru-Guzik, "The Elephant in the Lab: Synthesizability in Generative Small-Molecule Design″ Current Opinion," *Chemical Engineering* 51 (2026): 101217, https://doi.org/10.1016/j.coche.2025.101217.

240. J. L. Tyler, D. Trauner, and F. Glorius, "Reaction Development: A Student's Checklist," *Chemical Society Reviews* 54 (2025): 3272, https://doi.org/10.1039/d4cs01046a.

241. P. S. Gromski, A. B. Henson, J. M. Granda, and L. Cronin, "How to Explore Chemical Space Using Algorithms and Automation," *Nature Reviews Chemistry* 3 (2019): 119, https://doi.org/10.1038/s41570-018-0066-y.

242. M. H. S. Segler and M. P. Waller, "Modelling Chemical Reasoning to Predict and Invent Reactions," *Chemistry—A European Journal* 23 (2017): 6118, https://doi.org/10.1002/chem.201604556.

243. D. Caramelli, J. M. Granda, S. H. M. Mehr, D. Cambié, A. B. Henson, and L. Cronin, "Discovering New Chemistry With an Autonomous Robotic Platform Driven by a Reactivity-Seeking Neural Network," *ACS Central Science* 7 (2021): 1821, https://doi.org/10.1021/acscentsci.1c00435.

244. V. Dragone, V. Sans, A. B. Henson, J. M. Granda, and L. Cronin, "An Autonomous Organic Reaction Search Engine for Chemical Reactivity," *Nature Communications* 8 (2017): 15733, https://doi.org/10.1038/ncomms15733.

245. J. M. Granda, L. Donina, V. Dragone, D.-L. Long, and L. Cronin, "Controlling an Organic Synthesis Robot With Machine Learning to

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Search for New Reactivity," *Nature* 559 (2018): 377, https://doi.org/10.1038/s41586-018-0307-8.

246. E. P. Gajewska, S. Szymkuć, P. Dittwald, et al., "Algorithmic Discovery of Tactical Combinations for Advanced Organic Syntheses," *Chemistry* 6 (2020): 280, https://doi.org/10.1016/j.chempr.2019.11.016.

247. R. Roszak, L. Gadina, A. Wołos, et al., "Systematic, Computational Discovery of Multicomponent and One-Pot Reactions," *Nature Communications* 15 (2024): 10285, https://doi.org/10.1038/s41467-024-54611-5.

248. K. Molga, S. Szymkuć, P. Gołębiowska, et al., "A Computer Algorithm to Discover Iterative Sequences of Organic Reactions," *Nature Synthesis* 1 (2022): 49, https://doi.org/10.1038/s44160-021-00010-3.

249. T. Klucznik, L.-D. Syntrivanis, S. Baś, et al., "Computational Prediction of Complex Cationic Rearrangement Outcomes," *Nature* 625 (2024): 508, https://doi.org/10.1038/s41586-023-06854-3.

250. B. Mahjour, J. Lu, J. Fromer, N. Casetti, and C. W. Coley, ChemRxiv preprint (2024), 10.26434/chemrxiv-2024-qfjh9-v2.

251. A. F. Zahrt, Y. Mo, K. Y. Nandiwale, R. Shprints, E. Heid, and K. F. Jensen, "Machine-Learning-Guided Discovery of Electrochemical Reactions," *Journal of the American Chemical Society* 144 (2022): 22599, https://doi.org/10.1021/jacs.2c08997.

252. H. Wang, T. Fu, Y. Du, et al., "Scientific Discovery in the Age of Artificial Intelligence," *Nature* 620 (2023): 47, https://doi.org/10.1038/s41586-023-06221-2.

253. C. J. Taylor, A. Pomberger, K. C. Felton, et al., "A Brief Introduction to Chemical Reaction Optimization," *Chemical Reviews* 123 (2023): 3089, https://doi.org/10.1021/acs.chemrev.2c00798.

254. J. Rein, S. D. Rozema, O. C. Langner, et al., "Generality-Oriented Optimization of Enantioselective Aminoxyl Radical Catalysis," *Science* 380 (2023): 706, https://doi.org/10.1126/science.adf6177.

255. C. N. Prieto Kullmer, J. A. Kautzky, S. W. Krska, T. Nowak, S. D. Dreher, and D. W. C. MacMillan, "Accelerating Reaction Generality and Mechanistic Insight Through Additive Mapping," *Science* 376 (2022): 532, https://doi.org/10.1126/science.abn1885.

256. C. C. Wagen, S. E. McMinn, E. E. Kwan, and E. N. Jacobsen, "Screening for Generality in Asymmetric Catalysis," *Nature* 610 (2022): 680, https://doi.org/10.1038/s41586-022-05263-2.

257. I. O. Betinol, J. Lai, S. Thakur, and J. P. Reid, "A Data-Driven Workflow for Assigning and Predicting Generality in Asymmetric Catalysis," *Journal of the American Chemical Society* 145 (2023): 12870, https://doi.org/10.1021/jacs.3c03989.

258. S. L. Sivilotti, D. M. Friday, and N. E. Jackson, "Active Learning High Coverage Sets of Complementary Reaction Conditions," *Digital Discovery* 4 (2025): 846, https://doi.org/10.1039/D4DD00365A.

259. T. Seo, D. Kim, S. Ham, et al., "Generality-Driven Optimization of Enantio- and Regioselective Mono-Reduction of 1,2-Dicarbonyls by High-Throughput Experimentation and Machine Learning," *Angewandte Chemie International Edition* 65 (2026): e19425, https://doi.org/10.1002/anie.202519425.

260. S. B. Zacate, J. A. Dantas, S. Lin, A. G. Doyle, and M. S. Sigman, "Considerations in Pursuing Reaction Scope Generality," *Angewandte Chemie International Edition* 64, no. 41 (2025): e202511091, https://doi.org/10.1002/anie.202511091.

261. P. M. Murray, F. Bellany, L. Benhamou, D.-K. Bučar, A. B. Tabor, and T. D. Sheppard, "The Application of Design of Experiments (DoE) Reaction Optimisation and Solvent Selection in the Development of New Synthetic Chemistry," *Organic & biomolecular chemistry* 14 (2016): 2373, https://doi.org/10.1039/c5ob01892g.

262. B. J. Wall, M. T. Koeritz, L. M. Stanley, and B. VanVeller, "A Practical Start-Up Guide for Synthetic Chemists to Implement Design of Experiments (DoE)," *ACS Catalysis* 15 (2025): 8885, https://doi.org/10.1021/acscatal.5c01626.

263. S. A. Weissman and N. G. Anderson, "Design of Experiments (DoE) and Process Optimization. A Review of Recent Publications," *Organic Process Research & Development* 19 (2015): 1605, https://doi.org/10.1021/op500169m.

264. Y. Wu, A. Walsh, and A. M. Ganose, "Race to the Bottom: Bayesian Optimisation for Chemical Problems," *Digital Discovery* 3 (2024): 1086, https://doi.org/10.1039/D3DD00234A.

265. J. Guo, B. Ranković, and P. Schwaller, "Bayesian Optimization for Chemical Reactions," *Chimia* 77 (2023): 31, https://doi.org/10.2533/chimia.2023.31.

266. S. Desimpel, M. Dorbec, K. M. van Geem, and C. V. Stevens, "Bayesian Optimization for Chemical Reactions," *Chemical Society Reviews* 55, no. 5 (2026): 2731–2775, https://doi.org/10.1039/d5cs00962f.

267. P. I. Frazier, arXiv preprint (2018), 10.48550/arxiv.1807.02811.

268. B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, "Taking the Human out of the Loop: A Review of Bayesian Optimization," *Proceedings of the IEEE* 104 (2016): 148, https://doi.org/10.1109/JPROC.2015.2494218.

269. Y. Han, M. Deng, K. Liu, et al., "Computer-Aided Synthesis Planning (CASP) and Machine Learning: Optimizing Chemical Reaction Conditions," *Chemistry—A European Journal* 30 (2024): e202401626, https://doi.org/10.1002/chem.202401626.

270. A. Senthil Vel, D. Cortés-Borda, and F.-X. Felpin, "A Chemist's Guide to Multi-Objective Optimization Solvers for Reaction Optimization," *Reaction Chemistry & Engineering* 9 (2024): 2882, https://doi.org/10.1039/D4RE00175C.

271. F. Häse, L. M. Roch, and A. Aspuru-Guzik, "Chimera: Enabling Hierarchy Based Multi-Objective Optimization for Self-Driving Laboratories," *Chemical Science* 9 (2018): 7642, https://doi.org/10.1039/c8sc02239a.

272. F. Häse, L. M. Roch, C. Kreisbeck, and A. Aspuru-Guzik, "Phoenics: A Bayesian Optimizer for Chemistry," *ACS Central Science* 4 (2018): 1134, https://doi.org/10.1021/acscentsci.8b00307.

273. R. J. Hickman, M. Aldeghi, F. Häse, and A. Aspuru-Guzik, "Bayesian Optimization With Known Experimental and Design Constraints for Chemistry Applications," *Digital Discovery* 1 (2022): 732, https://doi.org/10.1039/D2DD00028H.

274. F. Häse, M. Aldeghi, R. J. Hickman, L. M. Roch, and A. Aspuru-Guzik, "Gryffin: An Algorithm for Bayesian Optimization of Categorical Variables Informed by Expert Knowledge," *Applied Physics Reviews* 8, no. 3 (2021): 031406, https://doi.org/10.1063/5.0048164.

275. B. J. Shields, J. Stevens, J. Li, et al., "Bayesian Reaction Optimization as a Tool for Chemical Synthesis," *Nature* 590 (2021): 89, https://doi.org/10.1038/s41586-021-03213-y.

276. J. A. G. Torres, S. H. Lau, P. Anchuri, et al., "A Multi-Objective Active Learning Platform and Web App for Reaction Optimization," *Journal of the American Chemical Society* 144 (2022): 19999, https://doi.org/10.1021/jacs.2c08592.

277. R. J. Hickman, M. Sim, S. Pablo-García, et al., "Atlas: A Brain for Self-Driving Laboratories," *Digital Discovery* 4 (2025): 1006, https://doi.org/10.1039/D4DD00115J.

278. S. P. Schmid, E. M. Rajaonson, C. T. Ser, et al., arXiv preprint (2025), 10.48550/arXiv.2502.18966.

279. J. P. Dürholt, T. S. Asche, J. Kleinekorte, et al., arXiv preprint (2024), 10.48550/arxiv.2408.05040v1.

280. S. G. Baird, A. R. Falkowski, and T. D. Sparks, "Honegumi: An Interface for Accelerating the Adoption of Bayesian Optimization in the Experimental Sciences," *npj Computational Materials* (2026), https://doi.org/10.1038/s41524-026-02156-0.

281. D. Reker, E. A. Hoyt, G. J. Bernardes, and T. Rodrigues, "Adaptive Optimization of Chemical Reactions With Minimal Experimental Information," *Cell Reports Physical Science* 1 (2020): 100247, https://doi.org/10.1016/j.xcrp.2020.100247.

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

282. K. C. Felton, J. G. Rittig, and A. A. Lapkin, "Summit: Benchmarking Machine Learning Methods for Reaction Optimisation," *Chemistry—Methods* 1 (2021): 116, https://doi.org/10.1002/cmtd.202000051.

283. N. H. Angello, V. Rathore, W. Beker, et al., "Closed-Loop Optimization of General Reaction Conditions for Heteroaryl Suzuki-Miyaura Coupling," *Science* 378 (2022): 399, https://doi.org/10.1126/science.adc8743.

284. J. Y. Wang, J. M. Stevens, S. K. Kariofillis, et al., "Identifying General Reaction Conditions by Bandit Optimization," *Nature* 626 (2024): 1025, https://doi.org/10.1038/s41586-024-07021-y.

285. Q. Jiang, M. Tian, J. Liu, and J. Mo, "BiDir-GRCO: A Bidirectional General Reaction Conditions Optimization Framework Integrating Multi-Armed Bandit and Regression Model," *ACS Omega* 10 (2025): 36733, https://doi.org/10.1021/acsomega.5c06476.

286. S. Gallarati, P. van Gerwen, R. Laplaza, L. Brey, A. Makaveev, and C. Corminboeuf, "A Genetic Optimization Strategy With Generality in Asymmetric Organocatalysis as a Primary Target," *Chemical Science* 15 (2024): 3640, https://doi.org/10.1039/d3sc06208b.

287. S.-W. Li, S. Chen, J. C. A. Oliveira, S.-Q. Zhang, L. Ackermann, and X. Hong, "Staged Diversity-Constrained Machine Learning for High-Dimensional Reaction Condition Optimization," *Angewandte Chemie International Edition* 65, no. 17 (2026): e4418883, https://doi.org/10.1002/anie.4418883.

288. E. Shim, J. A. Kammeraad, Z. Xu, A. Tewari, T. Cernak, and P. M. Zimmerman, "Predicting Reaction Conditions From Limited Data Through Active Transfer Learning," *Chemical Science* 13 (2022): 6655, https://doi.org/10.1039/d1sc06932b.

289. G. Tom, S. P. Schmid, S. G. Baird, et al., "Self-Driving Laboratories for Chemistry and Materials Science," *Chemical Reviews* 124 (2024): 9633, https://doi.org/10.1021/acs.chemrev.4c00055.

290. S. X. Leong, C. E. Griesbach, R. Zhang, et al., "Steering Towards Safe Self-Driving Laboratories," *Nature Reviews Chemistry* 9 (2025): 707, https://doi.org/10.1038/s41570-025-00747-x.

291. A. Slattery, Z. Wen, P. Tenblad, et al., "Automated Self-Optimization, Intensification, and Scale-Up of Photocatalysis in Flow," *Science* 383 (2024): eadj1817, https://doi.org/10.1126/science.adj1817.

292. A. Tripp and J. M. Hernández-Lobato, "Diagnosing and Fixing Common Problems in Bayesian Optimization for Molecule Design," arXiv preprint (2024), arXiv:2406.07709.

293. F. Schäfer, L. Lückemeier, and F. Glorius, "Improving Reproducibility Through Condition-Based Sensitivity Assessments: Application, Advancement and Prospect," *Chemical Science* 15 (2024): 14548, https://doi.org/10.1039/d4sc03017f.

294. L. Pitzer, F. Schäfers, and F. Glorius, "Rapid Assessment of the Reaction-Condition-Based Sensitivity of Chemical Transformations," *Angewandte Chemie International Edition* 58 (2019): 8572, https://doi.org/10.1002/anie.201901935.

295. C. P. Haas, M. Lübbesmeyer, E. H. Jin, et al., "Open-Source Chromatographic Data Analysis for Reaction Optimization and Screening," *ACS Central Science* 9 (2023): 307, https://doi.org/10.1021/acscentsci.2c01042.

296. J. Obořil, C. P. Haas, M. Lübbesmeyer, et al., "Automated Processing of Chromatograms: A Comprehensive Python Package With a GUI for Intelligent Peak Identification and Deconvolution in Chemical Reaction Analysis," *Digital Discovery* 3 (2024): 2041, https://doi.org/10.1039/D4DD00214H.

297. G. Chen, M. Fleck, and T. Stuyver, "Leveraging Chemical Hidden-Space Representations Effectively in Bayesian Optimization for Experiment Design Through Dimension-Aware Hyperpriors," *Journal of Chemical Theory and Computation* 22, no. 11 (2026): 5594–5608, https://doi.org/10.1021/acs.jctc.6c00251.

298. M. Malu, G. Dasarathy, and A. Spanias, "Bayesian Optimization in High-Dimensional Spaces: A Brief Survey," "in 2021 12th International Conference on Information, Intelligence, Systems & Applications (IISA)," (2021): 1, https://doi.org/10.1109/IISA52424.2021.9555522.

299. L. Papenmeier, M. Poloczek, and L. Nardi, "Understanding High-Dimensional Bayesian Optimization," in Proceedings of the 42nd International Conference on Machine Learning, PMLR 267.

300. G. Tom, S. Lo, S. Corapi, A. Aspuru-Guzik, and B. Sanchez-Lengeling, "Ranking Over Regression for Bayesian Optimization and Molecule Selection," *APL Machine Learning* 3, no. 3 (2025): 036113, https://doi.org/10.1063/5.0272663.

301. R. Schlama, J. W. Sin, R. P. Burwood, K. Püntener, R. Bigler, and P. Schwaller, "Swarm Intelligence for Chemical Reaction Optimization," *Chem* (2026): 103035, https://doi.org/10.1016/j.chempr.2026.103035.

302. F. Boser, J. Spies, and F. Glorius, "ChemRxiv preprint," (2025), 10.26434/chemrxiv-2025-hq4rx.

303. M. C. Kozlowski, "On the Topic of Substrate Scope," *Organic Letters* 24 (2022): 7247, https://doi.org/10.1021/acs.orglett.2c03246.

304. D. Rana, P. M. Pflüger, N. P. Hölter, G. Tan, and F. Glorius, "Standardizing Substrate Selection: A Strategy Toward Unbiased Evaluation of Reaction Generality," *ACS Central Science* 10 (2024): 899, https://doi.org/10.1021/acscentsci.3c01638.

305. T. Gensch and F. Glorius, "The Straight Dope on the Scope of Chemical Reactions," *Science* 352 (2016): 294, https://doi.org/10.1126/science.aaf3539.

306. W. Gao, P. Raghavan, R. Shprints, and C. W. Coley, "Revealing the Relationship Between Publication Bias and Chemical Reactivity With Contrastive Learning," *Journal of the American Chemical Society* 147 (2025): 8959, https://doi.org/10.1021/jacs.5c01120.

307. K. D. Collins and F. Glorius, "A Robustness Screen for the Rapid Assessment of Chemical Reactions," *Nature Chemistry* 5 (2013): 597, https://doi.org/10.1038/nchem.1669.

308. K. D. Collins, A. Rühling, and F. Glorius, "Application of a Robustness Screen for the Evaluation of Synthetic Organic Methodology," *Nature Protocols* 9 (2014): 1348, https://doi.org/10.1038/nprot.2014.076.

309. P. S. Kutchukian, J. F. Dropinski, K. D. Dykstra, et al., "Chemistry Informer Libraries: A Chemoinformatics Enabled Approach to Evaluate and Advance Synthetic Methods," *Chemical Science* 7 (2016): 2604, https://doi.org/10.1039/c5sc04751j.

310. E. N. Bess, A. J. Bischoff, and M. S. Sigman, "Designer Substrate Library for Quantitative, Predictive Modeling of Reaction Performance," *Proceedings of the National Academy of Sciences of the United States of America* 111 (2014): 14698, https://doi.org/10.1073/pnas.1409522111.

311. S. K. Kariofillis, S. Jiang, A. M. Żurański, S. S. Gandhi, J. I. Martinez Alvarado, and A. G. Doyle, "Using Data Science To Guide Aryl Bromide Substrate Scope Analysis in a Ni/Photoredox-Catalyzed Cross-Coupling With Acetals as Alcohol-Derived Radical Sources," *Journal of the American Chemical Society* 144 (2022): 1045, https://doi.org/10.1021/jacs.1c12203.

312. S. Felten, C. Q. He, M. Weisel, M. Shevlin, and M. H. Emmert, "Accessing Diverse Azole Carboxylic Acid Building Blocks via Mild C-H Carboxylation: Parallel, One-Pot Amide Couplings and Machine-Learning-Guided Substrate Scope Design," *Journal of the American Chemical Society* 144 (2022): 23115, https://doi.org/10.1021/jacs.2c10557.

313. B. C. Haas, A. E. Goetz, A. Bahamonde, J. C. McWilliams, and M. S. Sigman, "Predicting Relative Efficiency of Amide Bond Formation Using Multivariate Linear Regression," *Proceedings of the National Academy of Sciences of the United States of America* 119 (2022): e2118451119, https://doi.org/10.1073/pnas.2118451119.

314. L. V. Hoff, G. A. Chesnokov, A. Linden, and K. Gademann, "Mechanistic Studies and Data Science-Guided Exploration of Bromotetrazine Cross-Coupling," *ACS Catalysis* 12 (2022): 9226, https://doi.org/10.1021/acscatal.2c01813.

315. L. van Dijk, B. C. Haas, N.-K. Lim, et al., "Data Science-Enabled Palladium-Catalyzed Enantioselective Aryl-Carbonylation of

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Sulfonimidamides," *Journal of the American Chemical Society* 145 (2023): 20959, https://doi.org/10.1021/jacs.3c06674.

316. I. N.-M. Leibler, S. S. Gandhi, M. A. Tekle-Smith, and A. G. Doyle, "Strategies for Nucleophilic C(sp3)-(Radio)Fluorination," *Journal of the American Chemical Society* 145 (2023): 9928, https://doi.org/10.1021/jacs.3c01824.

317. T. Tang, A. Hazra, D. S. Min, et al., "Interrogating the Mechanistic Features of Ni(I)-Mediated Aryl Iodide Oxidative Addition Using Electroanalytical and Statistical Modeling Techniques," *Journal of the American Chemical Society* 145, no. 15 (2023): 8689–8699, https://doi.org/10.1021/jacs.3c01726.

318. S. Deolka, D. Dalmau, A. Garcia Roca, S. Vasylevskyi, M. H. Samha, and M. S. Sigman, "Development of a Heteroleptic Palladium Catalyst for the Fluorination of Arylboronic Acids Using Data Science Tools," *Journal of the American Chemical Society* 147 (2025): 35705, https://doi.org/10.1021/jacs.5c11929.

319. K. Ichizawa, T. Nishii, and H. Gotoh, "Substrate Scope in Organic Reactions: Comparison of Fingerprint-Based and Reactivity-Based Similarity Metrics," *Journal of Chemical Information and Modeling* 66 (2026): 3878, https://doi.org/10.1021/acs.jcim.5c03167.

320. A. G. Beck, J. Fine, Y.-H. Lam, E. C. Sherer, E. L. Regalado, and P. Aggarwal, "Dedenser: A Python Package for Clustering and Downsampling Chemical Libraries," *Journal of Chemical Information and Modeling* 65 (2025): 1053, https://doi.org/10.1021/acs.jcim.4c01980.

321. N. S. Eyke, W. H. Green, and K. F. Jensen, "Iterative Experimental Design Based on Active Machine Learning Reduces the Experimental Burden Associated With Reaction Screening," *Reaction Chemistry & Engineering* 5 (2020): 1963, https://doi.org/10.1039/D0RE00232A.

322. S. Viet Johansson, H. Gummesson Svensson, E. Bjerrum, et al., "Using Active Learning to Develop Machine Learning Models for Reaction Yield Prediction," *Molecular Informatics* 41 (2022): e2200043, https://doi.org/10.1002/minf.202200043.

323. L. W. Souza, N. D. Ricke, B. C. Chaffin, et al., "Applying Active Learning Toward Building a Generalizable Model for Ni-Photoredox Cross-Electrophile Coupling of Aryl and Alkyl Bromides," *Journal of the American Chemical Society* 147 (2025): 18747, https://doi.org/10.1021/jacs.5c02218.

324. S. Roediger, M. Sigman, and A. G. Doyle, ChemRxiv preprint (2025), 10.26434/chemrxiv-2025-r0sst.

325. J. Li, X. Xiao, Q. Yang, B. Zhao, and S. Luo, "ScopeMap: An AI-Assisted, Human-in-the-Loop Workflow for Mapping Reaction Scope and Boundaries," *Angewandte Chemie International Edition* 65 (2026): e2455429 , https://doi.org/10.1002/anie.2455429.

326. S. Lin, S. Dikler, W. D. Blincoe, et al., "Mapping the Dark Space of Chemical Reactions With Extended Nanomole Synthesis and MALDI-TOF MS," *Science* 361 (2018): eaar6236, https://doi.org/10.1126/science.aar6236.

327. B. Mahjour, Y. Shen, and T. Cernak, "Ultrahigh-Throughput Experimentation for Information-Rich Chemical Synthesis," *Accounts of Chemical Research* 54 (2021): 2337, https://doi.org/10.1021/acs.accounts.1c00119.

328. D.-H. Liu, P. M. Pflüger, A. Outlaw, et al., "Late-Stage Saturation of Drug Molecules," *Journal of the American Chemical Society* 146 (2024): 11866, https://doi.org/10.1021/jacs.4c00807.

329. A. M. Żurański, S. S. Gandhi, and A. G. Doyle, "A Machine Learning Approach to Model Interaction Effects: Development and Application to Alcohol Deoxyfluorination," *Journal of the American Chemical Society* 145 (2023): 7898, https://doi.org/10.1021/jacs.2c13093.

330. I. L. Eshel, B. Shahar, S. Barranco, M. Pérez-Temprano, and A. Milo, ChemRxiv preprint (2025), 10.26434/chemrxiv-2025-z0djf.

331. P. Neves, B. Hao, S. Aikonen, et al., "Robust Out-of-Distribution Prediction of Buchwald–Hartwig Reactions," *Nature Computational Science* (2026), https://www.nature.com/articles/s43588-026-01017-6 .

332. K. Jorner, A. Tomberg, C. Bauer, C. Sköld, and P.-O. Norrby, "Organic Reactivity From Mechanism to Machine Learning," *Nature Reviews Chemistry* 5 (2021): 240, https://doi.org/10.1038/s41570-021-00260-x.

333. S. C. Smith, C. S. Horbaczewskyj, T. F. N. Tanner, J. J. Walder, and I. J. S. Fairlamb, "Automated Approaches, Reaction Parameterisation, and Data Science in Organometallic Chemistry and Catalysis: Towards Improving Synthetic Chemistry and Accelerating Mechanistic Understanding," *Digital Discovery* 3 (2024): 1467, https://doi.org/10.1039/D3DD00249G.

334. G. S. Tâmega, M. O. Costa, A. de Araujo Pereira, and M. A. Barbosa Ferreira, "Data Science Guiding Analysis of Organic Reaction Mechanism and Prediction," *Chemical Record* 24 (2024): e202400148, https://doi.org/10.1002/tcr.202400148.

335. P. R. Wells, "Linear Free Energy Relationships," *Chemical Reviews* 63 (1963): 171, https://doi.org/10.1021/cr60222a005.

336. L. P. Hammett, "The Effect of Structure Upon the Reactions of Organic Compounds. Benzene Derivatives," *Journal of the American Chemical Society* 59 (1937): 96, https://doi.org/10.1021/ja01280a022.

337. J. M. Crawford, C. Kingston, F. D. Toste, and M. S. Sigman, "Data Science Meets Physical Organic Chemistry," *Accounts of Chemical Research* 54, no. 16 (2021): 3136–3148, https://doi.org/10.1021/acs.accounts.1c00285.

338. S. Lundberg and S.-I. Lee, arXiv preprint (2017), 10.48550/arXiv.1705.07874.

339. A. Hoque and R. B. Sunoj, "Deep Learning for Enantioselectivity Predictions in Catalytic Asymmetric β-C–H Bond Activation Reactions," *Digital Discovery* 1 (2022): 926, https://doi.org/10.1039/D2DD00084A.

340. K. C. Harper, E. N. Bess, and M. S. Sigman, "Multidimensional Steric Parameters in the Analysis of Asymmetric Catalytic Reactions," *Nature Chemistry* 4 (2012): 366, https://doi.org/10.1038/nchem.1297.

341. S. H. Newman-Stonebraker, S. R. Smith, J. E. Borowski, et al., "Univariate Classification of Phosphine Ligation state and Reactivity in Cross-Coupling Catalysis," *Science* 374 (2021): 301, https://doi.org/10.1126/science.abj4213.

342. M. Lavin and M. A. Tekle-Smith, "Data-Driven Approach to Understanding Tetrabutylammonium Decatungstate-Catalyzed C(sp3)-H Functionalization Selectivity," *Angewandte Chemie International Edition* 65, no. 29 (2026): e8014515, https://doi.org/10.1002/anie.8014515.

343. D. S. Levine, M. Shuaibi, E. W. C. Spotte-Smith, et al., arXiv preprint (2025), 10.48550/arXiv.2505.08762v1.

344. B. M. Wood, M. Dzamba, X. Fu, et al., arXiv preprint (2025), 10.48550/arXiv.2506.23971.

345. J. S. Smith, O. Isayev, and A. E. Roitberg, "ANI-1: An Extensible Neural Network Potential With DFT Accuracy at Force Field Computational Cost," *Chemical Science* 8 (2017): 3192, https://doi.org/10.1039/c6sc05720a.

346. S. Zhang, R. Zubatyuk, Y. Yang, A. Roitberg, and O. Isayev, "ANI-1xBB: An ANI-Based Reactive Potential for Small Organic Molecules," *Journal of Chemical Theory and Computation* 21 (2025): 4365, https://doi.org/10.1021/acs.jctc.5c00347.

347. R. J. Urquhart, A. van Teijlingen, and T. Tuttle, "Application of Neural Network Potentials to Modelling Transition States," *Chemical Communications* 61 (2025): 11810, https://doi.org/10.1039/d5cc02090e.

348. A. Denzel and J. Kästner, "Gaussian Process Regression for Transition State Search," *Journal of Chemical Theory and Computation* 14 (2018): 5777, https://doi.org/10.1021/acs.jctc.8b00708.

349. J. de Landsheere, M. P.-P. Kovar, K. Mark, et al., "Toward on-the-Fly Prediction of Reaction Energetics for High-Throughput Screening," *Chemical Physics Reviews* 7, no. 2 (2026): 021310, https://doi.org/10.1063/5.0321836.

350. R. Barrett and J. Westermayr, "Reinforcement Learning for Traversing Chemical Structure Space: Optimizing Transition States and Minimum Energy Paths of Molecules," *Journal of Physical Chemistry Letters* 15 (2024): 349, https://doi.org/10.1021/acs.jpclett.3c02771.

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

351. X. Wang, Y. Mao, and Z. Wang, "Machine Learning Approaches for Transition state Prediction," *Chem Catalysis* 5 (2025): 101458, https://doi.org/10.1016/j.checat.2025.101458.

352. S. Kim, J. Woo, and W. Y. Kim, "Diffusion-Based Generative AI for Exploring Transition States From 2D Molecular Graphs," *Nature Communications* 15 (2024): 341, https://doi.org/10.1038/s41467-023-44629-6.

353. C. A. Grambow, L. Pattanaik, and W. H. Green, "Deep Learning of Activation Energies," *Journal of Physical Chemistry Letters* 11 (2020): 2992, https://doi.org/10.1021/acs.jpclett.0c00500.

354. S. Darouich, J. W. Toney, W. Luo, J. Kästner, M. Niepert, and H. J. Kulik, arXiv preprint (2026), 10.48550/arxiv.2601.16469.

355. J. F. Joung, M. H. Fong, N. Casetti, J. P. Liles, N. S. Dassanayake, and C. W. Coley, "Electron Flow Matching for Generative Reaction Mechanism Prediction," *Nature* 645 (2025): 115, https://doi.org/10.1038/s41586-025-09426-9.

356. P. Tuo, J. Chen, and J. Li, "Flow Matching for Reaction Pathway Generation," *Nature Communications* 17 (2026): 8769, https://doi.org/10.1038/s41467-026-75654-w.

357. T. A. Neukomm, Z. Joncev, and P. Schwaller, arXiv preprint (2025), 10.48550/arXiv.2512.05722.

358. J. Bradshaw, M. J. Kusner, B. Paige, M. H. S. Segler, and J. M. Hernández-Lobato, arXiv preprint (2018), 10.48550/arxiv.1805.10970.

359. M. Das, A. Hoque, M. Baranwal, and R. B. Sunoj, arXiv preprint (2025), 10.48550/arxiv.2509.15872.

360. J. F. Joung, M. H. Fong, J. Roh, Z. Tu, J. Bradshaw, and C. W. Coley, "Reproducing Reaction Mechanisms With Machine-Learning Models Trained on a Large-Scale Mechanistic Dataset," *Angewandte Chemie International Edition* 63 (2024): e202411296, https://doi.org/10.1002/anie.202411296.

361. M. Tavakoli, Y. T. T. Chiu, A. Shmakov, A. M. Carlton, D. van Vranken, and P. Baldi, "AI for Interpretable Chemistry: Predicting Radical Mechanistic Pathways via Contrastive Learning," *Advances in Neural Information Processing Systems* 36 (2023): 4080–4096, https://doi.org/10.52202/075280-0178.

362. M. A. Kayala and P. Baldi, "ReactionPredictor: Prediction of Complex Chemical Reactions at the Mechanistic Level Using Machine Learning," *Journal of Chemical Information and Modeling* 52 (2012): 2526, https://doi.org/10.1021/ci3003039.

363. B. B. Hoar, W. Zhang, Y. Chen, et al., "Redox-Detecting Deep Learning for Mechanism Discernment in Cyclic Voltammograms of Multiple Redox Events," *ACS Electrochemistry* 1 (2025): 52, https://doi.org/10.1021/acselectrochem.4c00014.

364. H. Sheng, J. Sun, O. Rodríguez, et al., "Autonomous Closed-Loop Mechanistic Investigation of Molecular Electrochemistry via Automation," *Nature Communications* 15 (2024): 2781, https://doi.org/10.1038/s41467-024-47210-x.

365. D. A. Rosser and K. C. Leonard, "High-Speed Cyclic Voltammetry Regressions Using Machine Learning," *ACS Electrochemistry* 1 (2025): 1038, https://doi.org/10.1021/acselectrochem.5c00012.

366. J. Burés and I. Larrosa, "Organic Reaction Mechanism Classification Using Machine Learning," *Nature* 613 (2023): 689, https://doi.org/10.1038/s41586-022-05639-4.

367. D. G. Blackmond, "Reaction Progress Kinetic Analysis: A Powerful Methodology for Mechanistic Studies of Complex Catalytic Reactions," *Angewandte Chemie International Edition* 44 (2005): 4302, https://doi.org/10.1002/anie.200462544.

368. D. G. Blackmond, "Kinetic Profiling of Catalytic Organic Reactions as a Mechanistic Tool," *Journal of the American Chemical Society* 137 (2015): 10852, https://doi.org/10.1021/jacs.5b05841.

369. OpenAI, "Introducing ChatGPT" Can be Found under https://openai.com/index/chatgpt/(accessed: 24.02.2026).

370. R. MacKnight, D. A. Boiko, J. E. Regio, L. C. Gallegos, T. A. Neukomm, and G. Gomes, "Rethinking Chemical Research in the Age of Large Language Models," *Nature Computational Science* 5 (2025): 715, https://doi.org/10.1038/s43588-025-00811-y.

371. K. K. L. Tharwani, R. Kumar, Sumita, N. Ahmed, and Y. Tang, arXiv preprint (2025), https://doi.org/10.48550/arXiv.2508.05427.

372. M. C. Ramos, C. J. Collison, and A. D. White, "A Review of Large Language Models and Autonomous Agents in Chemistry," *Chemical Science* 16 (2025): 2514, https://doi.org/10.1039/D4SC03921A.

373. A. M. Bran and P. Schwaller, *Drug Development Supported by Informatics*, ed. H. Satoh, K. Funatsu, and H. Yamamoto, (Springer Nature, 2024), 143–163.

374. Y. Su, X. Wang, Y. Ye, et al., "Automation and Machine Learning Augmented by Large Language Models in a Catalysis Study," *Chemical Science* 15 (2024): 12200, https://doi.org/10.1039/D3SC07012C.

375. A. M. Mroz, A. R. Basford, F. Hastedt, et al., "Cross-Disciplinary Perspectives on the Potential for Artificial Intelligence Across Chemistry," *Chemical Society Reviews* 54 (2025): 5433, https://doi.org/10.1039/D5CS00146C.

376. N. Janakarajan, T. Erdmann, S. Swaminathan, T. Laino, and J. Born in *Drug Development Supported by Informatics*, ed. H. Satoh, K. Funatsu, and H. Yamamoto, (Springer Nature, 2024), 121–141.

377. J. Wei, Y. Yang, X. Zhang, et al., arXiv preprint (2025), https://doi.org/10.48550/arXiv.2508.14111.

378. X. Qu, A. Damoah, J. Sherwood, et al., arXiv preprint (2025), https://doi.org/10.48550/arXiv.2508.11957.

379. S. Gao, R. Zhu, P. Sui, et al., arXiv preprint (2025), https://doi.org/10.48550/arXiv.2509.23426.

380. K. Yue, Q. Zhou, R. Bird, et al., "Trends and Opportunities in Organic Synthesis: Global State of Research Metrics and Advances in Precision, Efficiency, and Green Chemistry," *Journal of Organic Chemistry* 88 (2023): 4031, https://doi.org/10.1021/acs.joc.2c03057.

381. Q. Ai, F. Meng, J. Shi, B. Pelkie, and C. W. Coley, "Extracting Structured Data From Organic Synthesis Procedures Using a Fine-Tuned Large Language Model," *Digital Discovery* 3 (2024): 1822, https://doi.org/10.1039/D4DD00091A.

382. W. Zhang, Q. Wang, X. Kong, et al., "Fine-Tuning Large Language Models for Chemical Text Mining," *Chemical Science* 15 (2024): 10600, https://doi.org/10.1039/D4SC00924J.

383. S. Pagel, M. Jirasek, and L. Cronin, "Verification and Execution of the Scientific Literature via Chemputation Augmented by Large Language Models," *Communications Chemistry* 9 (2026): 191, https://doi.org/10.1038/s42004-026-01993-w.

384. J. Li, M. Li, Q. Yang, and S. Luo, "ReactionSeek: LLM-Powered Literature Data Mining and Knowledge Discovery in Organic Synthesis," *Nature Communications* 17 (2026): 3356, https://doi.org/10.1038/s41467-026-70180-1.

385. S. Bae, M. Jeon, and H. R. Moon, "Text Mining in MOF Research: From Manual Curation to Large Language Model-Based Automation," *Chemical Communications* 61 (2025): 11083, https://doi.org/10.1039/D5CC02511G.

386. Z. Zheng, O. Zhang, C. Borgs, J. T. Chayes, and O. M. Yaghi, "ChatGPT Chemistry Assistant for Text Mining and the Prediction of MOF Synthesis," *Journal of the American Chemical Society* 145 (2023): 18048, https://doi.org/10.1021/jacs.3c05819.

387. S. X. Leong, S. Pablo-García, B. Wong, and A. Aspuru-Guzik, "MERMaid: Universal Multimodal Mining of Chemical Reactions From PDFs Using Vision-Language Models," *Matter* 8 (2025): 102331, https://doi.org/10.1016/j.matt.2025.102331.

388. Y. Qian, J. Guo, Z. Tu, C. W. Coley, and R. Barzilay, "RxnScribe: A Sequence Generation Model for Reaction Diagram Parsing," *Journal of*

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Chemical Information and Modeling 63 (2023): 4030, https://doi.org/10.1021/acs.jcim.3c00439.

389. M. Caldas Ramos, S. S. Michtavy, A. D. White, and M. D. Porosoff, "Bayesian Optimization of Catalysis With in-Context Learning," *ACS Central Science* 12 (2026): 599, https://doi.org/10.1021/acscentsci.5c02418.

390. B. Rankovic, R.-R. Griffiths, and P. Schwaller, arXiv preprint (2025), https://doi.org/10.48550/arXiv.2504.06265.

391. A. Cissé, X. Evangelopoulos, V. V. Gusev, and A. I. Cooper, "Language-Based Bayesian Optimization Research Assistant (BORA," in Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (2025), 4967, https://doi.org/10.24963/ijcai.2025/553.

392. Y. Zhang, Y. Han, S. Chen, et al., "Large Language Models to Accelerate Organic Chemistry Synthesis," *Nature Machine Intelligence* 7 (2025): 1010, https://doi.org/10.1038/s42256-025-01066-y.

393. R. MacKnight, J. E. Regio, J. G. Ethier, L. A. Baldwin, and G. Gomes, arXiv preprint (2025), https://doi.org/10.48550/arXiv.2509.00103v1.

394. A. Cissé, M. E. Cooper, M. Zhu, X. Evangelopoulos, and A. I. Cooper, "Can We Automate Scientific Reasoning in Closed-Loop Experiments Using Large Language Models?," *Digital Discovery* 5, no. 3 (2026): 1132–1160, https://doi.org/10.1039/D5DD00520E.

395. Z. Zhou, S. Veeramani, F. Munguia-Galeano, H. Fakhruldeen, and A. I. Cooper, "Localization, Inspection, and Reasoning (LIRA) Module for Autonomous Workflows in Self-Driving Laboratories," *Communications Chemistry* 8 (2025): 384, https://doi.org/10.1038/s42004-025-01770-1.

396. Y. Ruan, C. Lu, N. Xu, et al., "An Automatic End-to-End Chemical Synthesis Development Platform Powered by Large Language Models," *Nature Communications* 15 (2024): 10160, https://doi.org/10.1038/s41467-024-54457-x.

397. A. M. Bran, S. Cox, O. Schilter, C. Baldassari, A. D. White, and P. Schwaller, "Augmenting Large Language Models With Chemistry Tools," *Nature Machine Intelligence* 6 (2024): 525, https://doi.org/10.1038/s42256-024-00832-8.

398. D. A. Boiko, R. MacKnight, B. Kline, and G. Gomes, "Autonomous Chemical Research With Large Language Models," *Nature* 624 (2023): 570, https://doi.org/10.1038/s41586-023-06792-0.

399. Y. Kang, S. X. Leong, H. S. Park, et al., "El Agente Seguro: An Agent for Chemical Safety," Working Paper (2026), https://doi.org/10.26434/chemrxiv.15004378/v2.

400. Z. Wang, H. Huang, H. Zhao, et al., arXiv preprint (2025), https://doi.org/10.48550/arXiv.2507.14267.

401. F. Neese, "Software Update: The ORCA Program System—Version 6.0," *WIREs Computational Molecular Science* 15, no. 2 (2025): e70019, https://doi.org/10.1002/wcms.70019.

402. Y. Zou, A. H. Cheng, A. Aldossary, et al., "El Agente: An Autonomous Agent for Quantum Chemistry," *Matter* 8 (2025): 102263, https://doi.org/10.1016/j.matt.2025.102263.

403. S. G. H. Kumar, Y. Zou, A. Wang, et al., arXiv preprint (2026), https://doi.org/10.48550/arXiv.2602.17886..

404. C. Choi, Y. Zou, M. Müller, et al., arXiv preprint (2026), https://doi.org/10.48550/arXiv.2602.04849.

405. I. Gustin, L. M. Calderón, J. B. Pérez-Sánchez, et al., "El Agente Cuantico: Automating Quantum Simulations," *Reports on Progress in Physics* 89 (2026): 077602, https://doi.org/10.1088/1361-6633/ae8933.

406. J. Bai, A. Aldossary, T. Swanick, et al., "El Agente Gráfico: Structured Execution Graphs for Scientific Agents," arXiv preprint (2026), https://doi.org/10.48550/arXiv.2602.17902.

407. J. B. Pérez-Sánchez, Y. Zou, J. A. Campos-Gonzalez-Angulo, et al., "El Agente Quntur: A Research Collaborator Agent for Quantum Chemistry," arXiv preprint (2026), https://doi.org/10.48550/arXiv.2602.04850.

408. Z. Zhang, A. Yin, A. Baweja, et al., "El Agente Forjador: Task-Driven Agent Generation for Quantum Simulation," arXiv preprint (2026), https://doi.org/10.48550/arXiv.2604.14609.

409. A. T. Kalai, O. Nachum, S. S. Vempala, and E. Zhang, "Why Language Models Hallucinate," arXiv preprint (2025), https://doi.org/10.48550/arXiv.2509.04664.

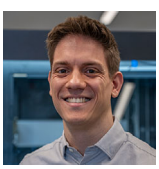
410. Malikussaid, H. H. Nuha, and I. Kurniawan, "VALID-Mol: A Systematic Framework for Validated LLM-Assisted Molecular Design," arXiv preprint (2025), https://doi.org/10.48550/arXiv.2506.23339.

411. A. Mirza, N. Alampara, S. Kunchapu, et al., "A Framework for Evaluating the Chemical Knowledge and Reasoning Abilities of Large Language Models Against the Expertise of Chemists," *Nature Chemistry* 17 (2025): 1027, https://doi.org/10.1038/s41557-025-01815-x.

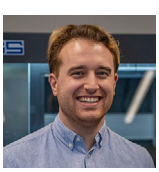
Biographies



Niklas Hölter received his BSc and MSc degrees in chemistry from the University of Münster. During his MSc, he conducted international research stays at Stockholm University (Prof. Kálmán Szabó), The University of Tokyo (Prof. Motomu Kanai) and University of California, Los Angeles (Prof. Abigail Doyle). Moreover, Niklas spent a 6-month internship in the reaction optimization department of Bayer AG. His doctoral studies in the group of Prof. Frank Glorius focus on the interface of (photo-)catalysis, data science and lab automation.



Felix Katzenburg is a postdoctoral fellow in the group of Alán Aspuru-Guzik and Varinia Bernales at the University of Toronto, working on AI for discovery and self-driving labs. Felix studied chemistry at the University of Münster, including a research stay at the Max-Planck-Institut für Kohlenforschung (with Prof. Josep Cornellà). He obtained his PhD under the supervision of Prof. Frank Glorius in 2026, after conducting research in molecular machine learning, reaction screening, and automation at the University of Münster and the University of Michigan (with Prof. Tim Cernak).



Florian Boser received his BSc in chemistry and biochemistry and his MSc in chemistry from Ludwig-Maximilians University (LMU) Munich. Following a research stay at the University of Glasgow (with Prof. Lee Cronin), he joined the group of Prof. Frank Glorius at the University of Münster for his master's thesis and PhD. His research focuses on the interface of organic synthesis, computer science, and lab automation, specializing in high-throughput experimentation, machine learning for reactivity prediction and agentic workflows.

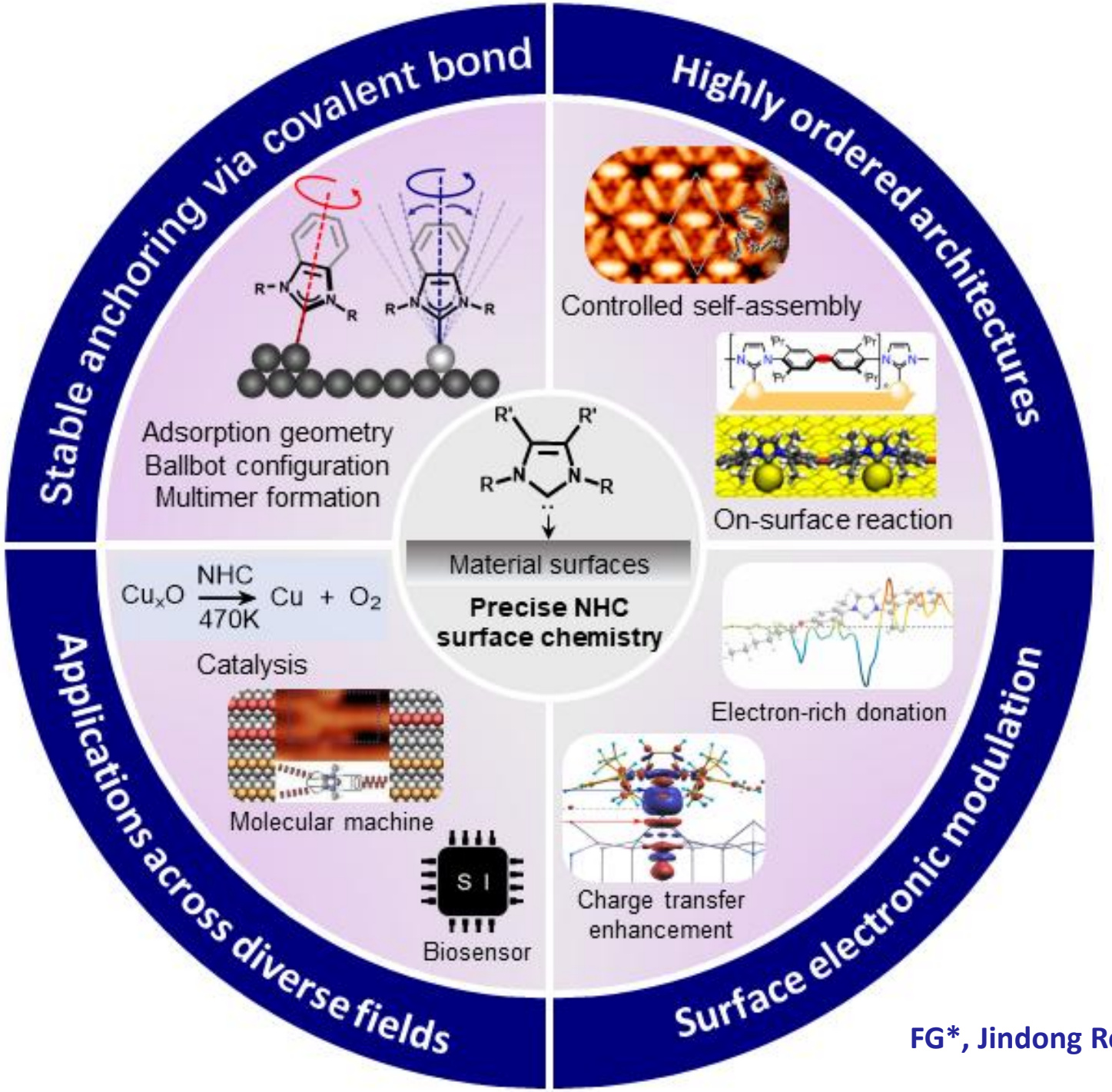


Frank Glorius is a full professor of organic chemistry at the University of Münster, Germany. His interdisciplinary research program focuses on new concepts for photocatalysis, C—H activation, as well as asymmetric heterogeneous catalysis, arene hydrogenation and organocatalysis. Moreover, his group is interested in functional molecules for surface modification, battery additives, or interfacing with biological membranes. Finally, his group develops data science and molecular machine learning tools with direct practical relevance to synthetic chemistry.

25109537, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anov.70032 by Universitäts- Und Landesbibliothek Münster, Wiley Online Library on [15/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License


②

Molecules are key to FUNCTION!!



FG*, Jindong Ren*: Chem. Soc. Rev. 2025.

Chem Soc Rev

TUTORIAL REVIEW

Published on 30 April 2025
Licence and permissions

Check for updates

Cite this: *Chem. Soc. Rev.*, 2025, **54**, 4626

Insights into the surface chemistry of N-heterocyclic carbenes

Yanyi Pan, [ab] Ankita Das,[c] Frank Glorius *[c] and Jindong Ren *[ab]

N-heterocyclic carbenes (NHCs) have emerged as a versatile and powerful class of ligands in surface chemistry, offering remarkable stability and tunability when bound to surfaces, including metals, metal oxides, and semiconductors. Understanding their surface and interfacial mechanisms at the atomic-level is essential for precise control of molecule–surface interaction, as well as intermolecular interactions, which directly influence material performance and functionalities. Research in surface chemistry focusing on molecular binding modes, self-assembly, on-surface reactions, and electronic properties is crucial for the rational design of efficient catalysts, customized materials, and high-performance devices. This review highlights these critical aspects of NHCs on surfaces, beginning with their robust and multiple binding modes, which underpin their stability and versatility. The covalent NHC–surface bonds allow NHCs to form stable attachments, often surpassing the strength of traditional thiol-based modifiers, promoting robust anchoring across diverse materials. Another focus is the self-assembly of NHCs into highly ordered monolayers, which facilitates the design of functional nanostructures. Emerging topics also include on-surface reactions, surface electronic properties, and interfacial charge transfer of NHCs, emphasizing their dependence on the substrate and NHC molecular structure. By consolidating recent advancements in the study of NHCs on surfaces, we aim to provide a comprehensive overview of their transformative potential in surface chemistry at the atomic scale, while also identifying key challenges and future directions in the field.

Received 27th February 2025

DOI: 10.1039/d4cs01299b

rsc.li/chem-soc-rev

Key learning points

(1) High-resolution analytical techniques enable precise molecular characterization.
(2) Diverse NHC structures and binding modes on material surfaces.
(3) Self-assembly and on-surface reactions of the NHCs.
(4) Electron-rich covalent binding to enhance the device performance in harsh environments.
(5) Perspectives and future directions in NHC related surface chemistry.

1. Introduction

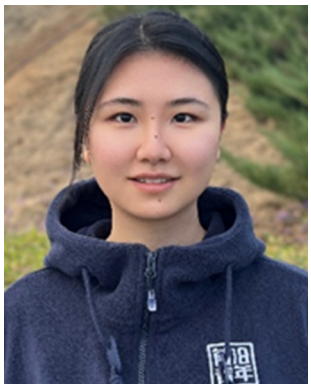
The surface chemistry of functional molecules is essential for advancing materials science and nanotechnology,[1] encompassing not only surface functionalization but also the complex interactions at the molecule–material interface, with emphasis on key aspects such as molecular binding mode,[2] self-assembly,[3] on-surface reactions,[4] and electrical properties.[5] These interactions together govern how organic molecules adhere to material surfaces and their ability to modify the physicochemical properties of the surface, which are key for the fabrication of engineered surfaces for applications in sustainable technologies and functional devices.[6] At the heart of surface chemistry is the binding mode, which defines the interaction between a functional molecule and the surface.[7] The adsorption geometry, whether flat-lying or upright, can significantly impact surface properties like interfacial charge transfer and responsiveness to external stimuli.[8] For instance, a flat adsorption can enhance charge transfer across the surface due to closer proximity between the adsorbate and substrate, whereas an upright adsorption could enable better interaction with external electric fields or light, affecting the material's electronic or optical properties.[9] Building on the binding mode,

[a] *CAS Key Laboratory of Standardization and Measurement for Nanotechnology, CAS Key Laboratory of Nanophotonic Materials and Devices, National Center for Nanoscience and Technology, Beijing 100190, P. R. China. E-mail: renjd@nanoctr.cn*
[b] *University of Chinese Academy of Sciences, Beijing 100190, P. R. China*
[c] *University of Münster, Organisch-Chemisches Institut, Münster 48149, Germany. E-mail: glorius@uni-muenster.de*

This journal is © The Royal Society of Chemistry 2025

the self-assembly of organic molecules on surfaces has garnered increasing attention. Self-assembly can be governed by various forces, including van der Waals (vdW) forces, electrostatic interactions, and hydrogen bonding.[10] Such forces guide the molecules to organize into well-defined monolayers, resulting in highly ordered nanoscale features with enhanced functionality. For example, self-assembled monolayers (SAMs) can form uniform coatings on surfaces, which serve as templates for subsequent chemical reactions or act as protective layers against corrosion.[11] While the self-assembly establishes the foundational organization, on-surface reactions enable precise molecular design through the formation or cleavage of chemical bonds.[12] These reactions, which can be triggered by external stimuli such as light, heat, or electric fields, can be selectively applied to functionalize surfaces as needed.[13] Surface chemistry further extends beyond the molecular binding mode, self-assembly, and on-surface reactions to include the study of electronic properties, which involves molecular electronic states, electron redistribution, and interfacial charge transfer.[14,15] As an illustration, organic molecules with strong electron-donating or electron-withdrawing groups can modulate the electronic properties of the surface, enhancing or suppressing catalytic reactivity, depending on the desired outcome.[16]

The versatility of organic molecules in surface chemistry, particularly their ability to interact with substrates in highly tunable ways, underpins their potential across a wide range of applications. By carefully selecting and modifying organic molecules, researchers can design surfaces with tailored



Yanyi Pan

Yanyi Pan received her BS in Microelectronics Science and Engineering from Sichuan University, China. She is currently pursuing her PhD in Integrated Circuit Science and Engineering at the National Center for Nanoscience and Technology under the supervision of Prof. Jindong Ren. Her work emphasizes enhancing device properties through molecular-scale surface modification. Leveraging surface-sensitive techniques and in situ spectroscopic analysis, she investigates surface defect modulation and interfacial charge redistribution to optimize device performance via atomic-level surface chemical engineering.



Ankita Das

Ankita Das earned her BSc in Chemistry from St. Xavier's College, Kolkata, and her MSc from the Indian Institute of Technology Hyderabad (IITH), India. During her master's, she conducted research under Prof. Tarun Kanti Panda on imidazolium salt-based electrolyte additives for Ca-ion batteries. Since 2021, she has been pursuing her doctoral studies in the group of Prof. Frank Glorius, where her research interest focuses on molecular design and diverse surface functionalization of N-Heterocyclic carbenes. Her research also includes developing novel electrolytes for advanced Li-ion batteries.



Frank Glorius

Frank Glorius is Full Professor of Organic Chemistry at the University of Münster, Germany. His diverse and interdisciplinary research program focuses on the development of new concepts for various areas of catalysis, such as skeletal editing, photocatalysis, C–H activation, N-heterocyclic carbenes (NHCs) in organocatalysis, (asymmetric) arene hydrogenation, and as unique surface modifiers for materials science and (enantioselective) heterogeneous catalysis. In addition, his group is developing molecular tools for the study and modification of biological systems and battery systems. Finally, data science and molecular machine learning are other current topics that the group is focusing on.



Jindong Ren

Jindong Ren is Full Professor of Physics, Chemistry, and Materials Science at the National Center for Nanoscience and Technology, China. His interdisciplinary research program focuses on nanoscale science and nanotechnology, particularly on-surface precision chemistry, interfacial charge transfer, and the optoelectronic properties of low-dimensional materials. Prof. Ren integrates advanced techniques, including scanning tunneling microscopy/spectroscopy, non-contact atomic force microscopy, and X-ray photoelectron spectroscopy, to push the boundaries of atomic-scale characterization. His group leads pioneering research in atomic-level surface functionalization, Ångström-scale spatial structure analysis, millielectronvolt-resolution electronic state characterization, and deca-Newton mechanical resolution, driving innovation in the precise control of nanoscale physical and chemical properties.

electronic, chemical, and optical properties, which are crucial for the development of next-generation nanotechnologies.[17] However, despite significant advancements in surface chemistry over the past few decades, traditional organic molecules still face challenges in balancing surface stability with mobility, which is indispensable for applications in harsh environments such as high voltage, current, or temperature conditions.[18] This highlights the need for novel surface modifiers that offer robust binding to surfaces, high surface mobility, and tunable electronic properties, making them ideal candidates for advancing functionalized surface development and molecular device integration.

NHCs are a class of new organic molecules characterized by a highly reactive carbon atom, which is stabilized by the surrounding heterocyclic ring structure to reduce the inherent instability of the free carbene and prevent excessive reactivity, surpassing traditional ligands like thiols in multifarious scenarios.[19–22] The carbon atom in the NHC is highly electron-rich, due to its lone pair of electrons and additional electron donation from neighbouring N-atoms, allowing it to serve as a strong σ-donor in coordination chemistry. Initially recognized for their role in homogeneous catalysis, NHCs have recently gained significant attention in surface chemistry due to their exceptional stability and versatility.

When anchored to solid surfaces—ranging from metals to metal oxides and semiconductors—NHCs chemically adsorb onto the surface, manifesting remarkable functionality in surface chemistry, including versatile binding mode, ordered self-assembly, controlled on-surface reactions, and flexible electronic properties.[23,24] An interesting feature is the "ballbot"-type configuration, which overcomes the challenge of simultaneously achieving high surface mobility and stability, a limitation commonly encountered with traditional ligands.[25,26] Despite the great progress achieved in NHC surface studies,[27–29] several critical questions remain open. How can the interaction between NHCs and surface materials be optimized, particularly to establish stable yet dynamic configuration like the ballbot type? How can the self-assembly and on-surface reaction of NHCs be controlled to unlock their full functional potential? What are the precise mechanisms driving charge transfer at the NHC–surface interface? Addressing these challenges is essential to fully unlock the transformative capabilities of NHCs in surface chemistry.

This review summarizes the current understanding of NHC surface chemistry, emphasizing their atomic-level interactions on surfaces (Fig. 1). First, various surface modification methods are outlined. Then, the binding modes and self-assembly processes of NHCs, particularly the ballbot configuration, are elucidated. Next, we introduce recent advances in chemisorbed NHC polymers *via* on-surface reactions and discuss how electron-rich NHCs modulate surface electronic properties, with a focus on their prospective applications in harsh environments. Finally, key challenges and emerging opportunities in this prospering field are summarized.

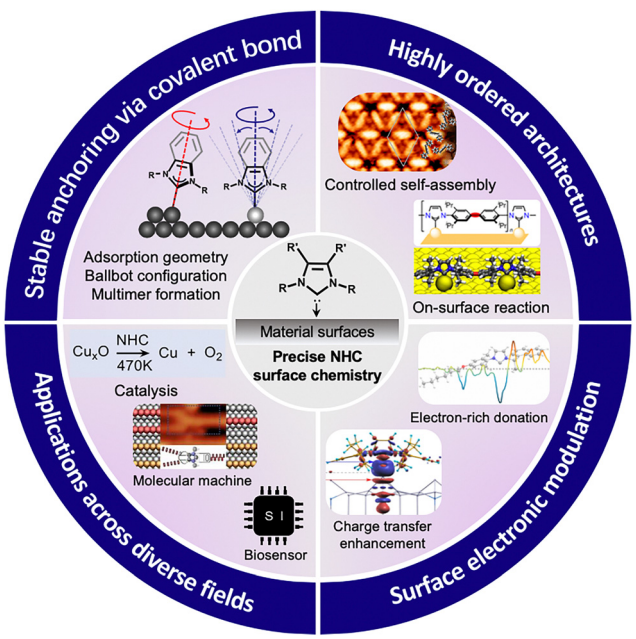


Fig. 1 Schematic diagram of the research progress on NHCs on surfaces.

2. Surface modification methods

NHCs are defined as heterocyclic compounds containing divalent carbon (carbene carbon) with at least one nitrogen within the ring structure (Fig. 2a).[30,31] The backbone and nitrogen substituents (*N*-substituents) of NHCs significantly affect their electronic structure and stability. Among them, the electronic attributes of the backbone structure are capable of modulating the electronic distribution across the entire NHC molecule, which in turn has a bearing on its propensity for stability and chemical reactivity. For example, NHC backbones that incorporate aromatic rings are endowed with the capacity to confer enhanced stability through the mechanism of conjugative stabilization.[32] Additionally, the electronic properties of the *N*-substituents—whether electron-donating or electron-withdrawing—regulate the charge density on the carbene carbon. Steric effects from these substituents further impact the coordination geometry when NHCs complex with metals.[33]

The early method for synthesizing NHCs, proposed by Arduengo, involved using strong bases to deprotonate the imidazolium salts in aprotic solvents, producing isolable NHCs.[34] This pioneering approach did not include any additional purification; however, the residual base interfered with the formation of NHC SAMs during surface modification (Fig. 2b).[35] A significant advancement was introduced by Crudden and co-workers, who demonstrated that the decomposition of solid imidazolium hydrogen carbonate, either in solution or the vapor phase, could yield high-quality NHC films by releasing CO_2 and H_2O.[36] Another notable method utilizes zwitterionic carbon dioxide complexes of NHCs, which release CO_2 to generate free NHCs. Johnson and colleagues employed this strategy to successfully anchor traceable NHCs onto gold surfaces for various applications.[27] Glorius, Fuchs, and co-workers advanced the thermally induced decarboxylation method to enable vacuum deposition of NHCs on surfaces, achieving high-quality,

This journal is © The Royal Society of Chemistry 2025

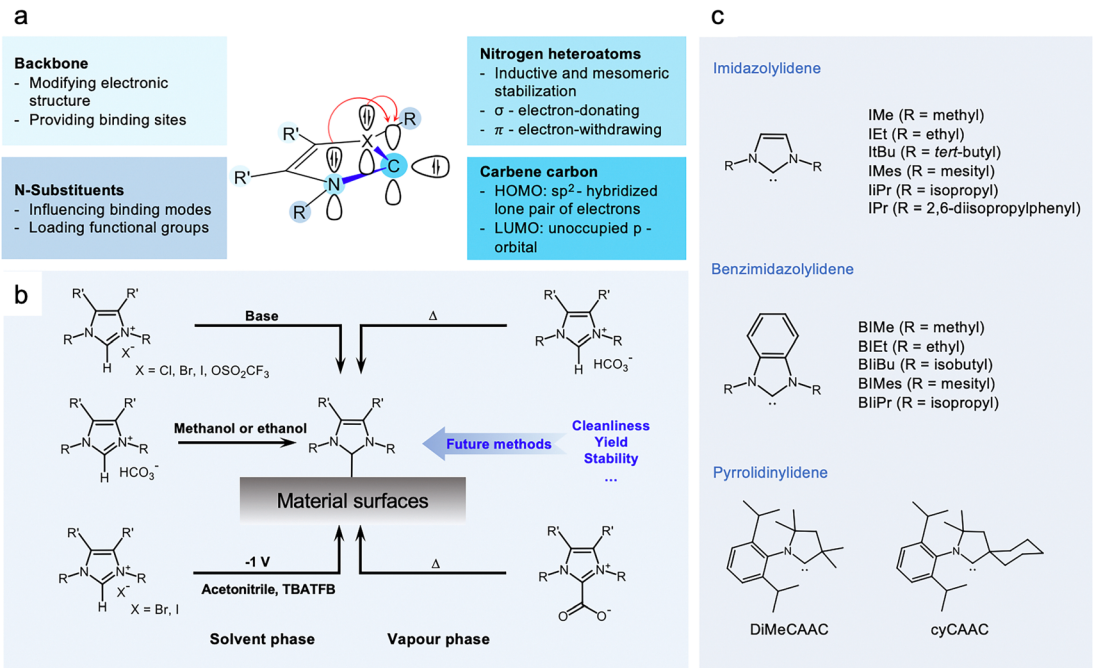


Fig. 2 Strategies for surface functionalization using NHCs. (a) General structural features of NHCs, illustrating the influence of the nitrogen heteroatoms, the ring backbone, the *N*-substituents, and the carbene carbon on the stability, reactivity, and electronic properties of NHCs. (b) Overview of deposition techniques employed to achieve NHC monolayers on diverse material surfaces. (c) Nomenclature and structural representation of commonly used NHCs, highlighting their versatility in surface functionalization and modification.

impurity-free metal surface modifications.[29] However, the synthesis of bicarbonate salts or CO_2 adducts involves cumbersome separation and ion-exchange steps.[37] In NHC-based surface modifications, the use of carboxylates often restricts the incorporation of base sensitive functional groups into NHCs. Additionally, incompatibilities arise between protonated functional groups and the synthesis of the hydrogen carbonate.[38] An alternative approach is the electrochemical deposition, which generates NHC SAMs by deprotonating imidazolium cations through the high local concentration of hydroxide ions at the metal electrode.[39] This method effectively addresses the issue of incompatibility between multiple functional groups and imidazolium salt synthesis in previous methods.[27,35–37] However, side reactions limit the adjustable range of solution concentration and pose stability concerns. Therefore, the ideal surface modification technique should enable selective carbene formation with good yield and will have wide functional group tolerance (Fig. 2c). Future advancements should focus on employing mild precursor formation techniques and efficient NHC release mechanisms, while avoiding aggressive reagents like strong alkalis or complex ion-exchange processes. A promising approach is to leverage the on-surface reaction to *in situ* convert NHC derivatives into NHCs.[40] This strategy will profoundly circumvent challenges such as the interference of carboxylates in the synthesis of NHCs with specific electronegative functional groups. By utilizing controlled surface reactions, it is possible to preserve the stability and reactivity of NHCs, simultaneously reducing the reliance on stringent chemicals and complex procedures and providing a more sustainable route for surface functionalization.

3. Binding mode of NHC adsorption: the role of the Ballbot configuration

3.1. Stable anchoring of NHCs *via* covalent bond

The stable anchoring of NHCs on metal, metal oxide and semiconductor surfaces is crucial for their application in materials design. For instance, on metal surfaces, the strong σ-donor affinity of the carbene carbon in NHCs, coupled with π back-donation from the metal, results in a strong covalent C–metal (C–M) bond that stabilizes the NHC on the metal surface.[37,41] This bonding is essential for stabilizing the NHC–substrate interface. Early investigations into NHC–metal (NHC–M) interactions primarily focused on gold surfaces,[27,28,37] due to gold's excellent chemical durability, resistance to oxidation or corrosion, and its outstanding conductivity and thermal properties. In 2013, Johnson *et al.*[27] reported the anchoring of NHCs on gold surfaces, computationally determining a bond length of 2.03 Å for the carbene carbon–gold (C–Au) interaction, with an associated dissociation energy of 67 kcal mol^{-1}. Further experimental studies by Crudden and co-workers demonstrated the thermal, chemical, and electrochemical stabilities of NHCs on gold surface.[28,37] Surprisingly, the C–Au bond exhibits significantly enhanced stability compared to traditional sulfur–gold (S–Au) bonds typically found in thiol-based SAMs. Density functional theory (DFT) simulations indicate that the most stable binding mode for NHCs on gold surfaces is *via* a single C–Au bond at atop sites, with a bond strength of approximately 150 kJ mol^{-1}. This is about 25 kJ mol^{-1} stronger than the S–Au bond strength, highlighting the superior binding strength of NHCs. Similar to the case of Au, NHCs establish stable

anchoring on other metal surfaces through strong covalent interactions. This facilitates robust coordination between the carbene centre and metal substrates, enabling NHCs to form stable attachments to a range of metal films and nanoparticles, such as Ag, Pd, Pt, Ru, Mg, and Cu.[25,26,29,42–44]

NHCs have also been shown to adsorb on metal oxides, a process of considerable interest due to the role of metal oxides in tailoring surface properties.[45,46] DFT simulations reveal that the NHCs bind on copper oxide predominantly through surface oxygen atoms rather than metal atoms, with a strong adsorption energy of 1.16 eV. This interaction results in a C–O bond distance of 1.26 Å, and the NHC ligand can break a Cu–O bond in the oxidized overlayer, pointing outward from the surface. More recently, NHCs have been successfully assembled not only on metal and metal oxide surfaces but also on hydrogen-terminated and boron-doped silicon surfaces, where the C–Si bond predominantly displays σ-character and features low thermodynamic barriers for bond rotation.[47,48] These covalent bonds, ranging from C–M, C–O, to C–Si, stabilize the NHCs on the surface, highlighting the adaptability and robustness of NHCs anchoring across various materials.

3.2. NHC adsorption geometry on the surface

The orientation of individual NHC on the surface can be either upright, tilted, or flat-lying, characterized by the deviation of the molecular major axis from the substrate normal. The steric properties, governed by the *N*-substituents, significantly influence the NHC orientation and can be quantified using the 'buried volume' parameter (%V_{bur}) developed by Nolan, Cavallo and co-workers (Fig. 3a).[49] The %V_{bur} value of an NHC is defined as the proportion of a spherical volume that overlaps with the NHC ligand when it coordinates with a metal atom at the core of the sphere. Ragone systematically revealed the complex relationship between NHC ligand flexibility and steric resistance through static/dynamic %V_{bur} calculations, breaking through the limitations of traditional static analysis.[50] In recent years, multiple research groups have explored the adsorption geometry of NHC molecules through a variety of advanced surface characterization techniques, shedding light on their binding configurations and the factors influencing the surface orientation.[33,51–59] Lovat *et al.* analyzed how the steric effects of *N*-substituents influence the binding orientation of NHCs on the Au(111) by near edge X-ray absorption fine-structure spectroscopy (NEXAFS) and DFT calculations (Fig. 3b).[51] IMe with small N substituents lies nearly parallel to the surface ($\theta \approx 72^\circ$), while Dipp with large substituents is almost perpendicular to the surface ($\theta \approx 13^\circ$), and BIiPr with intermediate-sized substituents were bound to the surface at an inclination of 40°. Similar results were obtained by Larrea *et al.* on transition metals like Cu substrate *via* high resolution electron energy loss spectroscopy (HREELS) experiments.[52] Bakker *et al.* reported using scanning tunneling microscopy (STM) and X-ray photoelectron spectroscopy (XPS) that longer alkyl substituents, as opposed to bulkier groups, favor the formation of flat-lying NHC–Au–NHC complexes.[33] Dery *et al.* demonstrated by NEXAFS that the reorientation of NO_2-functionalized NHCs from tilted to flat-lying conformation is induced by nitroso reduction.[43] Moreover, Camden, Jenkins, and Jensen collaborated to utilize surface-enhanced Raman spectroscopy (SERS) to identify the dominant contributions of the vibrational modes of substituent atoms to the SERS signal.[53] They demonstrated that the size of the substituents significantly influences the molecular orientation on the surface, as evidenced by the distinct adsorption behaviours of NHCs with methyl, *tert*-butyl, hexyl, and ethyl groups on gold nanoparticles (Fig. 3c and d).[54] Recently, Li *et al.* further investigated the adsorption configuration of NHCs using tip-enhanced Raman spectroscopy (TERS).[55] The chemical fingerprint of the molecule emphasizes the critical role of *N*-substituents in determining the surface binding mode of NHCs. The studies presented above indicate that the increase in the bulkiness of *N*-substituents leads to a shift in the molecular orientation from the flat-lying configuration to the upright conformation. Meanwhile, the elongation of *N*-substituents enhances vdW interactions with the surface, thereby favoring a greater tendency for the molecule to adopt a flat-lying configuration.[33]

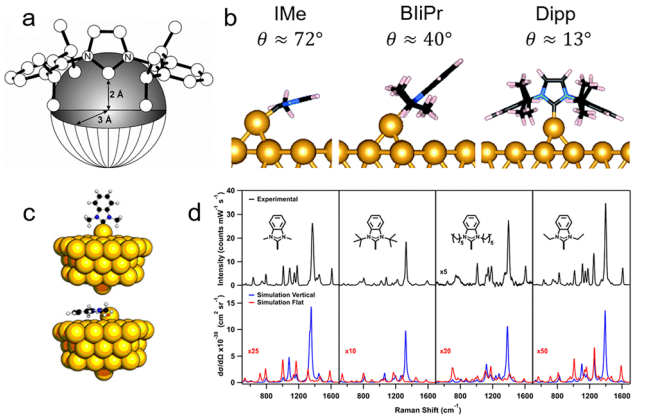


Fig. 3 Tunable surface geometry of NHCs. (a) Schematic illustration of the 'buried volume' parameter for analyzing the steric properties of NHCs (reproduced from ref. 49 with permission from American Chemical Society, copyright 2003). (b) Orientation of NHCs with varying steric hindrance of *N*-substituents on metal surfaces, as determined by NEXAFS spectra and computational analysis. θ Represents the inclination between the molecular principal axis and the substrate normal (reproduced from ref. 51 with permission from The Royal Society of Chemistry, copyright 2019). (c) and (d) SERS spectra reveal an increasing intensity of signals from the upright conformation as the size of the *N*-substituents in NHCs increases. (c) DFT simulations of isolated NHCs in upright and flat-lying configurations on the Au_{58} cluster. (d) Experimental (top) and theoretical (bottom) SERS spectra of NHCs with different *N*-substituents (reproduced from ref. 54 with permission from American Chemical Society, copyright 2023).

Molecular modulation between upright and flat-lying binding modes by annealing has been systematically reviewed in thiol systems.[57] In the case of NHCs, Zeng *et al.* employed reflectance absorbance infra-red spectroscopy (RAIRS) to investigate structural changes in BIiPr on Pt(111), proposing a unique intermediate binding mode.[56] In this mode, strong metal–adsorbate interactions cause the π-chemisorbed aromatic group to anchor the benzimidazolium moiety to the surface, while the N–C–N region tilts away. This configuration initially coexists with the upright surface-bound state, but transitions fully to the latter upon annealing or over time.

This journal is © The Royal Society of Chemistry 2025

Moreover, based on surface-specific vibrational sum-frequency generation (SFG) spectroscopy, Gutheil *et al.* reported that the orientation of NHC monolayers on Au thin films is highly dependent on deposition time.[59] Over time, a transition occurs from the kinetically favoured upright orientation to the thermodynamically favoured flat-lying binding mode. This discovery enables precise control over the monolayer structure based on adsorption time, thereby allowing for the tuning of its physicochemical properties.

Therefore, factors including steric hindrance of *N*-substituents, NHC–surface interaction, and orientation energy barriers mainly influence the adsorption geometry of NHCs. Through the precise modulation of the relevant variables, their binding modes can be controlled with accuracy, enabling the design of customized applications in catalysis and photoelectric devices.

3.3. Ballbot-type configuration featuring strong binding and high mobility

In addition to the molecular orientation transition discussed in the last section, researchers have also observed that, under certain conditions, the M atom at the other end of the NHC–M bond undergoes atomic-scale displacements relative to the first layer of the substrate, forming an NHC–gold adatom complex. In 2017, Wang *et al.* first reported the reorganization of NHC self-assembled islands through multiple STM scans of the same region (Fig. 4a).[29] The high surface mobility of IPr species on Au(111) contradicts the strong carbene C–Au binding, underscoring the need for a coherent theoretical model to reconcile these observations. They proposed the "ballbot" binding mode on the Au(111), where the NHC molecule facilitates the detachment of a single Au atom through chemical bonding, forming a stable NHC–Au complex. This interaction results in a unique surface configuration, in which the chemically adsorbed NHC molecule exhibits significant mobility across the surface, enabling dynamic movement and reorganization. DFT optimization revealed that the heights of IMe in the non-ballbot and ballbot configurations were optimized to 5.13 Å and 7.17 Å, respectively (Fig. 4b). Car–Parrinello molecular dynamics simulations and thermodynamic integration were used to describe the kinetic process by which IMe extracts an Au atom from the surface, demonstrating that IMe lowers the free-energy barrier for forming a single Au adatom. Based on these findings, the activation barrier for the diffusion of ballbot species on the surface was calculated as 10 kJ mol^{-1}, which is dramatically lower than that for the surface bond configuration. However, their experimental evidence supporting the ballbot model is

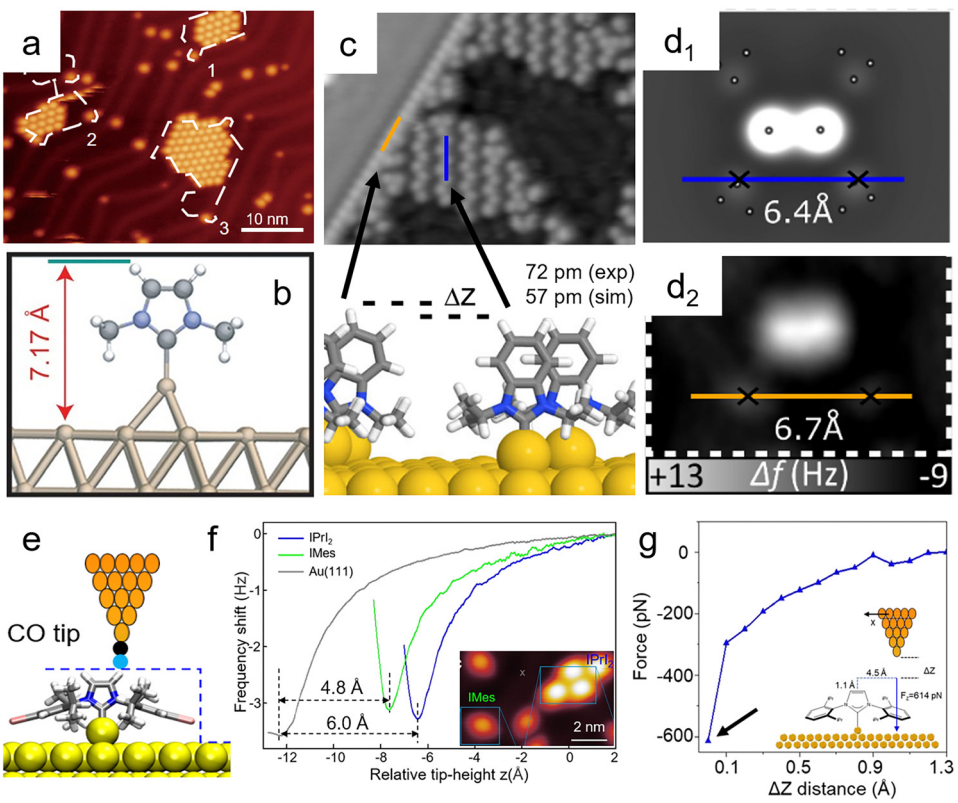


Fig. 4 Atomic-scale analysis of NHC molecular configurations: ballbot-type structure exhibiting strong surface binding and high surface mobility. (a) Mobility of NHC islands on surface. (b) The proposed ballbot configuration and theoretical optimization of NHC configuration at the Au interface (reproduced from ref. 29 with permission from Springer Nature, copyright 2017). (c) Height difference between BIiPr molecules at step edges and in the ballbot configuration (reproduced from ref. 60 with permission from Springer Nature, copyright 2021). (d_1) and (d_2) DFT optimization (d_1) and AFM image (d_2) showing the contrast and distance of the iPr groups. (e) and (f) nc-AFM height difference between ballbot $IPrI_2$ and non-ballbot IMes. (g) Lateral shear force threshold for the directed movement of a ballbot NHC (reproduced from ref. 23 with permission from Springer Nature, copyright 2023).

limited, particularly as it does not fully account for the complex perturbations to molecular motion induced by the STM tip,[58] underscoring the need for further experimental validation.

Inayeh *et al.* utilized DFT simulations and STM constant current mode to measure the height differences of BIiPr at the Au(111) step edges, comparing ballbot at 0.8 monolayer (ML) and non-ballbot below 0.4 ML binding modes (Fig. 4c).[60] The surface coverage was regulated by adjusting the NHC precursor solution molarity within the range of 5.0 to 42.5 mM. Their study revealed that the heights of the ballbot species are uniform, whereas those of the non-ballbot species differ by approximately the height of a gold atom. This finding supports the existence of a critical molecular coverage threshold for ballbot species formation. Recently, Ren *et al.* extensively investigated ballbot configuration using the highly sensitive qPlus-based noncontact atomic force microscopy (nc-AFM) technique, measuring molecular contrast, height, and lateral force.[23] The first direct evidence was obtained from the observation of adatom-induced elevation of the molecule, which caused the iPr side groups to relax towards the surface (Fig. 4d_1 and d_2). This was experimentally observed as a reduced contrast signal for the iPr group in the non-ballbot model. The relaxation also triggered the rotation of the iPr groups, causing the top hydrogen atoms to shift away from the molecular axis. The measured distance between iPr groups closely matched the predictions for the ballbot model, diverging from the non-ballbot model. The second piece of direct evidence involved probing the height of the topmost hydrogen atom in NHC molecules using a CO-terminated tungsten tip (Fig. 4e and f). The measured height of the $IPrI_2$ monomer aligned with the predicted ballbot configuration height of 6.0 Å, showing a 1.2 Å difference compared to the IMes monomer in the non-ballbot configuration. Finally, the third piece of evidence originated from measuring the lateral shear force threshold required to move an NHC–Au complex using an nc-AFM tip (Fig. 4g). The measured threshold force of −354 piconewtons (pN) was comparable to the force needed to move a physically adsorbed atom on a metal surface[61] and significantly lower than the force required to move a complex with strong covalent bond.[62] These results offer compelling experimental evidence in favor of the ballbot model, resolving the apparent contradiction between the strong chemisorption and the high mobility of NHCs on the metal surface.

Given the experimental support for the ballbot model of NHCs, further investigation into the atomic-scale formation is imperative to gain broader implications of this configuration. This involves examining the balance between the strong molecule–metal interaction, which facilitates the detachment of metal atoms, and the steric repulsion between the NHC molecule and the substrate, which stabilizes the resulting structure. For instance, when cyclohexyl cyclic (alkyl)(amino)carbenes (cyCAAC) are deposited on Au (111), they form a ballbot species. In contrast, on Ag(111), the cyCAAC molecule adopts a binding mode where the Ag atom is fixed at a lower adsorption height compared to the Ag adatom, preventing the full extraction of the metal atom from the surface. Similarly, on Cu(111), the cyCAAC molecule binds to a copper atom within the first layer of the substrate, contrasting with the ballbot configuration on Au(111) and further demonstrating the substrate-dependent nature of the binding mode.[63]

Hence, the formation of ballbot-type NHCs is primarily influenced by elements such as molecular coverage, NHC structure, and substrate species. While ballbot configuration has primarily been observed on metal surfaces, there is considerable potential to extend this concept to other materials. Exploring how various surface properties, such as those found in semiconductors, insulators, or 2D materials, influence the formation and stability of ballbot configuration could pave the way for novel approaches in materials design and surface functionalization.

3.4. NHC multimers on the surface

In addition to the binding of an individual molecule to a substrate atom, experiments have also observed the formation of different NHC complexes, where two or more NHCs bind to a single metal adatom on the substrate. In 2017, Wang *et al.* reported the discovery of the bis-NHC complex and simulated the morphology of $Au(IMe)_2$ (Fig. 5a) and $Au(IMe)_3$ (Fig. 5b).[29] A parallel study on Cu(111) revealed that at high coverage, BIMe can form a two-dimensional ordered structure consisting of two $(BIMe)_2Cu$

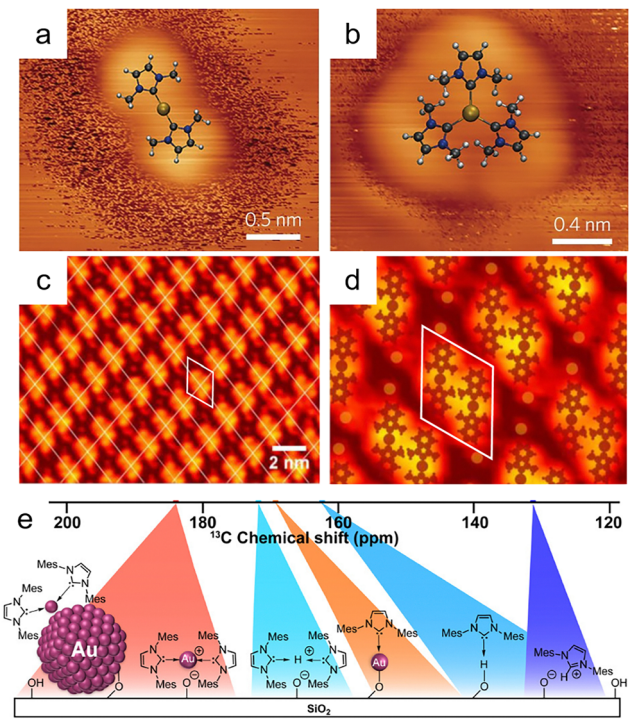


Fig. 5 Molecular arrangement of NHC multimers on various surfaces based on high-resolution atomic-scale structural analysis. (a) and (b) Observation of NHC multimers on Au(111) in the STM experiment, with molecular structures calculated using DFT (reproduced from ref. 29 with permission from Springer Nature, copyright 2017). (a) Dimeric IMe–Au complexes. (b) Trimeric IMe–Au complexes. (c) STM topography of dimeric BIMe–Cu complexes on Cu(111). (d) Schematic representation of the SAM shown in (c), where two BIMe–Cu complexes are indicated by the white rhombus (reproduced from ref. 52 with permission from Wiley-VCH Verlag, copyright 2017). (e) Solid-state NMR analysis confirms the coordination of bis-NHCs to metal sites on silica surface (reproduced from ref. 64 with permission from American Chemical Society, copyright 2024).

This journal is © The Royal Society of Chemistry 2025

complexes on the surface (Fig. 5c and d). These complexes are stabilized by the insertion of Cu adatoms and form a periodic arrangement matching the (7×4) superstructure.[52]

Later, Bakker *et al.* provided a more detailed discussion on the formation process of NHC dimers.[33] Theoretically, adjacent IMe molecules forming dimers are attached to the surface in a flat-lying configuration. These two molecules then bind to the same surface atom, followed by the extraction of a metal atom from the surface to form an NHC–M–NHC complex. Further calculations of the binding energy of different NHCs in both the adatom binding mode and the bis-NHC complex binding mode showed that NHCs with smaller side groups prefer to form bis-NHC complexes.[60] Notably, the minimal energy difference between the monomer and dimer species of BIiPr suggests that both cases are energetically accessible. Gradient experiment on BIiPr coverage indicates that the relative abundance of NHCs bound to surface atoms, adsorbate atoms, and bis-NHCs on the surface is significantly influenced by coverage. Dimer species appear at low coverage and transform into ballbot species as coverage increases, possibly due to competitive adsorption. Recently, Dery *et al.* identified the $Au(IMes)_2$ species on a silica surface using solid-state nuclear magnetic resonance (NMR) spectroscopy, benchmarked on ^{13}C labelled molecular models.[64] Their study successfully demonstrates that coordination of bis-NHCs on metal sites can be realized in numerous extended systems, ranging from well-defined crystal surfaces to metal sites in nanoparticles and those dispersed on oxide carriers such as silica.

Beyond these observations, NHC multimers might hold multiple potential implications and applications: (1) their unique structural properties can impart novel physical and chemical characteristics to the material. In optoelectronic materials, such multimeric structures can effectively tune the material's photoelectric conversion efficiency.[65] (2) Fixing these molecular multimers onto material surfaces can dramatically alter the surface characteristics, such as wettability and adhesion. In the field of biomedical materials, forming multimeric structures with bioactive molecules on the material surface can enhance biocompatibility and cellular affinity.[37] (3) The metal atoms in these molecular multimers can act as active centers that collaborate with the connected molecules, enabling efficient catalytic reactions.[66] This multimeric structure can create a unique reaction microenvironment, making it easier for reactant molecules to accumulate and react around the active centers. (4) The multifunctionality of molecular multimers allows for their integration into nano-devices, where they can perform complementary roles. In nanoelectronics, assembling conductive molecules and magnetic metal atoms into multimeric structures could lead to the development of nano-devices that simultaneously exhibit electrical and magnetic functions, holding promising possibility in high-density storage and quantum computing.[67] The exploration of NHC multimers thus opens new possibilities for material design, surface engineering, and catalysis, making it a promising area for future research and application.

4. Self-assembly of chemisorbed NHCs on the surface

4.1. Controlled self-assembly on the metal surface

For metallic substrates, which typically exhibit weak molecule–surface interactions, striking a balance is crucial to prevent organic molecules from being overly immobilized and to avoid excessive interactions that could cause molecular deformation or decomposition.[47] This moderate interaction facilitates molecular diffusion on the surface, allowing the molecules to explore and settle into optimal positions for self-assembly.

The study of NHC self-assembly on metal surfaces initially focused on Au(111).[29] For non-ballbot-type NHCs, like IMes, which directly bond to surface atoms, achieving an ordered structure on the surface is not feasible. In contrast, for ballbot-type NHCs, similar to other previously reported organic molecules,[68] at low coverage, IPr, IMe,[29] and BIiPr[60] preferentially bind to step edges and elbow positions of the $22 \times \sqrt{3}$ reconstructed regions. As the coverage increases, NHC islands form in fcc stacking regions. At full coverage, self-assembly including hexagonal (IPr), rhombic (IMe), and zigzag (BIiPr) patterns has been observed. After analyzing the NHC self-assembly on the gold surface, the focus will shift to their controlled self-assembly, with emphasis on how parameters including substrate species, crystallographic orientation, temperature, molecular structure, and intermolecular interaction influence the formation of well-ordered structures.

As demonstrated by the recent study on cyCAAC on various metal substrates such as Au(111), Ag(111), and Cu(111), the self-assembly is substrate-dependent.[63] On Au(111), the self-assembly of the trimer phase in low coverage and the dimer phase in high coverage was observed (Fig. 6a). In contrast, on Ag(111), cyCAAC molecules preferentially adopted a closely packed dimer phase, independent of coverage, due to the distinct binding mode of cyCAAC on the Ag surface. The self-assembly on Cu(111) is unique in that it exhibits well-separated stripe-like patterns at sub-monolayer coverage, with a dimer phase emerging at monolayer coverage.

The self-assembly of NHCs is not only influenced by the substrate species but also by the crystallographic orientation of the surface. For instance, Navarro *et al.* compared the self-assembly of IPr molecules on Cu(100) and Cu(111) surfaces, revealing distinct behaviours.[44] On Cu(100), at low coverage, IPr molecules form dispersed clusters. At higher coverage, small, close-packed patches without long-range order emerge. In contrast, IPr molecules on Cu(111) exhibit more ordered assemblies with increased coverage. At low coverages, triangular and hexagonal islands dominate, while "magic number" islands (12-membered) form at higher coverages (Fig. 6b). Molecular lattice (described as $\sqrt{21} \times \sqrt{21}$ R11°) commensurate with the substrate is observed in nearly 1 ML. DFT calculations suggest that Cu(100) has a high thermodynamic cost for generating mobile Cu adatoms, limiting molecular mobility and hindering ordered assembly. In contrast, Cu(111) reduces both thermodynamic and kinetic barriers, promoting higher mobility and more ordered structures.

In addition to the substrate species and its crystallographic orientation, substrate temperature greatly contributes to the

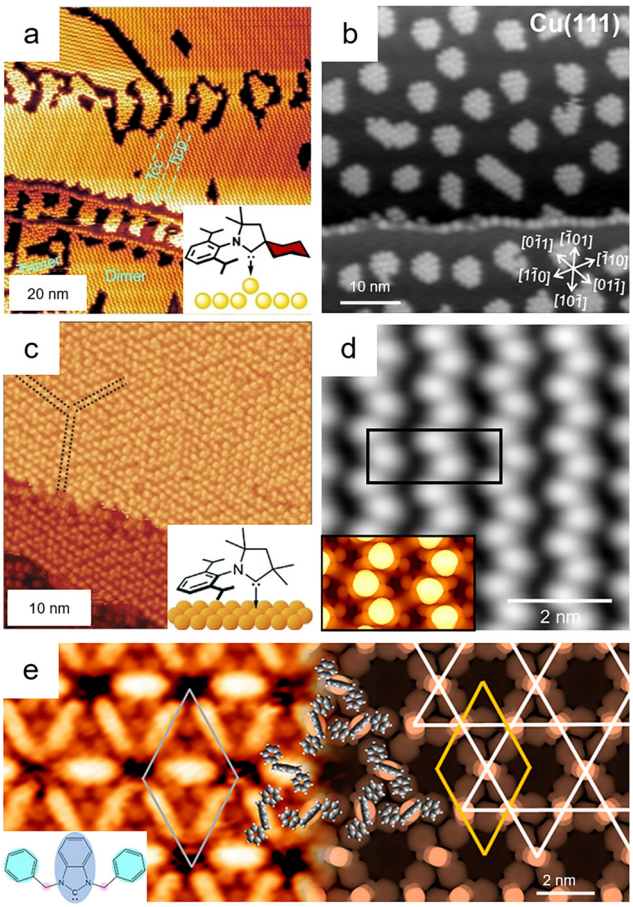


Fig. 6 Controlled self-assembly of NHCs on metal surfaces, influenced by substrates, molecular structures, and intermolecular interactions. (a) Reversible self-assembly of cyCAAC on Au(111), demonstrating conformational switching between axial chair and equatorial chair forms at varying coverages (reproduced from ref. 63 with permission from John Wiley and Sons, copyright 2022). (b) STM image showing IPr self-assembly on Cu(111): at 0.25 ML coverage, on Cu(111), IPr forms magic number islands with a consistent quantity of molecules (reproduced from ref. 44 with permission from John Wiley and Sons, copyright 2022). (c) STM image of DiMeCAAC assembly on Au(111), showing short-range ordered but long-range disordered SAMs compared to cyCAAC (reproduced from ref. 69 with permission from John Wiley and Sons, copyright 2020). (d) STM images of BIiPr self-assembly on Au(111), where intermolecular π–π and CH–π interactions result in different distances between adjacent molecules in a zig-zag lattice (reproduced from ref. 60 with permission from Springer Nature, copyright 2021). (e) Kagome lattice formed by NHCs on Cu(111) *via* π–π interactions among *N*-substituents (reproduced from ref. 70 with permission from American Chemical Society, copyright 2022).

modulation of the self-assembly. Annealing the samples provides the necessary energy for NHCs to overcome energy barriers and enhance ordering of the self-assembled structures.[71] For example, on gold surfaces, annealing drives a transition from an upright zig-zag arrangement to a more stable dimeric configuration for BIiPr self-assembly, which reduces molecular mobility and promotes the formation of well-ordered assemblies.[60]

Furthermore, the self-assembly of NHCs is also influenced by their molecular structures, particularly the *N*-substituents. A comparison of cyCAAC[63] and DiMeCAAC[69] reveals pronounced differences in their ability to form ordered assemblies on metal surfaces. While cyCAAC readily forms well-ordered self-assembled structures, DiMeCAAC fails to achieve similar organization (Fig. 6c).[69] This disparity primarily stems from the flexibility of their side groups. The cyclohexyl side groups of cyCAAC can undergo conformational adjustments, such as ring flipping, allowing it to adapt to different adsorption geometry and promote orderly molecular alignment, thereby facilitating regular self-assembly. In contrast, DiMeCAAC, with its two-methyl side group, has a more rigid structure. The small size and limited conformational adaptability of the methyl groups restrict the molecule's adjustment on the surface, preventing the formation of stable, long-range ordered arrangements.

Finally, intermolecular interactions, including π–π and CH–π stacking, also effectively regulate the self-assembly of NHCs on metal surfaces.[72,73] For instance, BIiPr forms a zig-zag lattice on Au(111), with unequal distances between d_{zig} and d_{zag}, caused by CH–π interactions between the substituents and the imidazole ring, as well as π–π interactions between imidazole moieties (Fig. 6d). In the case of BIiPr, π–π and CH–π interactions result in small tilting and rotation of molecules, disrupting translational symmetry and generating unequal zig and zag segments. This behaviour is confirmed by Inayeh *et al.*, who demonstrated that deuterated BIiPr-d_{14} forms less ordered overlayers due to the weakened CH–π interactions upon deuteration.[60] Additionally, π–π interaction between *N*-substituents can further influence self-assembly, as observed by Angove *et al.* in the DBZ molecules, where π–π stacking between benzyl substituents leads to the formation of a Kagome lattice (Fig. 6e).[70]

4.2. Self-assembly on metal oxide and semiconductor surfaces

Self-assembly on non-metallic substrates, such as metal oxides and semiconductor surfaces, holds significant promise for improving surface functionalization, and facilitating the development of advanced sensors, catalysis, and electronic devices with tunable interfacial properties.[74]

Navarro *et al.* explored NHC adsorption on the copper oxide (Cu_xO) surface, reporting covalent bonding between NHCs and oxygen atoms in the oxide layer for the first time (Fig. 7a).[45] The abundance of surface mobility, which is common on metallic substrates, often results in more stable but less dynamic self-assembled structures on metal oxides. Indeed, NHCs align along rows rather than forming closely packed islands (Fig. 7b). Further investigation elucidated the role of NHC ligands in the thermal reduction of copper oxide overlayers, demonstrating that these ligands actively participate in the reduction process. Such participation is evidenced by the promotion of copper oxide removal at significantly lower temperatures at 470 K. This finding highlights the advantage of NHC ligands to facilitate more energy-efficient and controlled reduction processes, which could be crucial for applications in catalysis and materials science.

Franz *et al.* demonstrated NHC deposition on boron-doped Si(111) surfaces, where NHCs covalently bind in an upright configuration, forming highly ordered monolayers with excellent thermal stability and considerable work function reduction.[75] The assembly of IPr was studied by computing the potential

 This journal is © The Royal Society of Chemistry 2025

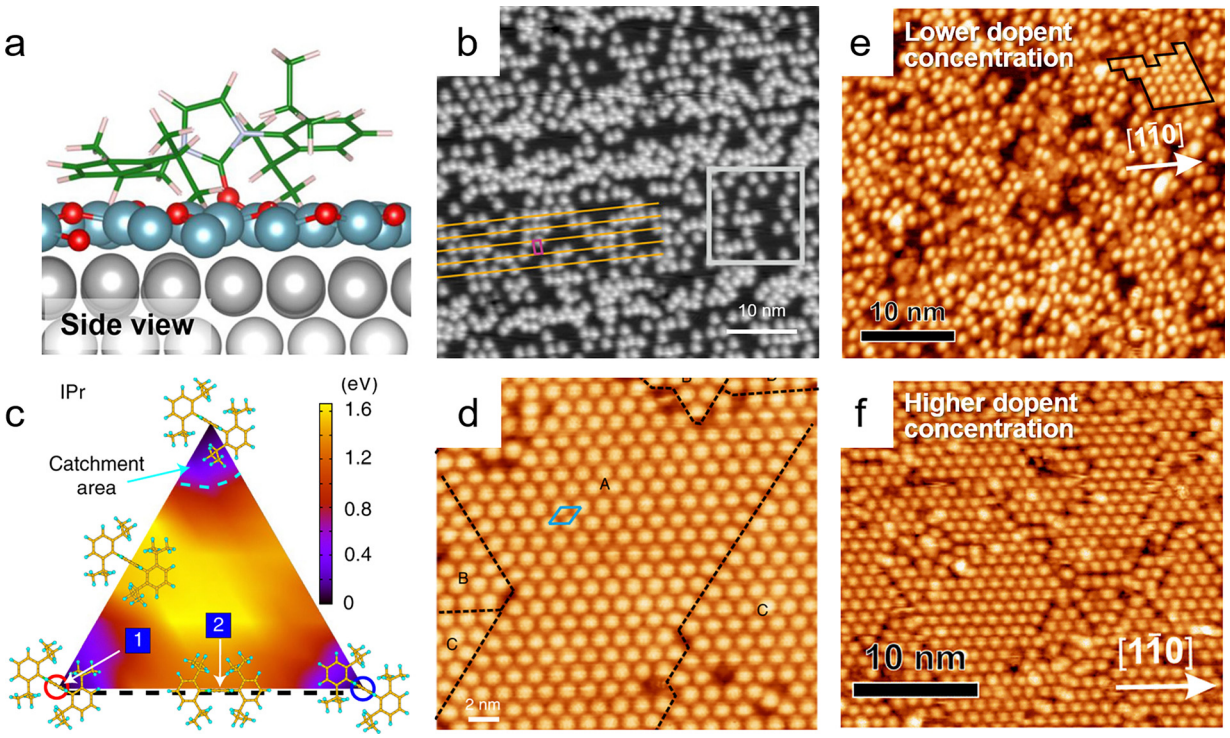


Fig. 7 Self-assembly of NHCs on metal oxide and semiconductor surfaces. (a) and (b) Self-assembly of IPr on the Cu_xO surface. (a) Interface between IPr and a copper oxide surface, where IPr interacts with oxygen atom. (b) In the "29" Cu_xO region (outlined by orange lines), the molecular arrangement adopts a striped pattern. In contrast, the molecular arrangement on the "41" Cu_xO region (highlighted by a grey rectangle) exhibits a less ordered molecular arrangement (reproduced from ref. 45 with permission from American Chemical Society, copyright 2022). (c) and (d) Potential energy surface (PES) and self-assembly of IPr on the Si(111)–B surface. (c) Color-mapped PES of IPr on Si(111)–B, illustrating the lowest energy states at the adatom sites (depicted in the darkest purple), with dashed horizontal lines indicating the minimum energy pathways for diffusion between adatoms. (d) STM image of the Si(111)–B surface with 1 ML IPr (reproduced from ref. 75 with permission from Nature Chemistry, copyright 2021). (e) and (f) Self-assembly of cyCAAC on the Si(111)–B surface at different doping concentration: (e) Dopant concentration above 2×10^{-19} cm^{-3}, and (f) dopant concentration above 6×10^{-19} cm^{-3} (reproduced from ref. 24 with permission from The Royal Society of Chemistry, copyright 2023).

energy surfaces (PESs) using DFT calculations. Their results revealed significant differences in the distribution of IPr adsorption energies (Fig. 7c): The Si adsorption atom sites (located at $\sqrt{3} \times \sqrt{3}$ surface lattice angular points) are the lowest-energy and most favorable adsorption sites, and the presence of a high-energy barrier around them prevents the molecules from moving laterally. The horizontal dashed line marks the minimum energy path (MEP) for diffusion of the molecules between adsorbed atoms, indicating that IPr needs to overcome a diffusion energy barrier of about 0.3 eV to migrate on the surface. This energy distribution feature suggests that unlike forming ballbot species on gold surfaces, IPr tends to be immovable at certain adsorption sites to form ordered monolayer structures on the silicon surface (Fig. 7d). Zielinski *et al.* investigated the self-assembly of cyCAAC on an Si(111)–B surface with different dopant concentrations. They found that in systems with higher dopant concentration (6×10^{-19} cm^{-3}), the NHC assembly exhibited larger domains and higher orderliness (Fig. 7e and f).[24] The predominance of doped silicon-based technologies in modern semiconductor manufacturing underscores the significance of this discovery, which establishes a framework for precision-engineered organic molecular functionalization of silicon interfaces and unlocks novel pathways for next-generation optoelectronic device development.

4.3. Co-assembly of NHCs with other molecules on the surface

While the self-assembly of NHCs has been extensively studied, exploring their co-assembly with guest molecules opens new avenues for designing multifunctional and synergistic surface structures. In co-assembly processes involving two molecular species, achieving a balance of intermolecular atomic-scale interactions is fundamental.[76] Typically, when two different kinds of molecules are co-deposited on a surface, they exhibit a tendency to segregate into single-component self-assembled islands, where each island contains only one species.[77] However, ballbot-type NHCs, characterized by their high surface mobility, can deviate from this behaviour to form mixed self-assembled lamellae structures.[78] In mixed self-assembly systems, the selection of molecular species is crucial, as their binding affinity, steric effects, and intermolecular interactions collectively determine the final arrangement. Knecht *et al.* introduced a strategy employing Ru porphyrins to regulate the orientation and lateral organization of NHCs on Ag(111) (Fig. 8a).[79] Experimental and theoretical structural analyses have shown that the NHC linkage to Ru–TPP induces a significant structure shift, with the Ru centre protruding 0.9 Å outward from the surface—an effect attributed to surface transfer. Within the NHC adsorption layer,

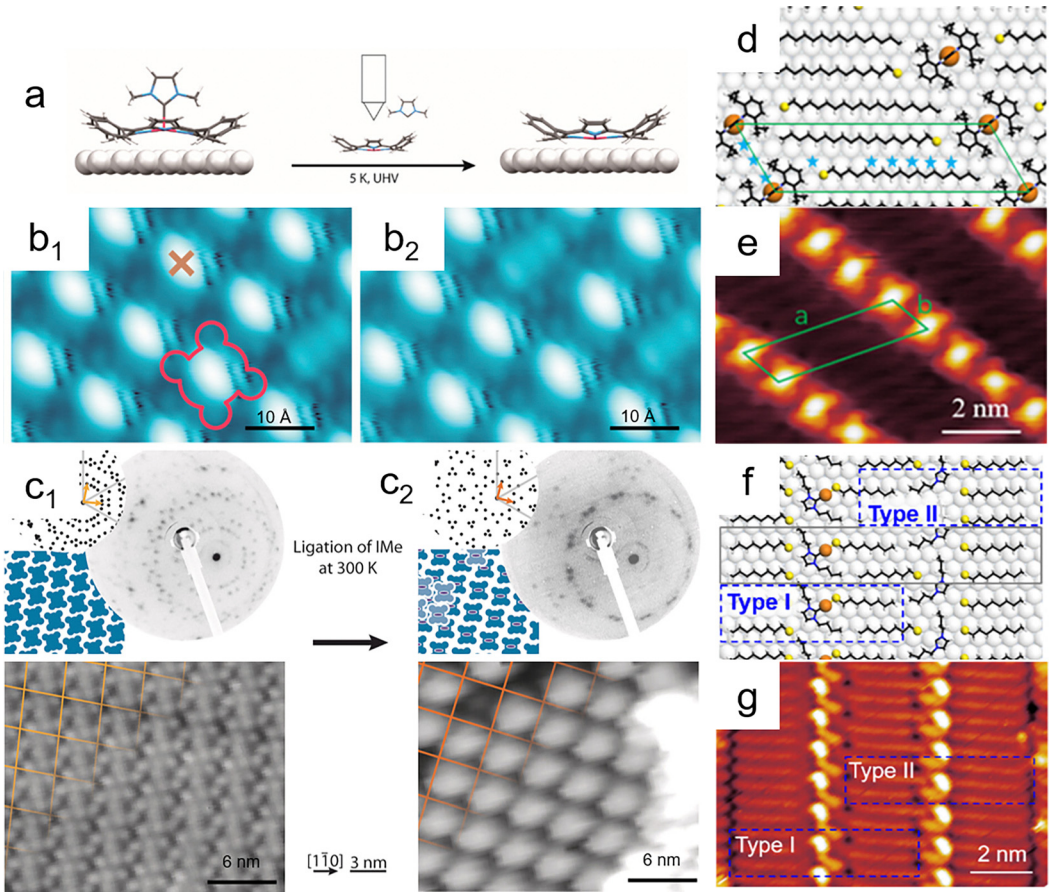


Fig. 8 Out-of-plane and in-plane co-assembly of NHCs with other molecules. (a)–(c) Out-of-plane co-assembly of IMe and Ru–TPP on the Ag(111). (a) Diagram illustrating the IMe functionalization and manipulation procedure. (b_1) STM image of IMe adsorbed on Ru–TPP/Ag(111), with a red outline highlighting a Ru–TPP molecule capped by IMe. The brown cross marks the location where STM tip manipulation occurred. (b_2) STM image of the same region as in (b_1) following the removal of a single IMe molecule, revealing the underlying Ru–TPP molecule (reproduced from ref. 79 with permission from American Chemical Society, copyright 2021). (c) Dynamic reorganization of Ru–TPP molecules on the Ag(111) driven by IMe. (c_1) Ru–TPP molecules self-assembled into a compressed phase. (c_2) Following IMe deposition, a diffraction pattern characteristic of the square phase Ru(IMe)–TPP is observed (reproduced from ref. 80 with permission from John Wiley and Sons, copyright 2021). (d)–(g) In-plane co-assembly of IPr and octadecanethiol (ODT) on the Ag(111) surface. (d) Top view of DFT-optimized atomic structures of the IPr/ODT SAM on the Ag(111) in (e). (e) STM image of the co-deposition of IPr/ODT molecules on Ag(111). (f) DFT optimization of supramolecular complexes indicated in (g). (g) STM image of the IBu/ODT_β SAM constructed by depositing ODT on an IBu-covered silver surface, containing IBu–Ag–ODT, IBu–Ag, and ODT–Ag complexes (reproduced from ref. 78 with permission from American Chemical Society, copyright 2024).

precise manipulation of individual NHCs in ordered arrays can be achieved using STM manipulation, enabling nanoscale "writing" (Fig. 8b_1 and b_2). Furthermore, the NHC functionalization process is reversible, allowing multiple cycles of adsorption and desorption on the same Ru–TPP layer. This well-defined model system facilitates the bottom-up creation of nanopatterns and supports further molecular manipulation using STM tips. Another study explored the role of NHCs as molecular porters in assembling Ru–porphyrins on Ag(111). This process elevates the Ru(NHC)–TPP complexes to a second layer, transforming the first layer from a compressed phase to a less dense tetragonal phase (Fig. 8c_1 and c_2).[80] The material transport, thermally activated and specific to the NHC ligands, weakens the bonding of ruthenium atoms to the silver surface. Notably, annealing promotes the removal of NHC ligands, and the structural reorganization is thermally reversible. Such dynamic regulation of molecular layer structures offers promising avenues for the development of new smart materials and molecular devices. In mixed self-assembly systems, the choice of molecular species is critical, as their respective binding affinity, steric effect, and intermolecular interaction play essential roles in determining the overall arrangement.[81] Looking forward, the unique characteristics of NHCs offer exciting opportunities to expand the scope of supramolecular surface functionalization, particularly in the development of reconfigurable interfaces and adaptive nanoscale devices.

Recently, Ren *et al.* investigated the co-assembly of NHCs and thiols on silver surface, unravelling an atomic-scale synergistic approach to forming multifunctional SAMs with enhanced stability, conductivity, and catalytic activity (Fig. 8d–g).[78] The electron-rich nature and high mobility of NHCs, combined with the strong binding strength of thiols, enable the formation of stable lamellar structures with superior anticorrosion performance and low surface resistance. STM and DFT calculations

 This journal is © The Royal Society of Chemistry 2025

reveal that the deposition sequence of flat-lying NHC and thiol impacts the SAM configurations, further influencing the uniformity and functionality of the assembled layers. This work underscores the potential of dual-ligand systems in advancing the precision of surface modification and tailoring material properties at the nanoscale.

4.4. Self-assembly of NHC derivatives

Building upon the foundational understanding of NHCs in surface functionalization, attention has expanded to include NHC derivatives, collectively termed NHX species, which encompass N-heterocyclic olefins (NHOs) and N-heterocyclic imines (NHIs) (Fig. 9a). These derivatives retain the characteristic electronic richness and high σ-donor capabilities of NHCs while introducing distinct structural and electronic features that diversify their interactions with material surfaces. By exploring these NHX species, researchers aim to unlock new molecular design strategies and broaden the scope of surface chemistry applications.

In fact, NHOs, which consist of N-heterocyclic and olefinic groups, exhibit highly polarized and electron-rich double bonds, with resonance structures encompassing both neutral olefins and ionic ylides. These electronic characteristics impart NHOs with stronger σ-donor capabilities in metal complexes compared to NHCs, while exhibiting weak π-back bonding properties. Similar to NHCs, NHOs form self-assembled monolayers through σ-bond interaction with substrate atoms.

The surface orientation of NHOs can also be modulated by *N*-substituents. For example, IMe–NHO binds to individual substrate atoms, forming planar adsorbed polymers that self-assemble into dimers, trimers, and tetramers, in contrast to IMe, where multiple molecules bind to a single substrate atom (Fig. 9b and c).[82] Comparative XPS studies of IMe–NHC and IMe–NHO reveal that NHO has higher thermal stability than NHC (Fig. 9d) and transfer a greater net electron charge to the surface. STM studies on Si(111)–B surfaces (Fig. 9e) revealed the existence of three molecular arrangement patterns in IMe–NHO monolayers: (1) a green labelled hexagonal $2\sqrt{3} \times 2\sqrt{3}$ periodic arrangement; (2) a blue labelled hexagonal 3×3 periodic arrangement; and (3) a yellow labelled rectangular $3 \times 2\sqrt{3}$ periodic arrangement (which corresponds to the substrate symmetry).[83] Their results indicate that the monolayer structure contains long-range disordered features but locally ordered molecular clusters. Ren *et al.* also examined the self-assembly of NHIs on metal surfaces such as Cu(111) and Ag(111).[84] The assembly order of NHIs is determined by both the molecular structure and substrate species. For example, IPr–NHI form regular hexagonal self-assembly in an upright conformation on the Cu(111) (Fig. 9f), whereas BIMe–NHI form striped self-assembly on Cu(111) and triangular self-assembly on Ag(111).

The reported NHX species highlight their capability in the rational design of multifunctional surfaces, combining the strengths of NHCs with the unique properties of NHOs and NHIs. One promising direction is the modification of the electron-rich carbene carbon in NHCs by incorporating different linkage atoms or functional groups. This approach could

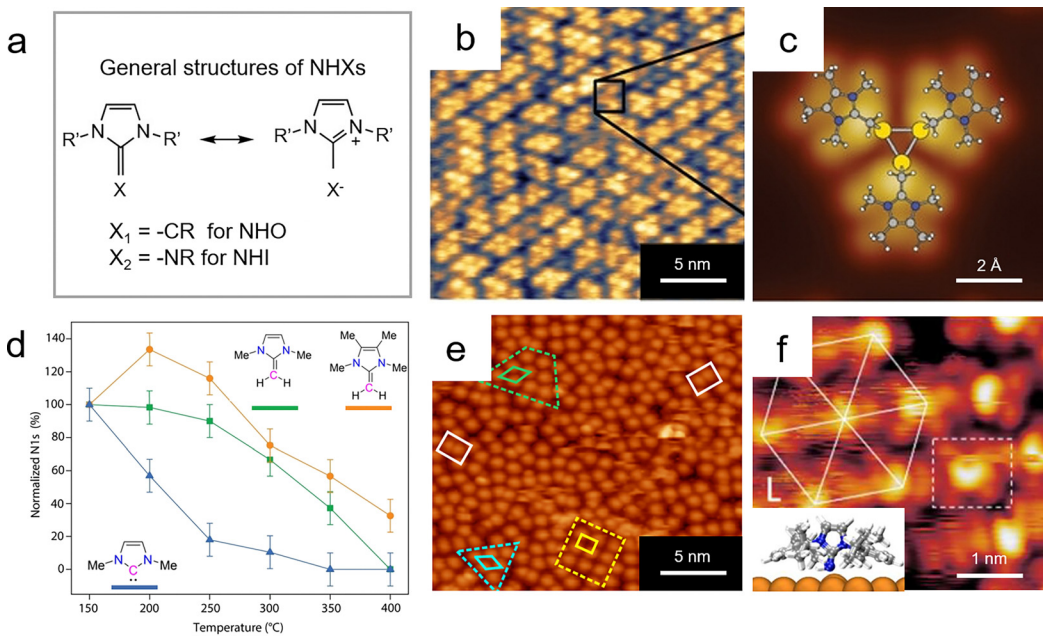


Fig. 9 NHX as molecular modifiers for material surfaces. (a) General structures of NHX, including NHO and NHI. (b) STM image showing NHO self-assembly on Au(111), forming trimeric and tetrameric structures. (c) NHO–Au trimer: each NHO is linked to a single gold atom, unlike the NHC trimer where three NHC molecules share one gold atom. (d) Normalized N1s XPS peak intensities for NHO-1 (green), NHO-2 (orange), and NHC (blue), highlighting the enhanced thermal stability of NHO in comparison to NHC (reproduced from ref. 82 with permission from John Wiley and Sons, copyright 2023). (e) Self-assembled NHO monolayers on Si(111), exhibiting well-defined periodic structures. The green and blue regions represent hexagonal arrangements with $2\sqrt{3} \times 2\sqrt{3}$ and 3×3 periodicities, respectively, while the yellow boxes mark areas with a rectangular pattern (reproduced from ref. 83 with permission from John Wiley and Sons, copyright 2023). (f) Commensuration of the self-assembled NHI monolayer on Cu(111), showing a highly ordered self-assembly pattern (reproduced from ref. 84 with permission from American Chemical Society, copyright 2024).

enable precise tuning of molecular adsorption and self-assembly further optimizing their surface functionalization capabilities.

5. Chemisorbed framework polymers *via* on-surface reactions

While molecular self-assembly has showcased the potential of NHCs in creating ordered surface structures, advancing beyond non-covalent interactions is essential for achieving robust, functionalized materials.[85] Through on-surface reactions, NHCs can function as building blocks for the construction of covalently bonded frameworks, such as metal–organic frameworks (MOFs) and covalent organic frameworks (COFs).

Ren *et al.* employed a modular approach to synthesize one-dimensional covalent-organic chains using ballbot-type NHC units (Fig. 10a).[23] This bottom-up synthesis involved the deposition of dihalogenated NHCs on gold surfaces, followed by annealing under ultrahigh vacuum condition to facilitate Ullmann coupling, thereby yielding linear ballbot-type 1D polymers (Fig. 10b and c). These polymers exhibited alignment along certain crystallographic directions on the Au(100) surface, with their mobility and orientation being controllable through adjustments in precursor activation energy and mobility. Ballbot-type NHCs, owing to their distinctive binding mode, play a pivotal role in enhancing polymer surface mobility, enabling them to overcome steric hindrance and facilitating the formation of incommensurate structures. The NHC polymers aligned along three distinct directions on the surface, separated by 60° intervals, consistent with the symmetry of the Au(100) surface. Statistical analysis of polymer distribution revealed a preferential alignment along the [011] direction, which was in agreement with the lowest energy configuration predicted by theoretical models. Notably, the use of $IPrBr_2$ as a precursor, in conjunction with mild annealing temperature, enabled controlled polymer alignment before the polymers became too long and immobile. The higher activation barrier for C–Br bond cleavage relative to C–I bond cleavage provided an additional degree of control over the polymerization process. These findings enhance our understanding of the on-surface synthesis of NHC-based polymers and open new opportunities for designing novel supramolecular architectures and advanced materials with applications in nanotechnology and catalysis.

Subsequently, Qie *et al.* made significant strides in this field by synthesizing and characterizing low-dimensional NHC

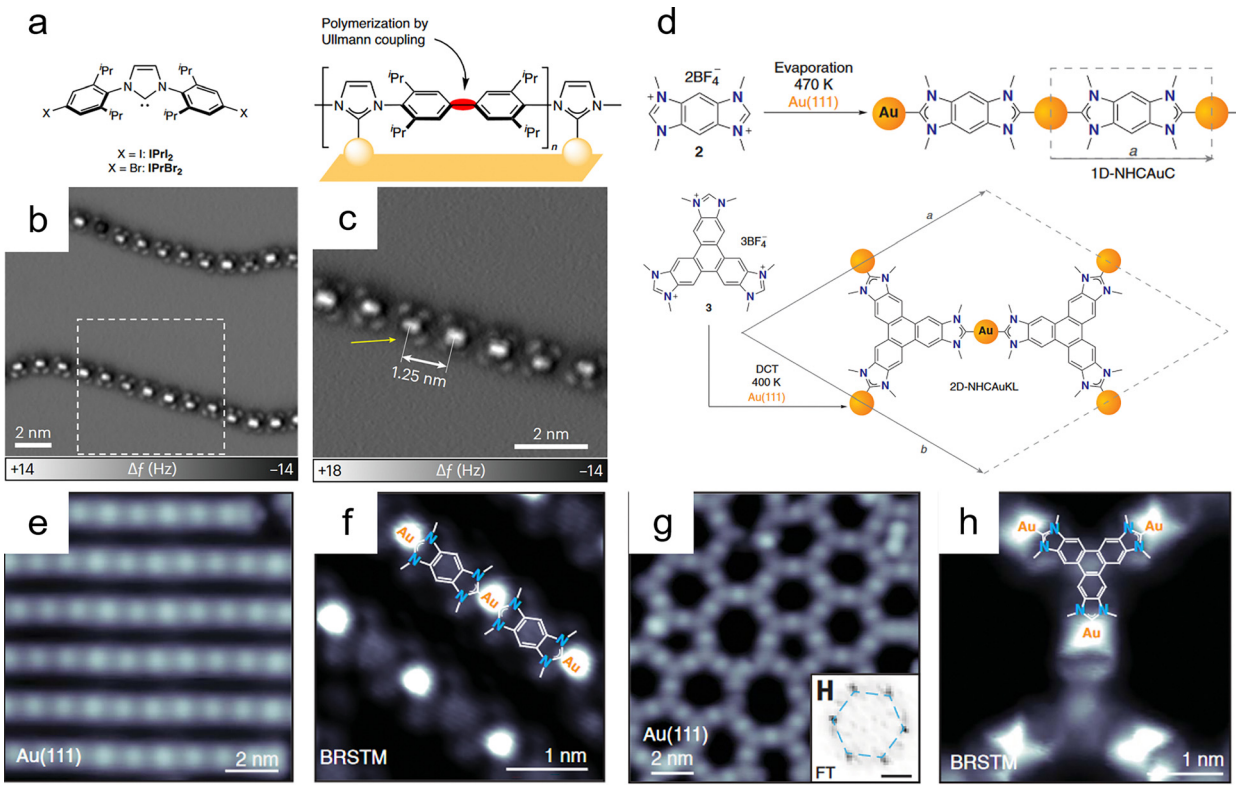


Fig. 10 On-surface reactions for the construction of upright and flat-lying NHC polymers on Au(111). (a) Bottom-up synthesis of ballbot-type 1D-NHC chains on Au(111) *via* Ullmann coupling reaction. (b) and (c) High-resolution nc-AFM images of the NHC polymers, with an inter-unit distance of 1.25 ± 0.05 nm, excluding metal–organic complexes containing Au adatom due to the larger inter-unit distance (reproduced from ref. 23 with permission from Springer Nature, copyright 2023). (d) Schematic representation of the surface-mediated *in situ* synthesis strategy for flat-lying NHC polymers. (e)–(h) STM topographic images of flat-lying 1D-NHC chains (e) and (f) and a flat-lying 2D-NHC network (g) and (h) on Au(111), showing atomic-scale NHC arrangements linked by Au atoms (reproduced from ref. 40 with permission from The American Association for the Advancement of Science, copyright 2024).

This journal is © The Royal Society of Chemistry 2025

lattices, thus further demonstrating the capability of NHCs for constructing extended MOFs (Fig. 10d).[40] They designed trivalent NHC ligands based on a triphenylene core, enabling the formation of 1D NHC–Au chains (Fig. 10e and f) and 2D NHC–Au Kagome lattices (Fig. 10g and h) through *in situ* synthesis on Au(111). Bond-resolved STM confirmed the lattice structures, allowing for detailed characterization of the periodic arrangements at the atomic level. These low-dimensional NHC lattices, synthesized *via* a modular approach, underscore the versatility and robustness of NHC ligands in constructing well-ordered extended molecular networks on the surface. The synthesis of these lattices was facilitated by the C–Au–C bonding interactions within NHC–Au–NHC junctions, which provided a stable foundation for the assembly of both 1D and 2D structures.

The modular design of the molecular precursors allowed for precise control over the formation of these lattices, with unique molecular interactions between NHC ligands and gold adatoms playing a crucial role in promoting the extended framework structure. The resulting 1D NHC–Au chains and 2D Kagome lattices exhibited distinct structural properties, with tunable alignment and periodicity that could be manipulated through variations in precursor design and synthesis conditions. This work highlights the significant potential of NHC-based molecular frameworks for the creation of highly ordered nanostructures on the surface, which can be further developed for a wide range of applications, including molecular electronics and catalysis. The ability to engineer these low-dimensional structures using a bottom-up approach offers exciting prospects for constructing advanced materials with tailored properties, driven by the unique interactions between NHC ligands and transition metal surfaces. The on-surface reaction of the molecule is primarily governed by key aspects such as the substrate, molecular structure, and external conditions.[86] The substrate's species and properties, including surface activity, defect density, and crystallographic orientation, play a pivotal role in determining reaction pathways and the stability of the resulting frameworks. Similarly, the molecular design of NHCs—encompassing factors such as the size, steric effect, and electronic properties—directly influences their reactivity. External factors, including temperature, light, and electronic fields, further control the kinetics and selectivity of on-surface reactions.

By systematically optimizing these parameters, the full potential of NHC-based surface chemistries can be realized, driving the development of transformative materials for nanoelectronics, catalysis, and high-performance advanced applications.

6. Electron-rich NHCs for modulating surface electronic properties and enhancing charge transfer

Having discussed the binding mode, self-assembly and on-surface reactions of NHCs, it is essential to understand the mechanisms of electronic state redistribution and charge transfer, and how their precise control can optimize catalysis and enable the design of novel electronic devices. The following section will explore the electronic properties of NHCs and the charge transfer dynamics between NHC molecules and their substrates.

Investigating the electronic properties of free NHC is indispensable for elucidating its surface and interfacial characteristics. The carbene carbon in NHCs adopts an sp^2 hybridization, and based on their frontier orbitals, forms both π (out-of-plane) and σ (in-plan) orbitals.[30] The in-plane σ orbital benefits from contributions from the carbon's 2s orbital, thus resulting in a lower energy compared to the out-of-plane π orbital. Singlet NHC carbenes favor the $\sigma^2\,\pi^0$ electronic configuration, where two nonbonding electrons are paired within a single σ orbital. The NHC molecule exhibits strong σ-donating capabilities due to the influence of the nitrogen atom in the carbene's substituent in the singlet ground state. Specifically, as a σ-electron-withdrawing group, the nitrogen atom enhances the s-character of the nonbonding σ orbital, thereby stabilizing the σ orbital while leaving the π orbital unchanged. More importantly, in (N,X)-substituted NHCs (Fig. 2a), the interaction with the symmetric combination of the substituents' lone pairs raises the energy of the vacant π orbital. Given that Hoffman established that a singlet ground state requires at least a 2 eV gap, the inductive effects and mesomeric effects widen the σ–π energy gap, further stabilizing the NHC's singlet ground state.[87]

6.1. NHCs on metal surfaces

When NHCs bind to metal surfaces in either upright or tilted configuration, the orbitals of the carbene carbon interact with the metal's atomic orbitals, resulting in the formation of a covalent bond. Chang *et al.* performed Bader charge and ALMO-EDA analyses of the electronic structure and demonstrated that NHCs, acting as a strong electron donor, transfer electron density from the lone pair of electrons at the carbene centre to the empty d orbital of gold, with negligible π-back-donation.[88] The binding energy for NHC adsorption on Au(111) is typically around 2.05 eV, with slight variations depending on the substituent and the size of the NHC ring. This unique binding mechanism makes NHCs highly effective for enhancing the electronic properties of gold surfaces. Ren *et al.* employed the integrated dual descriptor $F^{(2)}(x)$ to characterize the spatial distribution of the electronic properties of the NHC molecule.[89] As shown in Fig. 11a, a distinct negative peak at the center of the NHC ($x = 0$) indicates strong electron donation from the central carbon atom to the silver substrate through a vertical C–Ag bond. Furthermore, within the NHC molecule, the naphthyl group acts as an electron acceptor, while the mesityl group donates electrons.

As previously mentioned in Section 3.3, NHCs with larger *N*-substituents tend to form ballbot-type configuration. Ren *et al.* investigated the interfacial charge transfer properties between NHC molecules and the Au surface. Their results revealed that the frontier orbital energies of surface-bound NHC($IPrI_2$) are similar to those of the isolated molecule, with the only significant difference being the saturation of unpaired electrons at the carbene site (≈ -0.5 eV). The binding type (ballbot *vs.* non-ballbot) exhibits negligible influence on molecular orbital-projected density of states (PDOS), indicating minimal electronic structure modification at the Fermi level regardless of the binding type. Despite slight structural differences, both

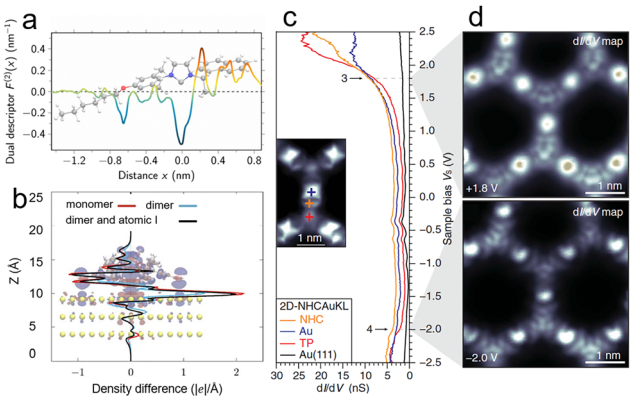


Fig. 11 Electronic properties of NHCs on metal surfaces. (a) Identification of electrophilic and nucleophilic characteristics of different functional groups within the NHC molecule using the dual descriptor function. The naphthyl group is an electron acceptor, while the mesityl derivative acts as an electron donor (reproduced from ref. 89 with permission from American Chemical Society, copyright 2020). (b) Study of charge transfer at the NHC interface, showing that the presence of iodine atoms significantly increases the charge transfer from the molecule to the surface. However, during dimerization, the charge transfer drastically decreases (reproduced from ref. 23 with permission from Nature Chemistry, copyright 2023). (c) and (d) Electronic structure of flat-lying NHC linker states on Au(111). (c) d*I*/d*V* point spectra recorded at specified positions marked by crosses in the inset image. (d) d*I*/d*V* map at +1.8 V (state 3), showing distinct nodal patterns corresponding to the triphenylene core of the NHC linker. The d*I*/d*V* map at −2.0 V (state 4) reveals additional nodal features and a faint signature at the position of the Au atoms, suggesting the involvement of the Au $5d_{xz}$ orbital in the interaction with the NHC linker (reproduced from ref. 40 with permission from The American Association for the Advancement of Science, copyright 2024).

configurations exhibit similar orbital interactions with the surface and share comparable electronic properties, particularly in terms of the unpaired electron saturation at the carbene site.[23] They further examined the interfacial charge transfer of NHCs with different side groups, as well as NHC before and after polymerization, demonstrating that iodine atoms significantly enhance charge transfer, whereas dimerization leads to a sharp reduction in charge transfer (Fig. 11b).

When NHCs with smaller substituents adsorb onto metal surfaces in a flat-lying orientation, they form polymeric structures of NHC–M–NHC linkages. Qie *et al.* utilized unrestricted DFT to investigate the frontier molecular orbitals of the IMe–Au–IMe dimer.[40] Their results suggest that the NHC–M interaction preserves the singlet carbene characteristics and gives rise to frontier orbitals with unusual symmetry. Specifically, the highest occupied molecular orbital (α-HOMO) of the α-electron is a singly occupied molecular orbital (SOMO), with an energy gap of up to 4 eV from the nearest occupied orbitals (β-HOMO and α-HOMO−1). Natural bonding orbital analysis shows that the SOMO in the C–Au–C bonding consists of 19% Au $6p_z$ natural atomic orbitals (NAO) and 20% contributions from each of the two carbon $2p_z$ NAOs. They also identified that the Kagome lattice structure formed by NHC–Au–NHC dimers exhibits unusual metal bands, contributing to an exceptionally low work function (2.3 eV) and enabling significant charge transfer from the Kagome lattice to the gold support (Fig. 11c and d).

Upon adsorption of NHCs onto surfaces, whether they form SAMs or construct MOF lattices, significant charge transfer to the metal substrate occurs, leading to notable changes in the metal's work function.[40,75] Kim *et al.* utilized UV photoelectron spectroscopy to show that the work function of Au surfaces modified with NHCs can be reduced to 3.3–3.5 eV, a decrease of nearly 2 eV compared to the unmodified Au surface.[90] In fact, such regulation of the metal work function primarily occurs through two mechanisms: the formation of an interfacial dipole and the contribution of the NHC molecular dipole. They analyzed the importance of these mechanisms in NHCs with different *N*-substituents through DFT calculations. They found that NHCs with smaller substituents (*e.g.*, BIPr) mainly alter the interfacial dipole through bonding of the carbene carbon with the metal, whereas larger substituents (*e.g.*, Dipp) result in a smaller interfacial dipole but are compensated by a larger molecular dipole.

The ability of NHCs to effectively and stably modulate the metal work function provides a foundation for examining how their surface electronic properties can be harnessed to optimize catalytic performance.[91,92] Ren and colleagues performed DFT calculations to explore the catalytic performance of self-assembled monolayers formed by IBu–Ag–IBu and IBu–Ag–ODT complexes in hydrogen dissociation and acetylene hydrogenation reactions.[84] For hydrogen dissociation, the calculated activation energy of the IBu–Ag–ODT complex is 2.07 eV, whereas the IBu–Ag–IBu complex exhibits a lower activation energy. This difference is attributed to the electronic properties of the Ag(111), which are fundamental to the reaction efficiency. DFT calculations revealed that the Ag(111) modified by the IBu–Ag–IBu complex has a lower work function than that modified by the IBu–Ag–ODT complex, making it easier for electrons to participate in the reaction, thereby leading to higher catalytic activity in hydrogen dissociation. In contrast, for acetylene hydrogenation, the IBu–Ag–ODT complex exhibited higher catalytic activity, with a reaction barrier of 1.42 eV, 0.20 eV lower than that of the IBu–Ag–IBu complex. Atomic structure analysis of the transition state showed that the electronic structure of the Ag adsorption site significantly influences the acetylene hydrogenation efficiency. The projection of the d_{z^2} orbital on the Ag adsorption atom indicated that in the IBu–Ag–ODT complex, the Ag atom's d_{z^2} orbital was closer to the Fermi level, contributing to its higher catalytic activity. The reaction rate-determining step for the IBu–Ag–ODT system was found to be the H_2 dissociation on Ag(111), whereas for the IBu–Ag–IBu system, it was the acetylene hydrogenation step. This difference underscores the significant impact of the molecular configuration of NHC-based complexes on reaction mechanisms, indicating that the structural design and tuning of these complexes can adequately optimize catalytic performance and provide valuable insights for the development of efficient catalysts.

6.2. NHCs on metal oxide and semiconductor surfaces

Having explored the surface electronic properties and interfacial charge transfer of NHCs on metal surfaces, it is also

 This journal is © The Royal Society of Chemistry 2025

important to examine how these two key electronic features are modified when NHCs interact with metal oxide surfaces.

For IPr adsorbed on CuO_x,[45] XPS revealed the emergence of a new high-binding-energy component at 531.3 eV in the O 1s spectrum, accompanied by a shift of the original peak from 529.5 eV to 529.7 eV (Fig. 12a). This indicates that the bonding between IPr and oxygen atoms in copper oxide efficiently modulates the substrate's electronic structure. As highlighted in recent studies,[66,93,94] the adsorption sites and binding strength of NHCs on metal oxide films are strongly governed by the d orbital electron occupancy of metal cations in the oxide lattice. Titanium oxide, with its Ti(IV) cations and vacant d orbital, allows the formation of strong coordinative bonds between Ti and carbene carbon, which further impacts the surface properties, such as the charge transfer and work function (Fig. 12b). On the other hand, iron oxide presents a more complex situation, with the cationic charge states of Fe(II) and Fe(III) leading to varying degrees of interaction with NHCs (Fig. 12c). The partially filled d orbital configuration of Fe(III) cations induces a more covalent (albeit weaker) bonding character compared to TiO_x substrates, resulting in reduced NHC binding strength through diminished orbital hybridization efficiency.

The study of NHC–Si interface electronic properties is also pivotal for the precise modulation of silicon-based circuits. The pioneering work by Johnson and colleagues investigated the adsorption of various NHCs on hydrogen-terminated Si surfaces.[95] Unlike the adsorption of NHCs on metal surfaces, in this case, NHCs form H–C–Si linkages by inserting into the Si–H bond. This reaction requires the NHC to exhibit enhanced nucleophilicity and electrophilicity, exemplified by cyclic amino alkyl carbene (CAAC) and acyclic diamino carbene (ADAC). The replacement of one of the two electronegative and π-donating amino substituents in NHCs with a σ-donating alkyl group enhances the nucleophilicity (σ-donating ability) of CAACs while simultaneously increasing their electrophilicity (π-accepting character) compared to (N,N)-substituted NHCs.[96] This groundbreaking study spurred significant interest in exploring NHC modifications on Si surfaces. Five years later, Franz *et al.* successfully deposited IPr and IMe on Si(111)–B surfaces, investigating the electronic properties of NHC-modified Si(111)–B surfaces.[75] XPS data showed a shift of the C 1s peak to a higher binding energy, indicating bonding between the individual NHC molecule and the reconstructed Si surface's Si adatom. Differential charge densities for two NHCs on Si(111)–B showed that both IPr and IMe donate 0.26 |e| of electron density to the substrate upon adsorption (Fig. 12d_1 and d_2). Upon adsorption of IMe on a Si–$Si(S_5)$ defect site with partially filled Si dangling bond, a similar bonding geometry is observed when IMe interacts with the empty dangling bonds of adatoms on a perfect Si(111)–B surface. However, the electron transfer is significantly enhanced, with IMe contributing 0.33 |e| per molecule to the substrate (Fig. 12d_3). Furthermore, the formation of a monolayer of NHCs directly influences the electronic properties of the Si(111) surface, with significant reduction in the Si surface work function (IPr: 1.63 eV, IMe: 1.86 eV, cyCAAC: 1.86 eV).[24,75]

In 2023, Meier and Schmidt discussed the influence of CAACs on the electron properties and surface work function of monohydride Si(001) surfaces.[48] After inserting into the Si–H bond (type I), CAACs with small substituents generate new electronic states near the valence band maximum (VBM) through surface state interactions, while with bulky substituents

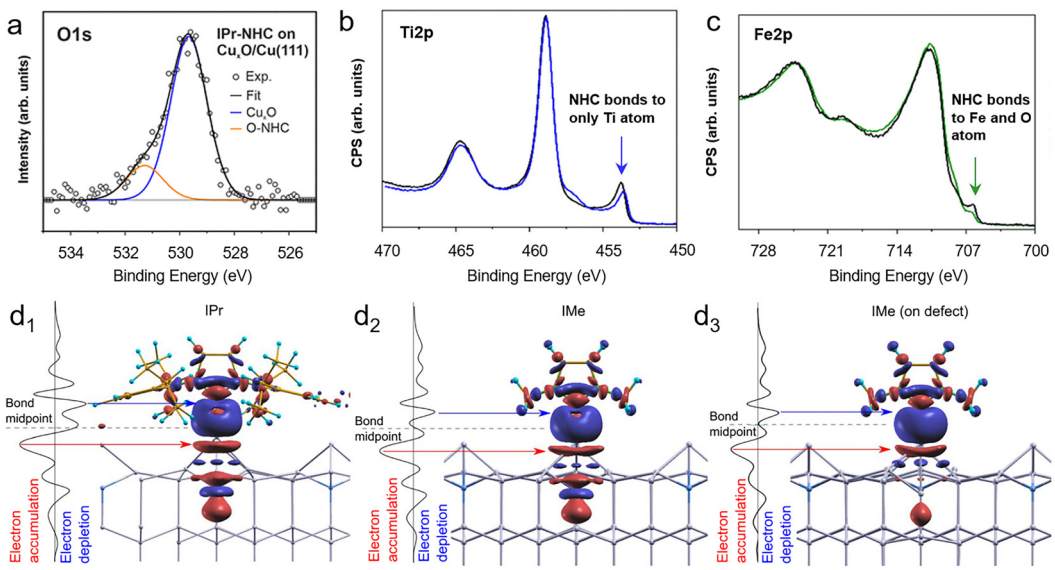


Fig. 12 Electronic properties of NHCs on metal oxide and semiconductor surfaces. (a) Binding of IPr with oxygen atoms on copper oxide, causing a shift in the O 1s core level from 529.5 eV to 529.7 eV (reproduced from ref. 45 with permission from American Chemical Society, copyright 2022). (b)and (c) Ti 2p (b) and Fe 2p (c) core level spectra showing peaks at 456.9 eV (b) and 708.2 eV (c), indicating binding of NHC with the Ti 3d orbital in TiO_x (b) and the Fe 3d orbital in FeO_x (c) (reproduced from ref. 94 with permission from American Chemical Society, copyright 2024). (d) Charge density difference plots for NHCs on Si(111)–B. (d_1) and (d_2) Adsorption of IPr (0.26 |e|) and IMe (0.26 |e|) molecules on a Si adatom. (d_3) Adsorption of IMe (0.33 |e|) on a Si defect site (reproduced from ref. 75 with permission from Springer Nature, copyright 2021).

suppress gap state formation due to steric shielding. When adsorbing on surface defects (type II), CAACs cause more pronounced changes in surface states and result in a significant reduction in work function (1.44 eV) compared to type I (0.5 eV), primarily due to the molecular dipole effect.

The exploration of NHCs at metal oxide and semiconductor surfaces reveals significant insights into the modulation of material electronic properties, particularly through the reduction of surface work function, which is fundamental for improving the efficiency of heterostructures and modulating energy barriers at heterojunction interfaces. Understanding how molecular modifications at these interfaces govern surface bonding, modulate charge transfer, and regulate electronic states is particularly beneficial for semiconductor device engineering, as these interfacial alterations dictate key performance metrics including carrier injection efficiency, light emission characteristics, thermionic emission properties, and adsorption dynamics.[97]

6.3. Conductance of NHC molecular junctions

While the previous section focused on the interface dynamics and electron transfer at the surface level, here we delve into the conductance characteristics of single-molecule junctions, where NHC molecules bridge metal electrodes.

These junctions provide essential understanding of the molecular-level regulation of electron flow and the influence of metal–substrate interactions. Doud *et al.* investigated the electrical conductance of single-molecule junctions formed by NHC molecules and different metal atoms (Au, Ag, Cu) using STM-based break junction (STM-BJ) technology (Fig. 13a).[98] The experimental results uncovered that the conductance of NHC molecular junctions is dependent on the metal atom coordinating with the molecule, with the conductance values varying in the sequence of Au > Ag > Cu (10^{-4}–10^{-3} G_0), where $G_0 = 2e^2/h$ is the quantum conductance (Fig. 13b). This metal dependence highlights the critical role of the metal–molecule interface in determining the electrical transport characteristics of molecular junctions. Additionally, by extending the backbone of the NHC molecule, the conductance value decays exponentially as the molecular length increases. Specifically, for each added phenylene unit, the conductance decreases by approximately 1.8/phenylene (Fig. 13c). Such structural sensitivity demonstrates the importance of molecular architecture in controlling the electrical properties of the junctions. DFT-based transport calculations were employed to elucidate the role of the NHC–M contact in the transport properties of the NHC–M junction. The transmission function analysis revealed several key insights: the highest transmission peak near the Fermi level is attributed to the LUMO of the NHC unit. The transmission spectra projected onto the NHC HOMO are much smaller than the total transmission, suggesting weak coupling between this orbital and both electrodes. In contrast, when the coupling matrices are projected onto the NHC LUMO, the LUMO contributes to over 1 eV of total electronic transmission. This work represents the first single-molecule study of NHC–M interface

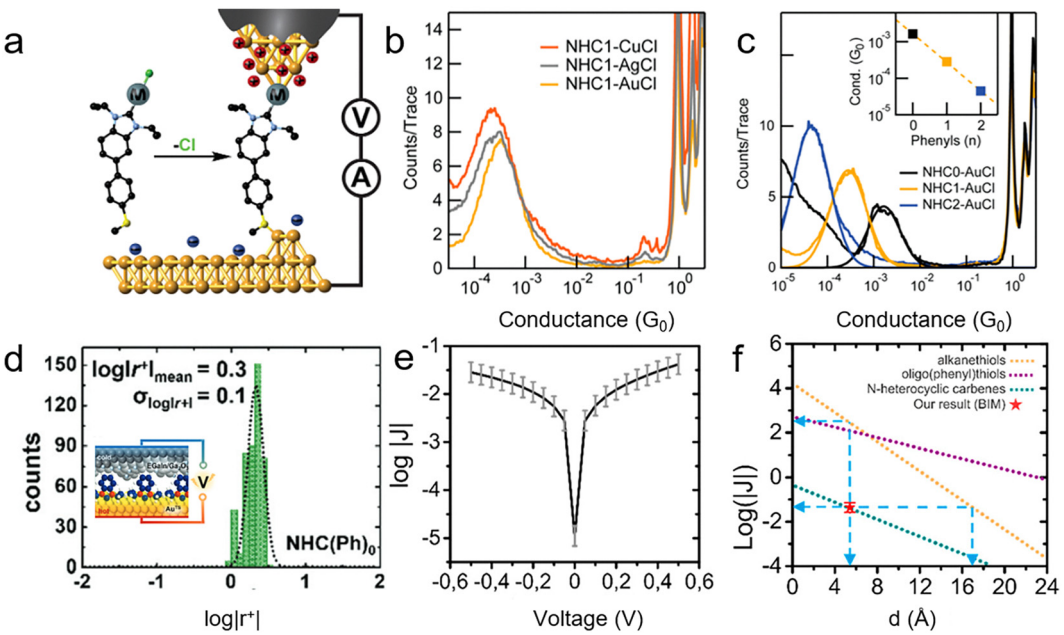


Fig. 13 Conductance of NHC molecular junctions with various structures and linkages. (a) Schematic of STM-BJ measurement setup for single-molecule NHC conductance. (b) 1D conductance histograms of NHC–M (M = Au/Ag/Cu) junctions, showing peak conductance in the order Au > Ag > Cu (10^{-4}–10^{-3} G_0). (c) Conductance histograms of NHC_n–Au (n = 0, 1, and 2) junctions, showing an exponential decrease in conductance with increasing molecular length (reproduced from ref. 98 with permission from American Chemical Society, copyright 2018). (d) A conductance plateau at $\sim 3 \times 10^{-4}$ G_0, confirming the stable formation of a BIiPr–AuCl single-molecule junction (reproduced from ref. 99 with permission from The Royal Society of Chemistry, copyright 2019). (e) Log|J(V)| plot for the BIMe monolayer in an EGaIn junction, showing low current densities in the ±0.5 V range. (f) Comparison of the conductivity of the BIMe/Au monolayer with other known SAMs (reproduced from ref. 100 with permission from American Chemical Society, copyright 2020).

 This journal is © The Royal Society of Chemistry 2025

electronic transfer and conductance regulation, achieving atomic-level determination of the final electrode structure through rational molecular design and synthesis combined with electrochemistry. Kang *et al.* learned the rectification properties of NHC-based SAMs.[99] They found that these SAMs exhibit moderate rectification ratios (between 1.9 and 2.1 at ±1.0 V for the rectification ratio $|r^+|$), suggesting that NHC-based SAMs have certain rectification capability in molecular junctions (Fig. 13d). Further study by Krzykawska *et al.* demonstrated that the conductance of NHC SAMs is approximately four orders of magnitude lower than that of equivalent-length alkylthiol SAMs when measured with EGaIn metal electrodes (Fig. 13e and f).[100] This low-conductance characteristic makes NHC SAMs promising candidates as high-capacitance gate insulators in organic field-effect transistors (OFETs). Overall, the conductance properties of NHC molecular junctions are significantly influenced by the metal atom species and molecular structure, positioning them as potential materials in the fields of nanoelectronics and molecular electronic devices.

7. Surface functionalization and durable organic devices in harsh environments

The remarkable interplay among the NHC molecular binding mode, self-assembly, on-surface reactions, and electronic properties offers a solid foundation for their application in robust organic devices tailored for harsh environments. The distinct properties of NHCs have enabled researchers to explore their integration into advanced materials and devices. These developments pave the way for expanding the role of NHC-based materials in designing high-performance solutions capable of maintaining functionality under extreme conditions.

Amirjalayer *et al.* reported that ballbot-type NHC can perform atomic-level etching and rearrangement on metal surfaces.[101] These nanotools execute two precise tasks on the surface: extracting single surface atoms to create atomic-level defects and pushing individual gold atoms. This pull–push operation enables NHC–Au molecular machines to achieve directed surface reconstruction at the atomic scale. Precise construction of surfaces and interfaces is necessary for enhancing catalytic activity, as the atomic and electronic structures of the catalytic surfaces govern the adsorption and desorption kinetics of reactant molecules.[102] Related studies have reported that the interface structure of nanoparticles can be altered by electron beams and the environmental factors.[103,104] This innovative approach offers new pathways to achieve atomic precision in catalytic surface control by designing and utilizing the mobility of molecular machines.

Molecular machines, when used as catalysts, have been comprehensively reviewed by Fletcher *et al.*, highlighting challenges such as product inhibition, which arises when synthesized products are not efficiently released, a common issue in catalysis that is particularly pronounced at the nanoscale.[105] Unlike macroscopic systems, molecular catalysts are influenced by diffusion and Brownian motion, complicating the development of nanoscale assembly lines where reactants must efficiently transition between catalytic sites. The exceptional stability and unique bonding capability of NHCs make them particularly well-suited for constructing rotor structures, as their ability to form a robust single carbon–metal bond provides a reliable foundation for such designs. Ren *et al.* successfully designed and realized a unidirectional molecular NHC-rotor through a combination of theoretical predictions, computational screening, and experimental validation.[89] By leveraging the structural tunability of NHCs, they identified candidates with asymmetric rotational potentials, a critical feature for achieving unidirectional rotation. The most effective rotor, featuring a mesityl *N*-substituent and a chiral naphthylethyl substituent, demonstrated unidirectional motion confirmed by STM/STS and supported by theoretical and molecular dynamics analyses. The development of unidirectional NHC-based molecular rotors presents a promising strategy for enhancing catalytic reaction efficiency. The unidirectional rotation of the NHC rotor induces dynamic changes in the local microenvironment, facilitating efficient mass transport. Specifically, the rotational motion reduces diffusion resistance by accelerating the delivery of reactants to the catalytic surface and promoting the removal of products from active sites.[106] Additionally, the directional driving force generated by the rotor's motion acts as a "molecular pump", guiding reactants toward active sites and expelling products, thereby minimizing product accumulation and maintaining active site availability.[107] The enhanced stability and tunability of NHC-based rotors, coupled with their ability to optimize interfacial charge transfer, establish them as a robust platform for designing advanced catalytic systems with superior performance in diverse and challenging environments.

NHC-modified surfaces have found significant applications in catalysis. Initially, NHCs were primarily utilized in homogeneous catalysis, with a notable example being the NHC-stabilized ruthenium catalysts used in the olefin metathesis reaction, which contributed to the 2005 Nobel Prize. More recently, the use of NHCs in heterogeneous catalysis has gained increasing attention. In heterogeneous catalysis, NHCs are predominantly employed to stabilize, activate and modify nanoparticles, leveraging their high surface area and recyclability, which enables the design of novel catalytic architectures with enhanced durability and performance. For example, Palazzolo *et al.* reported that the incorporation of NHCs onto Ru/C catalysts dramatically influences C–H activation processes, allowing for enhanced chemo- and regioselectivity in deuteration reactions.[108] The strong metal–carbene bonds formed between NHCs and Ru nanoparticles prevent catalyst deactivation by reducing metal leaching and aggregation. Additionally, the electronic modulation induced by NHCs fine-tunes the catalytic reactivity,[109] selectively favoring C–H deuteration over undesired reduction pathways. Similarly, the work by Valero *et al.* highlighted the impact of NHC-stabilized Ir nanoparticles in hydrogen isotope exchange (HIE) reactions, achieving highly selective deuterium and tritium labeling of anilines with exceptional stability and efficiency.[110] In the field of electrocatalysis, Zhang *et al.* demonstrated that polymeric NHC ligands serve as

robust stabilizers for Au and Pd nanoparticles, preventing nanoclustering under reductive conditions while enhancing CO_2 electroreduction activity and selectivity over proton reduction.[111] Unlike conventional ligands such as thiols and amines, polymeric NHCs form stable metal–carbon bonds that maintain nanoparticle dispersion and catalytic integrity even under prolonged electrochemical cycling. Additionally, the electronic influence of NHCs increases the surface electron density of metal nanoparticles, thereby promoting CO_2 activation while suppressing competing hydrogen evolution, leading to significantly improved faradaic efficiency. Furthermore, Lewis *et al.* illustrated how NHC-functionalized Pd catalysts significantly improve the direct synthesis of H_2O_2, surpassing conventional systems by electronically modulating Pd speciation.[66] The introduction of NHCs alters the electronic properties of surface Pd sites, mimicking the effect of alloying with secondary metals and thereby enhancing catalytic selectivity towards H_2O_2 formation over unproductive hydrogen dissociation. Moreover, the covalent attachment of NHCs to Pd nanoparticles prevents sintering and deactivation, ensuring long-term catalyst reusability and sustained high performance. While several review articles have discussed the role of NHCs in nanoparticle-based catalysis,[19,112] the atomic-level interactions of NHCs with surfaces remain less explored. Studying these interactions on nanoparticles is particularly challenging due to current analytical limitations, which are more pronounced compared to investigations on flat, single-crystal surfaces. To bridge this gap, NHCs on well-defined surfaces are studied to gain insights into their fundamental interactions, which can also provide information on NHC-on-nanoparticle systems, ultimately advancing our understanding of NHCs' role in nanoparticle catalysis at the atomic scale.

The unique electronic structure of NHCs and their coupling with surface electrons offer promising applications in the field of electronic/photoelectronic devices. While the thiol-based molecular junctions have been explored as miniaturized thermoelectric conversion devices, their thermal degradation limits operation to mild thermal conditions (below 323 K).[113–116] Park *et al.* addressed this limitation by replacing thiol anchors with NHCs to develop ultra-stable thermoelectric molecular junctions (Fig. 14a).[117] These junctions maintained a consistent Seebeck effect and thermovoltage up to approximately |1900 μV| at 573 K, exhibiting length-dependent properties (Fig. 14b). This advancement not only deepens the understanding of the role of molecule–electrode interfaces in the Seebeck effect but also provides new strategy for designing high-temperature resistant, efficient organic thermoelectric devices. Rai *et al.* achieved a significant breakthrough in surface plasmon-coupled phosphorescence (SPCP) emission engineering using NHC–platinum(II) complexes.[118] This study employed four classes of NHC–Pt(II) complexes with distinct alkynyl derivatives to achieve, for the first time, an 18-fold phosphorescence enhancement without requiring complex plasmonic architectures. The NHC ligands promoted molecular self-organization into excimers through a stable square-planar geometry, endowing the system with dual-emission characteristics: monomer emission at 498 nm under low coverage and red-shifted excimer emission at 572 nm at high coverage. This unique self-assembly behavior not only modified the dispersion characteristics but also enabled highly directional s-polarized emission (accounting for 80%) *via* the SPCP platform, overcoming the limitation of p-polarization dominance in traditional SPCP techniques. The significant enhancement of phosphorescence quantum efficiency by the NHC ligands, combined with their excimer-forming ability, rendered the complexes promising for applications in white-light-emitting materials and multi-analyte sensing, such as serving as both a sensing probe and an internal reference standard through dual-peak emission. This work provides new strategies for integrating NHC metal complexes into optoelectronic materials and bio-detection.

NHCs have also been employed in surface patterning and fabrication of three-dimensional structures on gold surfaces. Bosse *et al.* performed a cost-effective and environmentally friendly method for producing patterned electronic devices and nanostructures (Fig. 14c).[119] At low concentrations, IPr and IMe exhibit positive resist behaviour, accelerating the electrochemical etching process of NHC-modified gold surfaces (NHC@Au) (Fig. 14d$_1$). This is attributed to reduced gold surface work functions and the formation of NHC–Au complexes, which render gold atoms easier to remove from the surface lattice. Conversely, at higher concentrations, BIMe demonstrates negative resist behaviour by forming densely packed monolayers with low steric hindrance, effectively protecting gold surfaces from etchants (Fig. 14d$_2$). This concentration-dependent property enables precise control over gold surface etching. Amit *et al.* achieved selective deposition of NHC monolayers onto gold microelectrode arrays using pulsed electrochemical deposition.[121] This mask-free technique allows precise deposition of different types of NHC monolayers on individual electrodes, enabling independent adjustment of chemical and electronic properties across an electrode array. Such capability is significant for developing multifunctional molecular electronic devices, permitting localized integration of molecular components with specific electronic responses in integrated circuits.

Since its introduction in the 1990s, surface plasmon resonance (SPR) biosensing has become a pivotal analytical technique for real-time, label-free detection of biomolecular interactions.[122] This optical phenomenon is based on detecting refractive index changes at a gold sensor surface during molecular binding events, allowing direct measurement of binding kinetics and affinity constants without the need for fluorescent or radioactive labeling. At the heart of SPR technology is sensor surface functionalization, which is essential for preserving biomolecular integrity and preventing random orientation upon binding to the gold surface.[123] Conventional systems predominantly utilize thiol-based SAMs due to their ease of preparation and structural adaptability. However, the intrinsic instability of the Au–S bond remains a major limitation, compromising sensor durability and reproducibility.[124] To address these challenges, the development of more robust SPR biosensors leveraging the strong metal-binding affinity of NHCs presents a highly promising alternative.

Gold nanoparticles (Au NPs) protected by NHC ligands exhibit unique SPR characteristics. As demonstrated by the

 This journal is © The Royal Society of Chemistry 2025

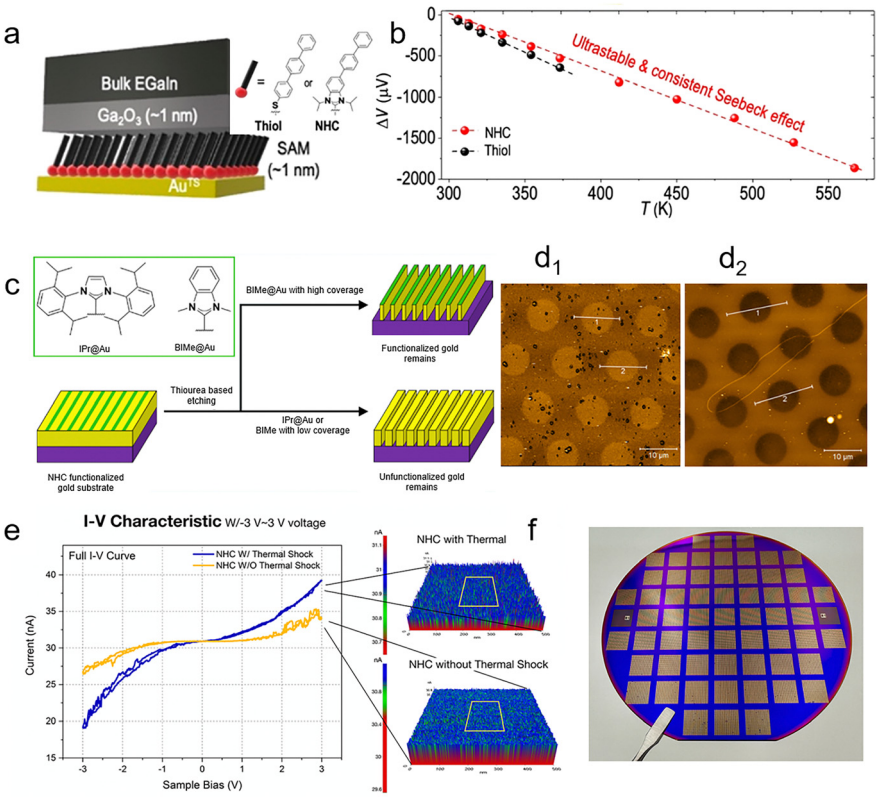


Fig. 14 Surface functionalization and robust organic electronic devices for harsh environments. (a) and (b) High-temperature stability and thermopower of NHC-based molecular junctions. (a) Structural design of NHC and thiol-based SAM molecular junctions, using EGaIn liquid metal top and template-stripped gold bottom electrodes to compare anchoring group properties while maintaining consistent molecular backbone and interfacial contact. (b) The linear relationship between overheating voltage (ΔV), heating temperature (T), and temperature difference (ΔT) reveals the stable Seebeck effect of NHC-based SAMs up to 570 K in vacuum, while thiol-based SAMs degrade at 373 K (reproduced from ref. 117 with permission from American Chemical Society, copyright 2022). (c) and (d) NHC as positive etch resists for patterning planar gold surfaces. (d_1) and (d_2) Optical microscopy and AFM images of gold surfaces patterned with BIMe(H)[HCO_3] at different concentrations (20 mM and 5 mM) *via* microcontact printing (μCP) after 3 min of etching. Higher concentration (20 mM) patterns show negative etch resist behaviour, while lower concentration (5 mM) shows positive etch resist behaviour (reproduced from ref. 119 with permission from American Chemical Society, copyright 2023). (e) and (f) Ultrastable 3D heterogeneous integration enabled by NHC self-assembled nanolayers. (e) Current–voltage (I–V) curves from conductive AFM showing that the NHC nanolayer's electrical conductivity stability significantly surpasses that of traditional benzotriazole passivation after thermal shock. (f) Photograph of the NHC nanolayer uniformly deposited on a 4-inch copper electrode *via* electrochemical deposition, demonstrating feasibility for large-scale integration (reproduced from ref. 120 with permission from American Chemical Society, copyright 2024).

research of Salorinne *et al.*, by adjusting the synthesis conditions of NHC–Au NPs (*e.g.*, reaction time and precursor ratios), the nanoparticle core size can be finely tuned within the range of 2.0–4.7 nm, accompanied by gradual evolution of SPR characteristics.[125] Experimental findings reveal that when the particle size reaches approximately 2.4 nm, a weak SPR band emerges at 520 nm, while particles of 3.3 nm exhibit an SPR peak that red-shifts to 533 nm, indicating entry into the typical plasmonic size range. Notably, even under weak SPR conditions (such as 2.4 nm particles), NHC–Au NPs can still absorb 532 nm pulsed laser energy, and the localized surface plasmon resonance effect converts light energy into heat energy. This thermoelastic expansion process generates detectable ultrasonic signals *via* the photoacoustic effect, which may be attributed to the unique electronic properties of the NHC ligands that modulate the localized surface plasmon resonance. Moreover, the exceptional protective nature of the NHC ligands endows them with excellent stability—larger particles containing SPR remain stable for 7 days in physiological saline (150 mM NaCl, pH 8), exhibiting only a sharpening of the SPR peak, whereas nanoparticles protected by traditional thiol ligands tend to aggregate or be etched under these conditions. In acidic environments, NHC–Au NPs can undergo ripening in organic solvents to form larger plasmonic particles (4.7 nm), a process that is reversible and offers a new strategy for constructing dynamically tunable plasmonic systems. These characteristics demonstrate that NHC ligands not only enable precise SPR modulation through size control, but their unique chemical stability and surface adaptability further expand the application boundaries of plasmonic nanomaterials in bioimaging.

Moreover, based on the study by Crudden *et al.*, NHC films demonstrate remarkable advantages for SPR biosensing.[37] Their unique upright molecular orientation and densely packed monolayer structure effectively minimizes nonspecific adsorption. For instance, during phospholipid bilayer formation, vesicle residues are nearly undetectable, significantly enhancing

detection sensitivity (Bovine serum albumin adsorption reduced by over 50%). Compared to traditional thiol-based films, NHC films exhibit superior chemical and thermal stability, with a desorption energy of 158 ± 10 kJ mol^{-1}, and maintain functional integrity under extreme pH conditions. Experimental results show that NHC-based sensors exhibit excellent linearity in melittin concentration response ($R^2 > 0.99$) and higher reproducibility across diverse buffer systems (pH 5.0–10.0), with a relative standard deviation (RSD) below 15%. Furthermore, the thermal stability of NHC films supports long-term storage at 65 °C and extends the sensor's operational lifespan to over 9 months, markedly surpassing the 3-month validity period of conventional thiol-based chips. These properties position NHC films as an ideal candidate for next-generation biosensing interfaces.[126,127] In 2024, Kim *et al.* introduced a novel biosensor platform combining NHC SAMs with graphene nanotransistors (GMFETs).[128] This biosensor exhibited ultra-stable pathogen detection, with a limit of detection (LOD) improved by two orders of magnitude compared to commercial methods, alongside enhanced reproducibility. Lee's study in 2024 demonstrated an ultrastable 3D heterogeneous integration strategy using NHC self-assembled nanolayers for Cu–polymer bonding.[120] Electrochemical deposition of NHC on 4-inch Cu electrodes forms covalent Cu–C bonds with exceptional thermal stability (no phase transition up to 270 °C). Compared to benzotriazole (BTA), NHC exhibits lower post-thermal shock surface roughness (5.973 nm *vs.* 7.557 nm) and stronger adhesion (34.77 nN *vs.* 16.31 nN) due to lateral polymerization-enhanced crystallinity. Fig. 14e reveals that NHC maintains stable current–voltage (*I–V*) characteristics (40.12 nA at −1.07 nm tip distance) after cyclic thermal shock, while BTA shows localized conduction (31.16 nA) and microvoid-induced degradation due to thermal instability. Kelvin probe force microscopy (KPFM) confirms NHC's enhanced charge transfer efficiency (work function reduction by 0.447 eV). Wafer-scale transmission electron microscopy and energy-dispersive spectroscopy (TEM-EDS) validation demonstrates defect-free Cu–NHC–Cu vertical stacking with 19.41 nm interfacial layers, enabling CMP-compatible 1-minute bonding at 170 °C. Such an approach provides a robust platform for high-performance 3D IC packaging requiring thermal stability and high-speed signal transmission. (Fig. 14f) These findings underscore the capability of NHC SAMs in advancing biosensor technologies by improving stability, sensitivity, and specificity, paving the way for future innovations in biosensing.

The potential of NHC surface chemistry extends far beyond its current applications. For instance, in sensor technology,[129] the integration of NHC-modified surfaces with advanced transduction mechanisms could lead to sensors with exceptional sensitivity, selectivity, and rapid response times. Furthermore, utilizing molecular self-assembly and nanofabrication techniques,[130] integrated NHC molecular devices and sensor arrays with enhanced accuracy and reliability could be developed. These advancements hold the promise of revolutionizing fields, such as nanoelectronics, nanophononics, and nanobiotechnology, while opening new avenues for precision materials science and technology development.

8. Conclusions, perspective, and directions

In this review, we have explored the surface chemistry of NHCs on metal, metal oxide, and semiconductor surfaces at the atomic scale, focusing on key aspects such as molecular binding modes, self-assembly, on-surface reactions, and electronic properties. The strong covalent bonding at the NHC–surface interface contributes to good stability, and the interplay of factors such as molecular coverage, NHC structure, and substrate composition governs the NHC geometry on surfaces. Ballbot-type NHCs exhibit both strong binding and notable mobility across the surface, enabling dynamic surface interactions. Additionally, the highly ordered self-assembly of chemisorbed NHCs on various surfaces has been systematically discussed, with mechanisms from substrate species, crystallographic orientation, environmental conditions, molecular structures, and intermolecular force highlighted. To address challenges associated with direct deposition of high molecular weight polymers onto surfaces, recent developments in on-surface reactions have facilitated the formation of robust, covalent NHC-based polymers. Furthermore, the electronic properties of NHCs, their charge transfer dynamics with substrates, and the conductance characteristics of single-molecule junctions have been examined, particularly for electron-rich NHCs. Finally, leveraging atomic-scale interface information of NHCs offers promising solutions for maintaining functionality under extreme conditions.

While NHCs demonstrate remarkable stability and tunability on various surfaces, challenges remain in fully understanding the intricate dynamics of NHC–surface interactions and their impact on material properties. Moving forward, it will be crucial to explore the interplay between NHCs and substrates, and further investigate their possibility in catalysis, electronics, and nanomaterials. Looking ahead, the future of NHC research lies in addressing key challenges and exploring untapped opportunities:

(1) Beyond simple NHCs: current studies predominantly focus on simple, commercially available NHC molecules such as IPr and IMes. Future research should explore novel NHC frameworks with diverse electronic, steric and functional properties to unlock new surface chemistry phenomena. Rational design of innovative NHC structures enables them to serve as (electronic) modifiers, surface anchors, networkers or protectors and will facilitate the development of versatile functional surfaces and expand their potential applications in catalysis, electronics, and molecular engineering.

(2) Advancing precise surface characterization methods: to gain a more comprehensive understanding of surface chemistry of NHC, it is essential to integrate multiple atomic-scale characterization techniques in a complementary manner. The synergy of different characterization techniques allows for more accurate probing of surface configurations, electronic states, and intermolecular interactions. By combining high-resolution imaging and spectroscopic methods, it becomes possible to precisely identify adsorption sites, spatial configurations, and interfacial electronic interactions. This holistic approach can not only improve experimental efficiency but also

 This journal is © The Royal Society of Chemistry 2025

uncover novel surface structures and properties hidden within molecular systems. Moreover, understanding surface potential modulation and its influence on molecular conformation switching, adsorption/diffusion dynamics, charge transfer processes, and energy dissipation mechanisms will significantly enhance our ability to manipulate interfacial properties at the atomic scale.

(3) Exploring on-surface reactions based on NHCs: the recently emerging field of NHC-mediated on-surface reactions presents exciting opportunities for tailoring surface functionalities. Future studies should focus on understanding the interaction between reactants and surfaces, optimizing reaction intermediates and transition states, and refining reaction pathways to enhance efficiency and selectivity. Investigating activation mechanisms, spatial constraints, and energy barriers will provide insights into reaction yield optimization and the precise control of surface reaction pathways. Additionally, leveraging surface activity—such as atomic-scale reconstructions, confinement effects, and site-specific reactivity—could enable the design of NHC-based surface polymerization strategies for precision synthesis of functional materials.

(4) Developing robust NHC-based materials: the exceptional robustness of NHC–surface bonds presents an opportunity for designing novel composite materials with unique properties. Exploring the surface chemistry of NHC-functionalized materials under extreme conditions will provide critical insights into their stability and performance for specialized applications. Future studies should explore NHC–surface interactions under extreme conditions, including high temperature, reactive atmosphere, and electrochemical environment, to broaden the applicability of NHC-functionalized surfaces across diverse fields such as chemistry, physics, materials science, engineering, and biology.

By consolidating recent advancements, this review provides a comprehensive framework for exploring the transformative impact of NHCs in surface chemistry and materials science. This research field poses many challenges and requires the collaboration of scientists with different skillset and tools in a highly interdisciplinary manner. As research progresses, NHCs are poised to drive innovation across disciplines, enabling molecule-based atomic-level precision in functional materials design and fostering the development of next-generation technologies in catalysis, information, and materials science.

Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors acknowledge generous financial support from the National Key Research and Development Program of China (No. 2022YFA1204100), the National Natural Science Foundation of China (No. 52250402, 22372046), the Deutsche Forschungsgemeinschaft (SFB 1459) and the Ministry for Culture and Science of North Rhine Westphalia, Germany (International Graduate School for Battery Chemistry, Characterization, Analysis, Recycling and Application (BACCARA).

References

1 K. Wieszczycka, K. Staszak, M. J. Woźniak-Budych, J. Litowczenko, B. M. Maciejewska and S. Jurga, *Coord. Chem. Rev.*, 2021, **436**, 213846.
2 D.-H. Kim, S.-J. Woo, C. P. Huelmo, M.-H. Park, A. M. Schankler, Z. Dai, J.-M. Heo, S. Kim, G. Reuveni, S. Kang, J. S. Kim, H. J. Yun, J. Park, J. Park, O. Yaffe, A. M. Rappe and T.-W. Lee, *Nat. Commun.*, 2024, **15**, 6245.
3 J. C. Love, L. A. Estroff, J. K. Kriebel, R. G. Nuzzo and G. M. Whitesides, *Chem. Rev.*, 2005, **105**, 1103–1170.
4 H. Jiang, Y. He, J. Lu, F. Zheng, Z. Zhu, Y. Yan and Q. Sun, *ACS Nano*, 2024, **18**, 1118–1125.
5 C. Tang, T. Stuyver, T. Lu, J. Liu, Y. Ye, T. Gao, L. Lin, J. Zheng, W. Liu, J. Shi, S. Shaik, H. Xia and W. Hong, *Nat. Commun.*, 2023, **14**, 3657.
6 H. Sirringhaus, *Adv. Mater.*, 2014, **26**, 1319–1335.
7 M. Zhu, C. M. Aikens, F. J. Hollander, G. C. Schatz and R. Jin, *J. Am. Chem. Soc.*, 2008, **130**, 5883–5885.
8 X. Chen, X. Xuan, W. Guo and Z. Zhang, *npj 2D Mater. Appl.*, 2025, **9**, 10.
9 W. Chen, H. Huang, S. Chen, Y. L. Huang, X. Y. Gao and A. T. S. Wee, *Chem. Mater.*, 2008, **20**, 7017–7021.
10 C. Du, Z. Li, X. Zhu, G. Ouyang and M. Liu, *Nat. Nanotechnol.*, 2022, **17**, 1294–1302.
11 D. Li, Q. Lian, T. Du, R. Ma, H. Liu, Q. Liang, Y. Han, G. Mi, O. Peng, G. Zhang, W. Peng, B. Xu, X. Lu, K. Liu, J. Yin, Z. Ren, G. Li and C. Cheng, *Nat. Commun.*, 2024, **15**, 7605.
12 Q. Chen, J. R. Cramer, J. Liu, X. Jin, P. Liao, X. Shao, K. V. Gothelf and K. Wu, *Angew. Chem., Int. Ed.*, 2017, **56**, 5026–5030.
13 D. M. Packwood and T. Hitosugi, *Nat. Commun.*, 2018, **9**, 2469.
14 Y. Hu, W. Yang, Y. Ma, Y. Qiu, W. Wei, B. Wu, K. Li, Y. Li, Q. Zhang, R. Xiao, C. Hou and H. Wang, *Nat. Commun.*, 2025, **16**, 1692.
15 Z. He, H. Bi, B. Liang, Z. Li, H. Zhang and Y. Wang, *Light: Sci. Appl.*, 2025, **14**, 75.
16 B. Huang, X. Tang, Y. Hong, L. Li, T. Hu, K. Yuan and Y. Chen, *Angew. Chem., Int. Ed.*, 2023, **62**, e202306667.
17 J. MacLeod, A. Saywell and P. Yu, *Commun. Chem.*, 2024, 7(291), s42004.
18 D. Tang, G. Lu, Z. Shen, Y. Hu, L. Yao, B. Li, G. Zhao, B. Peng and X. Huang, *J. Energy Chem.*, 2023, **77**, 80–118.
19 M. Koy, P. Bellotti, M. Das and F. Glorius, *Nat. Catal.*, 2021, **4**, 352–363.
20 Q. Chen, J. R. Cramer, J. Liu, X. Jin, P. Liao, X. Shao, K. V. Gothelf and K. Wu, *Angew. Chem., Int. Ed.*, 2017, **56**, 5026–5030.

21 M. N. Hopkinson, C. Richter, M. Schedler and F. Glorius, *Nature*, 2014, **510**, 485–496.
22 P. Bellotti, M. Koy, M. N. Hopkinson and F. Glorius, *Nat. Rev. Chem.*, 2021, **5**, 711–725.
23 J. Ren, M. Koy, H. Osthues, B. S. Lammers, C. Gutheil, M. Nyenhuis, Q. Zheng, Y. Xiao, L. Huang, A. Nalop, Q. Dai, H.-J. Gao, H. Mönig, N. L. Doltsinis, H. Fuchs and F. Glorius, *Nat. Chem.*, 2023, **15**, 1737–1744.
24 R. Zielinski, M. Das, C. Kosbab, M. T. Nehring, M. Dähne, N. Esser, M. Franz and F. Glorius, *J. Mater. Chem. C*, 2023, **11**, 7377–7382.
25 J. B. Ernst, S. Muratsugu, F. Wang, M. Tada and F. Glorius, *J. Am. Chem. Soc.*, 2016, **138**, 10718–10721.
26 S. Dery, S. Kim, D. Haddad, A. Cossaro, A. Verdini, L. Floreano, F. D. Toste and E. Gross, *Chem. Sci.*, 2018, **9**, 6523–6531.
27 A. V. Zhukhovitskiy, M. G. Mavros, T. Van Voorhis and J. A. Johnson, *J. Am. Chem. Soc.*, 2013, **135**, 7418–7421.
28 C. M. Crudden, J. H. Horton, I. I. Ebralidze, O. V. Zenkina, A. B. McLean, B. Drevniok, Z. She, H.-B. Kraatz, N. J. Mosey, T. Seki, E. C. Keske, J. D. Leake, A. Rousina-Webb and G. Wu, *Nat. Chem.*, 2014, **6**, 409–414.
29 G. Wang, A. Rühling, S. Amirjalayer, M. Knor, J. B. Ernst, C. Richter, H.-J. Gao, A. Timmer, H.-Y. Gao, N. L. Doltsinis, F. Glorius and H. Fuchs, *Nat. Chem.*, 2017, **9**, 152–156.
30 D. Bourissou, O. Guerret, F. P. Gabbaï and G. Bertrand, *Chem. Rev.*, 2000, **100**, 39–92.
31 P. De Frémont, N. Marion and S. P. Nolan, *Coord. Chem. Rev.*, 2009, **253**, 862–892.
32 J. Li, Z. Dong, Y. Chen, Z. Yang, X. Yan, M. Wang, C. Li and C. Zhao, *Nat. Commun.*, 2024, **15**, 2338.
33 A. Bakker, A. Timmer, E. Kolodzeiski, M. Freitag, H. Y. Gao, H. Mönig, S. Amirjalayer, F. Glorius and H. Fuchs, *J. Am. Chem. Soc.*, 2018, **140**, 11889–11892.
34 A. J. Arduengo, R. L. Harlow and M. Kline, *J. Am. Chem. Soc.*, 1991, **113**, 361–363.
35 S. Dery, S. Kim, G. Tomaschun, I. Berg, D. Feferman, A. Cossaro, A. Verdini, L. Floreano, T. Klüner, F. D. Toste and E. Gross, *J. Phys. Chem. Lett.*, 2019, **10**, 5099–5104.
36 M. Fèvre, J. Pinaud, A. Leteneur, Y. Gnanou, J. Vignolle, D. Taton, K. Miqueu and J.-M. Sotiropoulos, *J. Am. Chem. Soc.*, 2012, **134**, 6776–6784.
37 C. M. Crudden, J. H. Horton, M. R. Narouz, Z. Li, C. A. Smith, K. Munro, C. J. Baddeley, C. R. Larrea, B. Drevniok, B. Thanabalasingam, A. B. McLean, O. V. Zenkina, I. I. Ebralidze, Z. She, H.-B. Kraatz, N. J. Mosey, L. N. Saunders and A. Yagi, *Nat. Commun.*, 2016, **7**, 12654.
38 J. F. DeJesus, L. M. Sherman, D. J. Yohannan, J. C. Becca, S. L. Strausser, L. F. P. Karger, L. Jensen, D. M. Jenkins and J. P. Camden, *Angew. Chem., Int. Ed.*, 2020, **59**, 7585–7590.
39 E. Amit, L. Dery, S. Dery, S. Kim, A. Roy, Q. Hu, V. Gutkin, H. Eisenberg, T. Stein, D. Mandler, F. Dean Toste and E. Gross, *Nat. Commun.*, 2020, **11**, 5714.
40 B. Qie, Z. Wang, J. Jiang, Z. Zhang, P. H. Jacobse, J. Lu, X. Li, F. Liu, A. N. Alexandrova, S. G. Louie, M. F. Crommie and F. R. Fischer, *Science*, 2024, **384**, 895–901.
41 H. Jacobsen, A. Correa, A. Poater, C. Costabile and L. Cavallo, *Coord. Chem. Rev.*, 2009, **253**, 687–703.
42 L. Stephens, J. D. Padmos, M. R. Narouz, A. Al-Rashed, C.-H. Li, N. Payne, M. Zamora, C. M. Crudden, J. Mauzeroll and J. H. Horton, *J. Electrochem. Soc.*, 2018, **165**, G139–G145.
43 S. Dery, S. Kim, G. Tomaschun, D. Haddad, A. Cossaro, A. Verdini, L. Floreano, T. Klüner, F. D. Toste and E. Gross, *Chem. – Eur. J.*, 2019, **25**, 15067–15072.
44 J. J. Navarro, M. Das, S. Tosoni, F. Landwehr, M. Koy, M. Heyde, G. Pacchioni, F. Glorius and B. Roldan Cuenya, *Angew. Chem., Int. Ed.*, 2022, **61**, e202202127.
45 J. J. Navarro, M. Das, S. Tosoni, F. Landwehr, J. P. Bruce, M. Heyde, G. Pacchioni, F. Glorius and B. Roldan Cuenya, *J. Am. Chem. Soc.*, 2022, **144**, 16267–16271.
46 J. J. Navarro, M. Das, S. Tosoni, F. Landwehr, M. Heyde, G. Pacchioni, F. Glorius and B. Roldan Cuenya, *J. Phys. Chem. C*, 2022, **126**, 17528–17535.
47 S. Zhang, *Nat. Biotechnol.*, 2003, **21**, 1171–1178.
48 L. Meier and W. G. Schmidt, *J. Phys. Chem. C*, 2023, **127**, 1973–1980.
49 A. C. Hillier, W. J. Sommer, B. S. Yong, J. L. Petersen, L. Cavallo and S. P. Nolan, *Organometallics*, 2003, **22**, 4322–4326.
50 F. Ragone, A. Poater and L. Cavallo, *J. Am. Chem. Soc.*, 2010, **132**, 4249–4258.
51 G. Lovat, E. A. Doud, D. Lu, G. Kladnik, M. S. Inkpen, M. L. Steigerwald, D. Cvetko, M. S. Hybertsen, A. Morgante, X. Roy and L. Venkataraman, *Chem. Sci.*, 2019, **10**, 930–935.
52 C. R. Larrea, C. J. Baddeley, M. R. Narouz, N. J. Mosey, J. H. Horton and C. M. Crudden, *ChemPhysChem*, 2017, **18**, 3536–3539.
53 M. J. Trujillo, S. L. Strausser, J. C. Becca, J. F. DeJesus, L. Jensen, D. M. Jenkins and J. P. Camden, *J. Phys. Chem. Lett.*, 2018, **9**, 6779–6785.
54 R. L. Thimes, A. V. B. Santos, R. Chen, G. Kaur, L. Jensen, D. M. Jenkins and J. P. Camden, *J. Phys. Chem. Lett.*, 2023, **14**, 4219–4224.
55 L. Li, S. Mahapatra, J. F. Schultz, X. Zhang and N. Jiang, *Chem*, 2025, **11**, 102290.
56 Y. Zeng, T. Zhang, M. R. Narouz, C. M. Crudden and P. H. McBreen, *Chem. Commun.*, 2018, **54**, 12527–12530.
57 F. Schreiber, *Prog. Surf. Sci.*, 2000, **65**, 151–257.
58 J. K. Gimzewski and C. Joachim, *Science*, 1999, **283**, 1683–1688.
59 C. Gutheil, G. Roß, S. Amirjalayer, B. Mo, A. H. Schäfer, N. L. Doltsinis, B. Braunschweig and F. Glorius, *ACS Nano*, 2024, **18**, 3043–3052.
60 A. Inayeh, R. R. K. Groome, I. Singh, A. J. Veinot, F. C. De Lima, R. H. Miwa, C. M. Crudden and A. B. McLean, *Nat. Commun.*, 2021, **12**, 4034.
61 M. Ternes, C. P. Lutz, C. F. Hirjibehedin, F. J. Giessibl and A. J. Heinrich, *Science*, 2008, **319**, 1066–1069.
62 F. Huber, J. Berwanger, S. Polesya, S. Mankovsky, H. Ebert and F. J. Giessibl, *Science*, 2019, **366**, 235–238.
63 J. Ren, M. Freitag, Y. Gao, P. Bellotti, M. Das, B. Schulze Lammers, H. Mönig, Y. Zhang, C. G. Daniliuc, S. Du,

 This journal is © The Royal Society of Chemistry 2025

H. Fuchs and F. Glorius, *Angew. Chem., Int. Ed.*, 2022, **61**, e202115104.
64 S. Dery, W. Cao, C. Yao and C. Copéret, *J. Am. Chem. Soc.*, 2024, **146**, 6466–6470.
65 Y. Zhou, M. Li, H. Lu, H. Jin, X. Wang, Y. Zhang, S. Shen, Z. Ma, J. Song and Z. Bo, *Adv. Funct. Mater.*, 2021, **31**, 2101742.
66 R. J. Lewis, M. Koy, M. Macino, M. Das, J. H. Carter, D. J. Morgan, T. E. Davies, J. B. Ernst, S. J. Freakley, F. Glorius and G. J. Hutchings, *J. Am. Chem. Soc.*, 2022, **144**, 15431–15436.
67 M. L. Perrin, E. Burzurí and H. S. J. Van Der Zant, *Chem. Soc. Rev.*, 2015, **44**, 902–919.
68 P. Maksymovych and J. T. Yates, *J. Am. Chem. Soc.*, 2008, **130**, 7518–7519.
69 A. Bakker, M. Freitag, E. Kolodzeiski, P. Bellotti, A. Timmer, J. Ren, B. Schulze Lammers, D. Moock, H. W. Roesky, H. Mönig, S. Amirjalayer, H. Fuchs and F. Glorius, *Angew. Chem., Int. Ed.*, 2020, **59**, 13643–13646.
70 E. Angove, F. Grillo, H. A. Früchtl, A. J. Veinot, I. Singh, J. H. Horton, C. M. Crudden and C. J. Baddeley, *J. Phys. Chem. Lett.*, 2022, **13**, 2051–2056.
71 A. Kühnle, *Curr. Opin. Colloid Interface Sci.*, 2009, **14**, 157–168.
72 C. Janiak, *J. Chem. Soc., Dalton Trans.*, 2000, 3885–3896.
73 K. L. Hudson, G. J. Bartlett, R. C. Diehl, J. Agirre, T. Gallagher, L. L. Kiessling and D. N. Woolfson, *J. Am. Chem. Soc.*, 2015, **137**, 15152–15160.
74 O. Acton, D. Hutchins, L. Árnadóttir, T. Weidner, N. Cernetic, G. G. Ting, T. Kim, D. G. Castner, H. Ma and A. K. Y. Jen, *Adv. Mater.*, 2011, **23**, 1899–1902.
75 M. Franz, S. Chandola, M. Koy, R. Zielinski, H. Aldahhak, M. Das, M. Freitag, U. Gerstmann, D. Liebig, A. K. Hoffmann, M. Rosin, W. G. Schmidt, C. Hogan, F. Glorius, N. Esser and M. Dähne, *Nat. Chem.*, 2021, **13**, 828–835.
76 D. Li, S. Huang, J. Ge, Z. Zhuang, L. Zheng, L. Jiang, Y. Chen, C. Chu, Y. Zhang, J. Pan, B. Cheng, J.-D. Huang, H. Lin, W. Han and G. Liu, *J. Am. Chem. Soc.*, 2024, **146**, 33461–33474.
77 H.-Y. Gao, P. A. Held, S. Amirjalayer, L. Liu, A. Timmer, B. Schirmer, O. Díaz Arado, H. Mönig, C. Mück-Lichtenfeld, J. Neugebauer, A. Studer and H. Fuchs, *J. Am. Chem. Soc.*, 2017, **139**, 7012–7019.
78 J. Ren, M. Das, Y. Gao, A. Das, A. H. Schäfer, H. Fuchs, S. Du and F. Glorius, *J. Am. Chem. Soc.*, 2024, **146**, 32558–32566.
79 P. Knecht, B. Zhang, J. Reichert, D. A. Duncan, M. Schwarz, F. Haag, P. T. P. Ryan, T.-L. Lee, P. S. Deimel, P. Feulner, F. Allegretti, W. Auwärter, G. Médard, A. P. Seitsonen, J. V. Barth and A. C. Papageorgiou, *J. Am. Chem. Soc.*, 2021, **143**, 4433–4439.
80 P. Knecht, J. Reichert, P. S. Deimel, P. Feulner, F. Haag, F. Allegretti, M. Garnica, M. Schwarz, W. Auwärter, P. T. P. Ryan, T. Lee, D. A. Duncan, A. P. Seitsonen, J. V. Barth and A. C. Papageorgiou, *Angew. Chem., Int. Ed.*, 2021, **60**, 16561–16567.
81 B. H. Farnum, K.-R. Wee and T. J. Meyer, *Nat. Chem.*, 2016, **8**, 845–852.
82 I. Berg, L. Schio, J. Reitz, E. Molteni, L. Lahav, C. G. Bolaños, A. Goldoni, C. Grazioli, G. Fratesi, M. M. Hansmann, L. Floreano and E. Gross, *Angew. Chem., Int. Ed.*, 2023, e202311832.
83 M. Das, C. Hogan, R. Zielinski, M. Kubicki, M. Koy, C. Kosbab, S. Brozzesi, A. Das, M. T. Nehring, V. Balfanz, J. Brühne, M. Dähne, M. Franz, N. Esser and F. Glorius, *Angew. Chem., Int. Ed.*, 2023, e202314663.
84 J. Ren, M. Das, H. Osthues, M. Nyenhuis, B. Schulze Lammers, E. Kolodzeiski, H. Mönig, S. Amirjalayer, H. Fuchs, N. L. Doltsinis and F. Glorius, *J. Am. Chem. Soc.*, 2024, **146**, 7288–7294.
85 H. Furukawa, K. E. Cordova, M. O'Keeffe and O. M. Yaghi, *Science*, 2013, **341**, 1230444.
86 H. Chen, H. Zhu, Z. Huang, W. Rong and K. Wu, *Adv. Mater.*, 2019, **31**, 1902080.
87 R. Hoffmann, A. Imamura and W. J. Hehre, *J. Am. Chem. Soc.*, 1968, **90**, 1499–1509.
88 K. Chang, J. G. Chen, Q. Lu and M.-J. Cheng, *J. Phys. Chem. A*, 2017, **121**, 2674–2682.
89 J. Ren, M. Freitag, C. Schwermann, A. Bakker, S. Amirjalayer, A. Rühling, H.-Y. Gao, N. L. Doltsinis, F. Glorius and H. Fuchs, *Nano Lett.*, 2020, **20**, 5922–5928.
90 H. K. Kim, A. S. Hyla, P. Winget, H. Li, C. M. Wyss, A. J. Jordan, F. A. Larrain, J. P. Sadighi, C. Fuentes-Hernandez, B. Kippelen, J.-L. Brédas, S. Barlow and S. R. Marder, *Chem. Mater.*, 2017, **29**, 3403–3411.
91 Y. Zhou, C. Fuentes-Hernandez, J. Shim, J. Meyer, A. J. Giordano, H. Li, P. Winget, T. Papadopoulos, H. Cheun, J. Kim, M. Fenoll, A. Dindar, W. Haske, E. Najafabadi, T. M. Khan, H. Sojoudi, S. Barlow, S. Graham, J.-L. Brédas, S. R. Marder, A. Kahn and B. Kippelen, *Science*, 2012, **336**, 327–332.
92 C. G. Tang, M. C. Y. Ang, K.-K. Choo, V. Keerthi, J.-K. Tan, M. N. Syafiqah, T. Kugler, J. H. Burroughes, R.-Q. Png, L.-L. Chua and P. K. H. Ho, *Nature*, 2016, **539**, 536–540.
93 F. Zhang, H. S. Sasmal, C. G. Daniliuc and F. Glorius, *J. Am. Chem. Soc.*, 2023, **145**, 15695–15701.
94 E. Amit, R. Mondal, I. Berg, Z. Nairoukh and E. Gross, *Langmuir*, 2024, **40**, 10374–10383.
95 A. V. Zhukhovitskiy, M. G. Mavros, K. T. Queeney, T. Wu, T. Van Voorhis and J. A. Johnson, *J. Am. Chem. Soc.*, 2016, **138**, 8639–8652.
96 M. Melaimi, R. Jazzar, M. Soleilhavoup and G. Bertrand, *Angew. Chem., Int. Ed.*, 2017, **56**, 10046–10068.
97 M. Garg, B. R. Tak, V. Ramgopal Rao and R. Singh, *IEEE Trans. Electron Devices*, 2019, **66**, 2036–2039.
98 E. A. Doud, M. S. Inkpen, G. Lovat, E. Montes, D. W. Paley, M. L. Steigerwald, H. Vázquez, L. Venkataraman and X. Roy, *J. Am. Chem. Soc.*, 2018, **140**, 8944–8949.
99 S. Kang, S. Park, H. Kang, S. J. Cho, H. Song and H. J. Yoon, *Chem. Commun.*, 2019, **55**, 8780–8783.
100 A. Krzykawska, M. Wróbel, K. Kozieł and P. Cyganik, *ACS Nano*, 2020, **14**, 6043–6057.

101 S. Amirjalayer, A. Bakker, M. Freitag, F. Glorius and H. Fuchs, *Angew. Chem., Int. Ed.*, 2020, **59**, 21230–21235.
102 W. Gao, Z. D. Hood and M. Chi, *Acc. Chem. Res.*, 2017, **50**, 787–795.
103 Y. Kuwauchi, H. Yoshida, T. Akita, M. Haruta and S. Takeda, *Angew. Chem., Int. Ed.*, 2012, **51**, 7729–7733.
104 P. Liu, T. Wu, J. Madsen, J. Schiøtz, J. B. Wagner and T. W. Hansen, *Nanoscale*, 2019, **11**, 11885–11891.
105 L. Van Dijk, M. J. Tilby, R. Szpera, O. A. Smith, H. A. P. Bunce and S. P. Fletcher, *Nat. Rev. Chem.*, 2018, **2**, 0117.
106 P.-L. Wang, S. Borsley, M. J. Power, A. Cavasso, N. Giuseppone and D. A. Leigh, *Nature*, 2025, **637**, 594–600.
107 L. Van Dijk, M. J. Tilby, R. Szpera, O. A. Smith, H. A. P. Bunce and S. P. Fletcher, *Nat. Rev. Chem.*, 2018, **2**, 0117.
108 A. Palazzolo, T. Naret, M. Daniel-Bertrand, D. Buisson, S. Tricard, P. Lesot, Y. Coppel, B. Chaudret, S. Feuillastre and G. Pieters, *Angew. Chem., Int. Ed.*, 2020, **59**, 20879–20884.
109 J. B. Ernst, C. Schwermann, G. Yokota, M. Tada, S. Muratsugu, N. L. Doltsinis and F. Glorius, *J. Am. Chem. Soc.*, 2017, **139**, 9144–9147.
110 M. Valero, D. Bouzouita, A. Palazzolo, J. Atzrodt, C. Dugave, S. Tricard, S. Feuillastre, G. Pieters, B. Chaudret and V. Derdau, *Angew. Chem., Int. Ed.*, 2020, **59**, 3517–3522.
111 L. Zhang, Z. Wei, S. Thanneeru, M. Meng, M. Kruzyk, G. Ung, B. Liu and J. He, *Angew. Chem., Int. Ed.*, 2019, **58**, 15834–15840.
112 M. Ghosh and S. Khan, *ACS Catal.*, 2023, **13**, 9313–9325.
113 P. Reddy, S.-Y. Jang, R. A. Segalman and A. Majumdar, *Science*, 2007, **315**, 1568–1571.
114 P. Gehring, J. K. Sowa, C. Hsu, J. De Bruijckere, M. Van Der Star, J. J. Le Roy, L. Bogani, E. M. Gauger and H. S. J. Van Der Zant, *Nat. Nanotechnol.*, 2021, **16**, 426–430.
115 L. Cui, R. Miao, K. Wang, D. Thompson, L. A. Zotti, J. C. Cuevas, E. Meyhofer and P. Reddy, *Nat. Nanotechnol.*, 2018, **13**, 122–127.
116 S. Park, S. Kang and H. J. Yoon, *ACS Cent. Sci.*, 2019, **5**, 1975–1982.
117 S. Park, S. Kang and H. J. Yoon, *Nano Lett.*, 2022, **22**, 3953–3960.
118 B. Rai, D. Suter, V. Srinivasan, K. Venkatesan and S. S. Ramamurthy, *J. Phys. Chem. C*, 2021, **125**, 16681–16688.
119 F. Bosse, C. Gutheil, D. T. Nguyen, M. Freitag, M. Das, B. J. Tyler, T. Adolphs, A. H. Schäfer, H. F. Arlinghaus, F. Glorius and B. J. Ravoo, *ACS Appl. Mater. Interfaces*, 2023, **15**, 36831–36838.
120 J. Lee, G. Woo, G. Lee, J. Jeon, S. Lee, Z. Wang, H. Shin, G.-W. Lee, Y.-J. Kim, D.-H. Lee, M.-J. Kim, E. Kim, H. Seok, J. Cho, B. Kang, Y.-S. No, W.-J. Jang and T. Kim, *ACS Appl. Mater. Interfaces*, 2024, **16**, 35505–35515.
121 E. Amit, I. Berg, W. Zhang, R. Mondal, H. Shema, V. Gutkin, T. Kravchuk, F. D. Toste, Z. Nairoukh and E. Gross, *Small*, 2024, **20**, 2302317.
122 M. Malmqvist, *Nature*, 1993, **361**, 186–187.
123 M. Mrksich, *Chem. Soc. Rev.*, 2000, **29**, 267–273.
124 J. C. Love, L. A. Estroff, J. K. Kriebel, R. G. Nuzzo and G. M. Whitesides, *Chem. Rev.*, 2005, **105**, 1103–1170.
125 K. Salorinne, R. W. Y. Man, C. Li, M. Taki, M. Nambo and C. M. Crudden, *Angew. Chem., Int. Ed.*, 2017, **56**, 6198–6202.
126 Z. Li, M. R. Narouz, K. Munro, B. Hao, C. M. Crudden, J. H. Horton and H. Hao, *ACS Appl. Mater. Interfaces*, 2017, **9**, 39223–39234.
127 Z. Li, K. Munro, M. R. Narouz, A. Lau, H. Hao, C. M. Crudden and J. H. Horton, *ACS Appl. Mater. Interfaces*, 2018, **10**, 17560–17570.
128 K. H. Kim, S. E. Seo, S. J. Park, J. Kim, C. S. Park, T. Le, C. Lee, Y. K. Kim, H. Kim, S. Jun, J. Kwak, Y. K. Lee, H. Yoon, H. S. Song and O. S. Kwon, *Adv. Funct. Mater.*, 2024, **34**, 2310377.
129 K. Tao, P. Makam, R. Aizen and E. Gazit, *Science*, 2017, **358**, eaam9756.
130 L. Miao, B. Huang, H. Fang, J. Chai, Z. Liu and Y. Zhai, *Adv. Mater.*, 2023, 2305159.

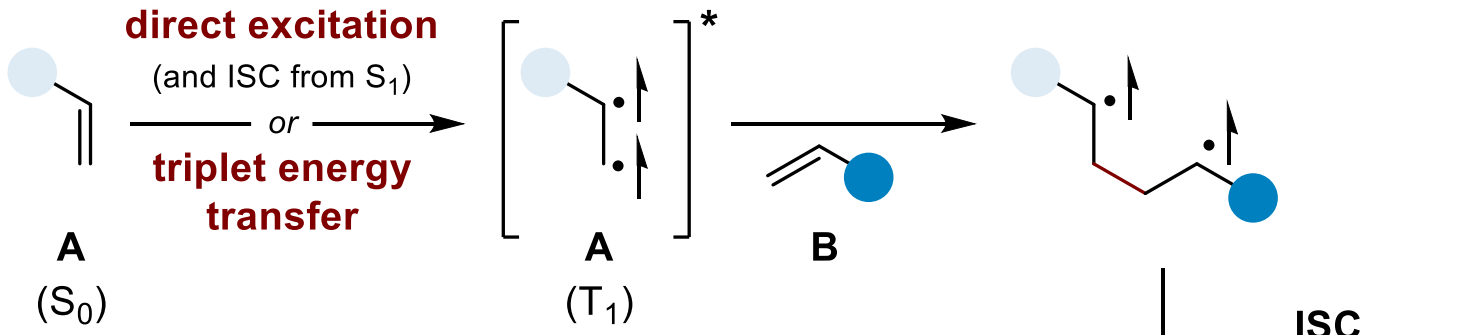
This journal is © The Royal Society of Chemistry 2025


3

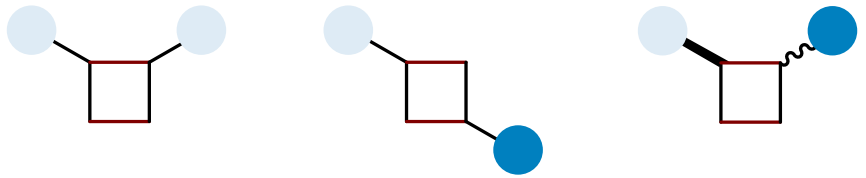
New catalytic tools!!

Known: [2 + 2]

Triplet excited state [2+2]-photocycloadditions

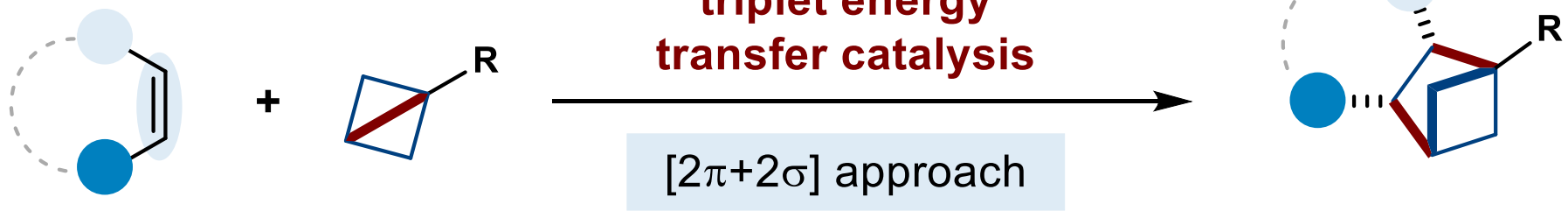


General selectivity challenges:



New & rational

EnT-enabled intermolecular [2π+2σ]-photocycloaddition: This work.



expanding chemical space of [2+2]-photocycloadditions

2π or not 2π...

Kleinmans, … FG, *Nature* **2022**, *605*, 477-482. See also *Nat. Chem.* **2023**, *15*, 535 from **David Procter** and *JACS* **2022**, *144*, 7988 from **Kevin Brown**

Chem Soc Rev

TUTORIAL REVIEW

Published on 03 January 2024
Licence and permissions

Check for updates

Cite this: *Chem. Soc. Rev.*, 2024, **53**, 1068

Energy transfer photocatalysis: exciting modes of reactivity

Subhabrata Dutta, Johannes E. Erchinger, Felix Strieth-Kalthoff, Roman Kleinmans and Frank Glorius *

Received 14th March 2023

DOI: 10.1039/d3cs00190c

rsc.li/chem-soc-rev

Excited (triplet) states offer a myriad of attractive synthetic pathways, including cycloadditions, selective homolytic bond cleavages and strain-release chemistry, isomerizations, deracemizations, or the fusion with metal catalysis. Recent years have seen enormous advantages in enabling these reactivity modes through visible-light-mediated triplet–triplet energy transfer catalysis (TTEnT). This tutorial review provides an overview of this emerging strategy for synthesizing sought-after organic motifs in a mild, selective, and sustainable manner. Building on the photophysical foundations of energy transfer, this review also discusses catalyst design, as well as the challenges and opportunities of energy transfer catalysis.

Key learning points
(1) Concise photophysical background of triplet–triplet energy transfer photocatalysis.
(2) Methods for distinguishing between operative mechanisms in solution.
(3) Decision making on several parameters of triplet–triplet energy transfer processes.
(4) Triplet energies and conceptual advancement in the application of triplet–triplet energy transfer processes.
(5) Direction towards potential underexplored field in triplet–triplet energy transfer photocatalysis.

1. General introduction and fundamentals

Recent years have seen a paradigm shift towards harnessing photons as a traceless source of energy, which has proved to play a vital role in extending the boundaries of sustainable synthesis.[1] Being an environmentally benign energy source for activating substrates and intermediates, visible light can bring added advantages of improved atom economy and green catalysis. Visible-light photocatalysis, despite the implication of being catalytic in photons, is described as a chemical phenomenon occurring under the action of catalytic substance – the photocatalyst (PC) – which absorbs visible light (380–780 nm)[2] and is involved in the transformation of reactants.[3] With the revolutionary introduction of energy conservation measures, researchers started employing visible-light sources like CFLs and LEDs over mercury lamps. This brings numerous scientific and environmental merits – (a) longer lamp life, (b) more energy efficient and low operational costs, (c) better reproducibility and reactivity control, (d) better temperature control, (e) minimized waste disposal, and (f) safer to handle in laboratories. This topic, in general, can be classified majorly based on two mechanistic patterns. In photoredox catalysis,[4–6] a substrate engages in the transfer of electrons (or proton-coupled electron transfer) with an excited state of redox-active photocatalyst. This reactivity mode has been widely applied as a synthetic platform for more than a decade, with multiple reviews discussing a detailed bedrock understanding. Additionally, the merger of photoredox chemistry with HAT catalysis,[7] organocatalysis,[8,9] biocatalysis,[10] and transition-metal catalysis[11] proved to be an enormously beneficial tool for mild cross-coupling reactions, selective functionalization of inert C–H bonds, and many more. Energy transfer catalysis[12–15] (EnT), specifically triplet–triplet energy transfer (TTEnT), enables the transfer of excited state energy from a donor moiety (the photocatalyst) to a suitable acceptor substrate. This is especially attractive if the acceptor is unable to absorb light itself, but exhibits excited triplet state chemistry on its own (Fig. 1A). For clarity—photosensitization—another term often used interchangeably for energy transfer, is technically defined as the photophysical transformation occurring in one entity under the influence of another photoexcited species.[3] In simple words, once the photocatalyst gets excited to its triplet state, it sensitizes or transfers the energy to a suitable substrate, which therefore changes the photophysical properties of the substrate and, in our case, converts it to the triplet state. Compared to photoredox

University of Münster, Organisch-Chemisches Institut, Corrensstraße 36, 48149 Münster, Germany. E-mail: glorius@uni-muenster.de

This journal is © The Royal Society of Chemistry 2024

catalysis, EnT catalysis is still in its early stages and has witnessed a multitude of (conceptual) advancements over the last few years.

This tutorial review aims to summarize recent advances in the field of visible-light-mediated triplet–triplet EnT catalysis from a synthetic and an operational standpoint. We aim to place major emphasis on conceptual advancements in tailoring and designing novel modes of excited-state reactivity, as well as the resulting applications in target-oriented synthesis. To provide the reader with a toolbox for understanding and using energy transfer catalysis, this tutorial review will commence with a brief overview of photophysical and operational considerations of energy transfer catalysis. Eventually, we provide a discussion of existing challenges and possible future developments in the field.

Theoretical background

The fundamental advantage of energy transfer catalysis for synthetic transformations lies in its ability to harness the reactivity of excited (triplet) states. These triplets, owing to their biradical-like nature, offer a variety of unique reactivity modes including isomerizations, cycloadditions and bond cleavages, as we will discuss within the course of this review. However, such triplet states are often unattainable or highly inefficient through direct excitation. To illustrate this, let's delineate the photophysical processes that lead to the population of a triplet state – along with the mechanisms that can hinder its efficiency (Fig. 1B). Notably, this discussion does not only provide insights into the importance of energy transfer catalysis for harnessing these triplet states – it will simultaneously provide important guidelines for the design of efficient photocatalysts. First, when a molecule absorbs a photon, it transitions vertically from its singlet ground state (S_0) to an excited singlet state (S_n). After swift vibrational relaxation, the lowest-energy conformation of the S_1 state is reached, which can differ significantly from the S_0 equilibrium geometry.



Subhabrata Dutta

Subhabrata Dutta received his BSc degree in chemistry from St. Xavier's College, Kolkata, India. Later on, he obtained his MSc degree in chemistry from the Indian Institute of Technology Bombay (IITB), India. During his two-year master's course, he completed his thesis under the supervision of Prof. Debabrata Maiti, working on asymmetric C–H activation. Since 2021, he has been pursuing his doctoral studies in the group of Prof. Frank Glorius, where his research interest focuses on metallaphotoredox catalysis and strain-release-assisted photochemical transformations.



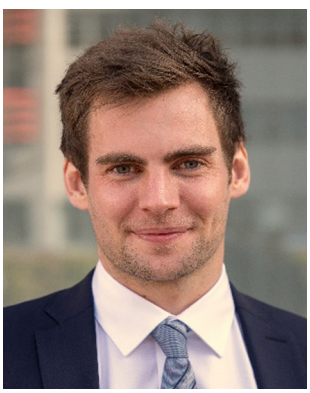
Johannes E. Erchinger

Johannes E. Erchinger studied chemistry at the University of Münster and received his BSc and MSc degree under the supervision of Prof. Frank Glorius. During his master's studies he completed a research internship at the University of Amsterdam in the group of Prof. Joost Reek and a six-month research stay at the Max-Planck-Institute of Coal Research, Mülheim an der Ruhr, in the group of Prof. Benjamin List. Since 2021 he is pursuing his doctoral studies in the group of Prof. Frank Glorius, during which he conducted a six-month research stay at Nagoya University in the group of Prof. Kenichiro Itami. His research interests are focused on photoredox and energy transfer catalysis.



Felix Strieth-Kalthoff

Felix Strieth-Kalthoff is a Schmidt Futures "AI in Science" postdoctoral fellow at the University of Toronto, specializing in laboratory automation and data-driven strategies for molecular discovery. Felix studied Chemistry at the University of Münster, where he obtained his PhD under the supervision of Prof. Frank Glorius in 2021. During his doctoral studies, Felix focused on systematic and data-driven approaches for discovering catalytic reactions, with a particular emphasis on the photophysics and mechanistic key steps in homogeneous photocatalysis.



Roman Kleinmans

Roman Kleinmans completed his BSc and MSc in chemistry at the University of Münster under the supervision of Prof. F. Glorius. During his master's studies, he spent six weeks in the group of Prof. Y. Nakao at Kyoto University as a research intern and later joined the group of Prof. K. Engle at Scripps Research in La Jolla for a six-month research stay, working on nickel catalysis. In 2023, he obtained his doctoral degree as a Kekulé fellow in the group of Prof. F. Glorius, where his research focused on the development of novel [2π+2σ]-photocycloadditions.

From here, a formal spin inversion process, known as intersystem crossing (ISC), leads to the formation of the desired triplet excited state. The spin-forbidden nature of the relaxation greatly enhances the lifetime of such triplet states in solution (often to >100 ns), which allows for efficient bimolecular interaction with substrates or reactants.[16]

The efficiency of populating such a triplet state can vary significantly – and can be assessed in two ways:[17]

(a) The quantum efficiency (or quantum yield) quantifies the number of incident photons that successfully lead to the formation of the desired triplet state.

(b) The energy efficiency relates to the fraction of an individual photon's energy conserved in the triplet state.

With respect to energy efficiency, both vibrational relaxation and ISC lead to thermal energy losses, particularly in organic molecules, where the corresponding (vibrational and electronic) energy gaps can be significant. In turn, the reduction of thermal energy loss by minimizing these energy gaps is a fundamental design principle for effective triplet photocatalysts. Whereas the energy loss due to vibrational relaxation can be tuned through *e.g.* structural rigidity, predicting and designing singlet–triplet energy gaps is a complex problem, requiring advanced photophysical considerations such as exchange and correlation energies, or (dynamic) spin polarization.[18]

On the other hand, the quantum efficiency is restricted by the rates of absorption and ISC, particularly for organic molecules which often possess low extinction coefficients (due to the low oscillator strength of the S_0–S_n transition), and therefore weak absorption. Added to this, ISC in organic molecules is often slow, – except in cases where specific ISC pathways exist (as, *e.g.*, in carbonyl compounds) – leading to quantum yield losses due to competing fluorescence or nonradiative deactivation pathways. This slowness can be attributed to the absence of heavy atoms that accelerate ISC through spin–orbit coupling, and the significant singlet–triplet energy gaps discussed above.[19] Efficient photocatalysts, in turn, must exhibit high extinction coefficients and rapid ISC, explaining the prevalence of heavy-atom Ru/Ir complexes and carbonyl compounds as efficient triplet sensitizers (see Fig. 3).

In contrast, typical organic molecules are not optimized for these criteria, and often feature high energy and/or quantum yield losses. Accessing their triplet states through direct excitation would thus require both short irradiation wavelengths (often down to the UV-B region; 280–315 nm)[2] and high light intensities, which not only raise concerns about cost and safety, but also result in unselective absorption and undesired side reactions, running contrary to the principles of green chemistry.[20,21] To harness the reactivity of such triplet states, triplet–triplet energy transfer catalysis offers a solution to circumvent these challenges by efficiently populating the catalyst's triplet state with minimal losses, allowing for its excitation under mild conditions. In a second step, the excited-state energy is transferred to the substrate, therefore indirectly "exciting" it without traversing the absorption-relaxation-ISC pathway, eventually enabling a more controlled and sustainable transformation. The next section will focus on the mechanism behind this 'indirect excitation' in detail.

Triplet–triplet energy transfer mechanisms

In molecular photophysics, there are two proposed mechanisms for an energy transfer between D* and A, namely Förster[22,23] (dipole–dipole interactions or Coulombic interactions) and Dexter energy transfer[24] (exchange or collisional interactions) (Fig. 2).

Since a triplet–triplet energy transfer (TTEnT) following a Förster-type mechanism violates Wigner's spin conservation rule (formally representing two independent and simultaneous spin reversal events), it does not hold any relevance for TTEnT catalysis. However, FRET – most often observed as a singlet–singlet energy transfer (SSEnT) – has found diverse applications outside catalysis targeting molecular interaction measurement, protein folding dynamics, reaction kinetics, and tracking signaling pathways, which continues to expand.[22] On the other hand, the Dexter theory describes energy transfer as the simultaneous exchange of electrons from the excited state donor (D*, this case excited photocatalyst PC*) to the ground state acceptor molecule (A or quencher) and *vice versa*. This process does formally not involve any change of the overall multiplicity, and is therefore spin-allowed. As the exciton transfer can be characterized as a (double) electron transfer, its kinetics can be also described by the Marcus theory.[25]

The overall summary of these two transfer mechanisms is as follows:

(a) The triplet–triplet Dexter energy transfer asserts the spectral overlap between the phosphorescence emission spectrum of D* and the singlet-to-triplet absorption spectrum of A. However, since it is not common to measure the S_0–T_1 excitation spectra, normal absorption, and emission spectroscopy are not enough to reliably characterize the feasibility of a Dexter-type EnT. Unlike FRET, it does not depend on the oscillator strength of either D or A.[24,26]

(b) The rate of Dexter energy transfer decays exponentially with increasing donor–acceptor separation. Since it relies on orbital overlap, significant rates of energy transfer can only be



Frank Glorius

Frank Glorius is a full professor of organic chemistry at the University of Münster. His research program focuses on the development of new concepts for diverse areas of catalysis such as in photocatalysis, C–H activation, data science, N-heterocyclic carbenes (NHCs) in organocatalysis, surface modifiers, and (asymmetric) arene hydrogenation and on the molecules that are designed and synthesized for applications in biological systems (membrane analogues and drugs) and materials science (semiconductors, battery additives, and heterogeneous catalysts).

This journal is © The Royal Society of Chemistry 2024

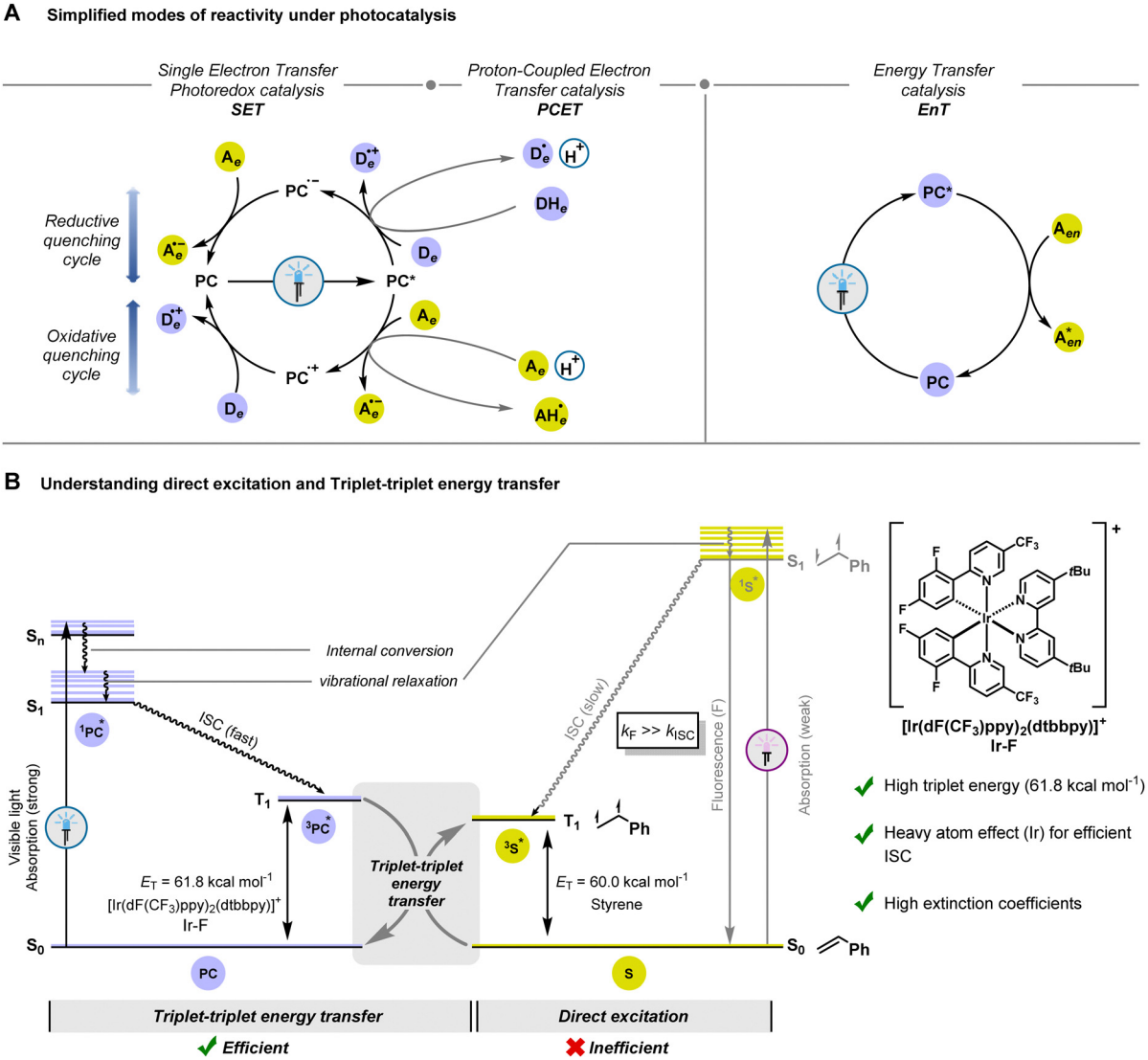


Fig. 1 (A) Mechanistic approaches of photocatalysis, (B) Schematic representation (with an example of styrene and Ir–F) describing the unique aspects of TTEnT and comparison with direct excitation. D: donor, A: acceptor, ISC: instersystem crossing, PC: photocatalyst, S: substrate, S_n: *n*th singlet state, and T_1: first triplet state.

observed for very small R_{DA} (on the order of Angstroms), usually requiring the formation of solvent-shared encounter complexes. As a consequence, TTEnT in solution becomes a diffusion-mediated process.

(c) In most scenarios, the rate of energy transfer is largely governed by its thermodynamic feasibility with ΔE_{EnT} ($E_{T,\,A}$ – $E_{T,\,D}$) < 0.

Commonly used photosensitizers

The choice of a suitable photosensitizer is the most crucial part for the success of energy transfer photocatalysis. There are a few obligatory notes for selecting a catalyst.

(a) The photocatalyst should have a higher triplet energy than the chosen substrate (acceptor). An overview of common catalysts and their triplet energies is given in Fig. 3. Experimental strategies to determine such triplet energies have been previously discussed in the literature.[12]

(b) The photocatalyst should have high intersystem crossing efficiency (S_1 to T_1).

(c) The photocatalyst should also possess a high extinction coefficient.

(d) It should not participate in side reactions with any substrate in its excited or ground state, *e.g.* hydrogen atom transfer or competing biradical reactions is common with ketone-based sensitizers,[27] photodecomposition of benzil,[28] *etc.*

(e) Notably, since excited states are inherently more redox-active than the corresponding ground states, many photosensitizers can also take part in photo-redox processes, depending on the substrate potentials *(vide infra)*.

The above-mentioned photosensitizers usually consist of a single visible-light harvesting unit which is currently limited to a single absorption band in the visible spectrum. The idea of broadband visible light absorption – using the entire solar energy spectrum – holds a promising future in accessing reactive

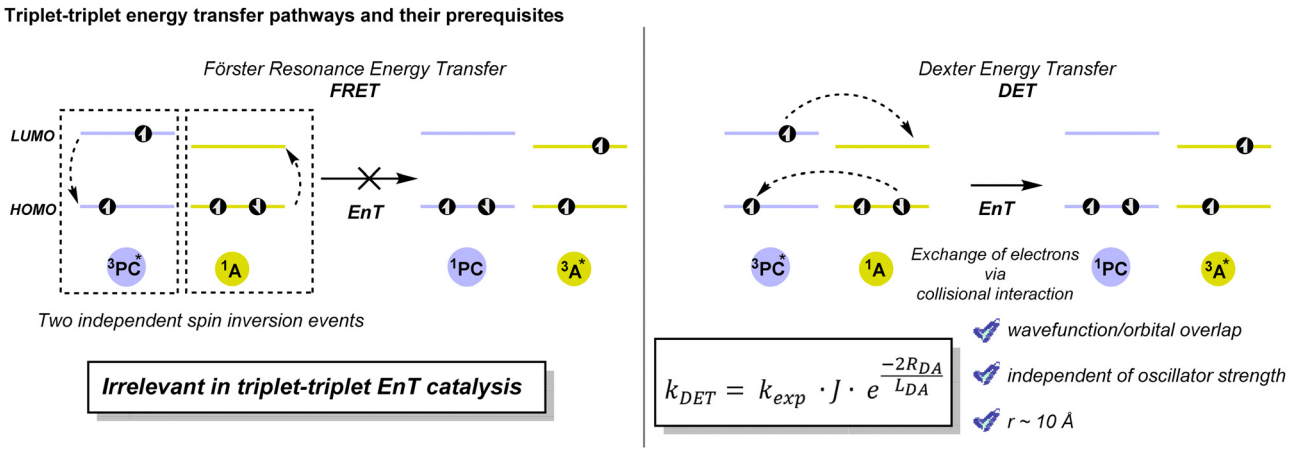


Fig. 2 Schematic and mathematical depiction of energy transfer mechanisms. k_{DET}: transfer rate constant, J: spectral overlap integral, R_{DA}: distance between D and A, L_{DA}: sum of van der Waals radii of D and A, HOMO: highest occupied molecular orbital, and LUMO: Lowest unoccupied molecular orbital.

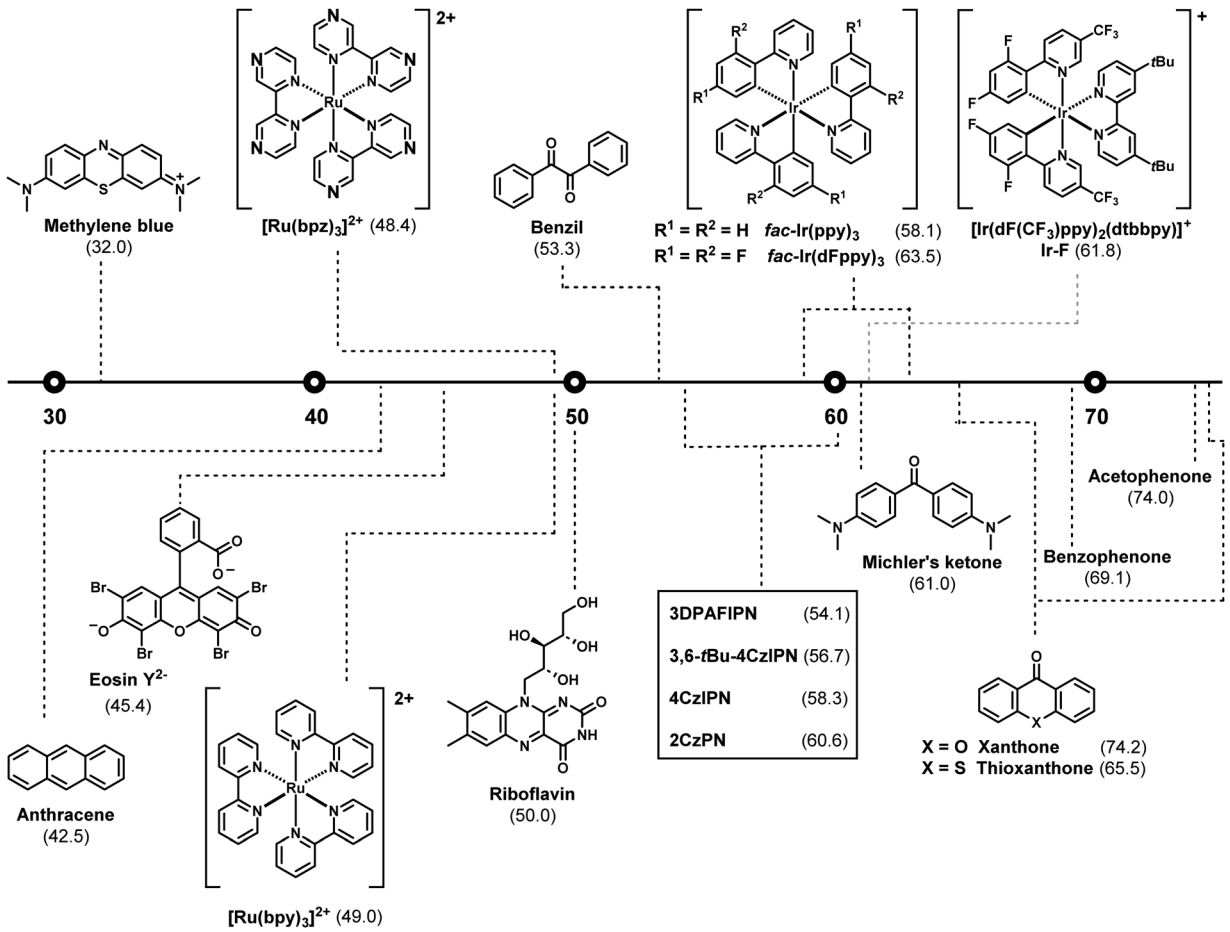


Fig. 3 Extended list of literature known photosensitizers with their triplet state energies (kcal mol^{-1}).[13,27,29–31]

triplet states. In the past few years, broadband light harvesting units in combination with spin converters (like C_{60}) have proven to be beneficial in harnessing triplet states that show phosphorescence up to 700 nm, thereby also covering the desired higher-wavelength end of the optical spectrum.[32] This is achieved using multichoromophore conjugates, where two or more strongly absorbing chromophores (a singlet energy donor and a singlet-energy acceptor as a spin converter) are linked covalently

 This journal is © The Royal Society of Chemistry 2024

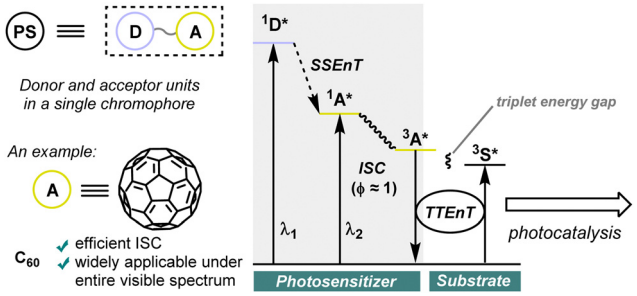


Fig. 4 SSEnT light harvesting for accessing triplet state of a substrate where C_{60} acts as a spin inverter.

(Fig. 4).[33,34] The donor will transfer the energy to the acceptor linker *via* SSEnT, promoting it to the excited singlet state. Owing to the high ISC quantum yield of **A**, the triplet state can be accessed. As an example, this principle has been used to circumvent the narrow absorption bands and long photooxygenation reaction time with $[Ru(bpy)_3]^{2+}$.[35] Moreover, the approach has been successfully applied for generating singlet oxygen[36] and further photochemical transformations.[37] However, these cases will not be discussed thoroughly in this present review, restricting it to TTEnT photocatalysis with mono-chromophore triplet photosensitizers.

Energy transfer catalysis (EnT) or photoredox catalysis (ET)?

The most common question while operating a photochemical reaction is the leading mechanism running behind the scenes. As mentioned before, most photocatalysts have a tendency to take part in both EnT and ET, which makes it challenging to unveil the true mode of reactivity. Nonetheless, there are a series of experimental ways to probe this.

(a) As a first step, the thermodynamic feasibility of electron or energy transfer can be estimated based on redox potentials and triplet energies. That being said, an electron transfer (ET) process may be ruled out based on the reported/measured redox potential values of the substrate and the photocatalyst.

(b) The detection of reactivity under direct excitation with UV light (in the absence of any photocatalyst) also indirectly hints towards reactivity from the substrate's excited state.

(c) A positive correlation between photocatalyst triplet energy and reaction yield is also an indicator of EnT.

(d) In an energy transfer scenario, the encounter complex[38] is independent of the solvent's polarity – unlike in the case of electron transfer, where solvent-stabilized charge separation plays a major role. Thus, the observation of similar quenching constants in Stern–Volmer studies with different solvents would suggest an EnT mode being operative. However, few solvent types can undergo side reactivity with PCs used (will be discussed in the next section).

(e) Transient absorption spectroscopy studies provide more in-depth insights into the nature of the catalyst–substrate interactions. Such studies allow for the simultaneous observation of the catalyst excited state decay, and the emergence of new states. The spectral fingerprint of the latter can then be assigned to a substrate excited state or a cation/anion through separate spectroscopic or spectro-electrochemical experiments.

Problems and troubleshooting in EnT catalysis and mechanisms

The efficiency of a reaction proceeding under energy transfer catalysis is governed by a range of parameters and reaction conditions, including the excitation wavelength, the choice of photosensitizer, solvent, reaction atmosphere (Fig. 5), or stoichiometries. The efficiency at a particular wavelength is often measured by the quantum yield of the process.

Photocatalyst. One of the most common problems in EnT catalysis is the choice of a suitable sensitizer (*vide supra*). The common practice of screening a series of catalysts for optimized reactivity brings down the productivity in general. As discussed above, the suitability of a sensitizer can also be examined with Stern–Volmer luminescence quenching studies, wherein the photoluminescence intensity drops gradually with an increase in the quencher (targeted substrate) amount. To further rule out the possibility of electron transfer from the same catalyst, testing the reactivity with a photosensitizer featuring a higher triplet energy and a limited redox window could serve as a potential option. For an additional mechanistic outlook, a reasonable decline in reactivity after adding an additional known triplet quencher could also suggest a triplet state reactivity.

Triplet energy. Before executing an experiment, the knowledge of triplet energies of both photosensitizer and target substrate provides an initial assessment of EnT feasibility. However, the limited number of available catalysts, *e.g.* owing to constraints in terms of triplet energies or side reactivity, can call for adjustments on the substrate end. The possibility of lowering the triplet energy of the substrate by a suitable transient entity like Lewis acid,[39–41] Brønsted acid,[42] iminium catalysis,[43] or H-bond donor[44] represents a promising direction for expanding the scope of sensitizable substrates. The triplet state energies of those activated systems can be predicted approximately based on screening different known sensitizers or quenchers as reference. DFT calculation contributes as an alternative to an experimentally rigorous and convoluted setup.[12]

Solvent and concentration. The understanding of mechanistic pathways and potential side reactions involved in a catalytic system can render reaction optimization more time-efficient. As an example, the involvement of highly localized

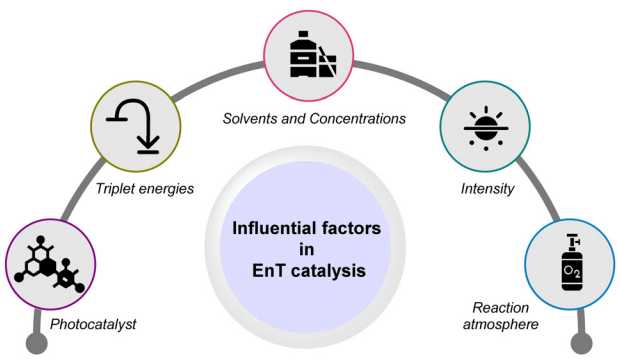


Fig. 5 Prominent decisive components in energy transfer catalysis.

biradical excited states of the substrates could potentially lead to undesired dimerization reactions (*e.g. via* a [2+2] cycloaddition), which can be impeded by reducing the reaction concentration or by adjusting the equivalents of reactants. The solvent type also plays a deciding role in the fate of excited state configurations of photosensitizers and their side reactions. For example, in polar solvents (good H-bond donors), both $^1\pi,\pi^*$ and $^3\pi,\pi^*$ states of thioxanthone (TXT) are stabilized.[27] This creates an energetic mismatch between the $^1\pi,\pi^*$ and $^3n,\pi^*$ which in turn favours fluorescence rather than triplet state population. Moreover, the stabilized $^3\pi,\pi^*$ also takes part in hydrogen atom abstraction. For this reason, alcohol-based solvents must be omitted in the case of TXT. Depending on the direction of change in polarity (either positive or negative) of the excited state, complementary solvatochromism is expected.

Intensity and photodevices. Catalytic reactions proceeding *via* energy transfer are often highly sensitive to the number of photons absorbed, and in turn, to the irradiation intensity. In certain scenarios, increasing the light intensity can compensate for low absorption coefficients or non-productive decay processes, *e.g.* with weakly absorbing di-alkyl ketone catalysts. Generally, the wavelength distribution of the light source (most commonly spectrally narrow LEDs), as well as efficient light transfer (good reflective surfaces, minimized losses through light scattering) are a few key points worth noting when setting up a photochemical reaction, along with efficient temperature control. As a cautionary side note, in specific cases, the reaction mechanism could also be influenced when changing the irradiation power.[45]

Reaction atmosphere. The generation of reactive singlet oxygen ($^3\Sigma^-_g \rightarrow {}^1\Delta_g$ = 22.5 kcal mol^{-1}) through sensitization by various low-energy triplet sensitizers is a cautionary remark.[46] Ground state triplet oxygen is a well-known triplet quencher, and therefore it warrants the reaction to take place in an inert atmosphere, especially for sensitizers with similar triplet excited state energies (*e.g.* ruthenium complexes, methylene blue).

2. Recent applications of triplet–triplet energy transfer in organic synthesis

This section presents a compendium of recently developed concepts in the arena of energy transfer photocatalysis. Starting from the emerging vision of controlling the reactivity of biradical excited states to handling excited state organometallic intermediates, this section will provide an outlook on the journey of different applicability domains of energy transfer photocatalysis. For the interest of readers, a chart of substrate triplet energies will be available for every subsection. This tutorial review aims to show selected examples comprising of breakthroughs, scholarly important findings, and recent advancements made in the past decade.

Cycloadditions

Merging two independent π-systems to a new, cyclic core motif gives an opportunity to rapidly build up molecular complexity and access unprecedented chemical space, lately acclaimed in pharmaceuticals, agrochemical, and materials industries.[47–50] In this vein, unique cycloaddition reactions proceeding *via* energy transfer (for triplet energies, see Fig. 6), importantly, including dearomatization reactions, have been developed in recent years showing the potential of this constructive strategy to build 3D frameworks.[51–53]

In common [2+2]-photocycloadditions, an olefinic motif **I** gets excited to its triplet state **II** *via* energy transfer from the photosensitizer, followed by radical addition to the second (ground-state) olefin partner forming triplet intermediate **III** (Fig. 7). This, upon relaxation (intersystem crossing) and recombination, forms the cycloadduct **V**.

Though direct excitation formally represents an alternative method of generating the required biradical **II**, generally low extinction coefficients and poor ISC quantum yield require shorter-wavelength irradiation and, in turn, decreased selectivity and functional group tolerance.

Dearomatization reactions. Owing to the disruption of aromatic systems, dearomatizing cycloadditions generally hold an endergonic profile. Escape from this thermodynamic equilibrium can be enabled through excited-state pathways. Early examples of cycloaddition reactions wherein molecular oxygen was utilized in its excited singlet state for pericyclic reactions have been discussed extensively.[55] In the past years, many pioneering contributions, from Yoon,[56] Bach,[57,58] and our group[59] demonstrated the synthetic power of this concept for dearomatizing different arenes and heteroarenes, offering bridge- and cage-like polycyclic motifs. Intramolecular reactions take the advantage of having both units in close proximity, but require extensive reactant synthesis, and therefore

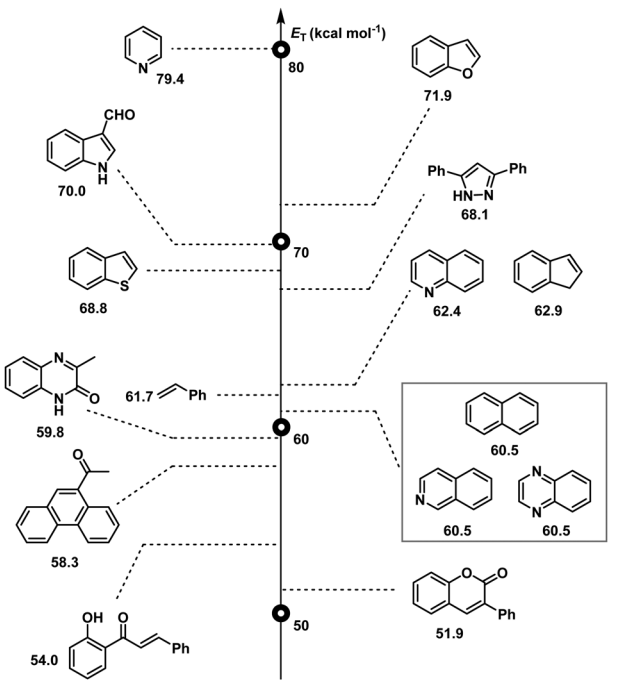


Fig. 6 Triplet energies of organic substrates for cycloaddition reactions.[39,54]

This journal is © The Royal Society of Chemistry 2024

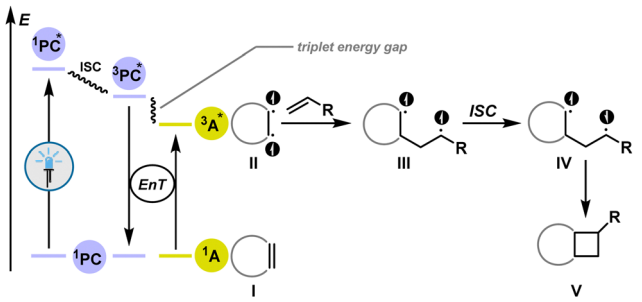


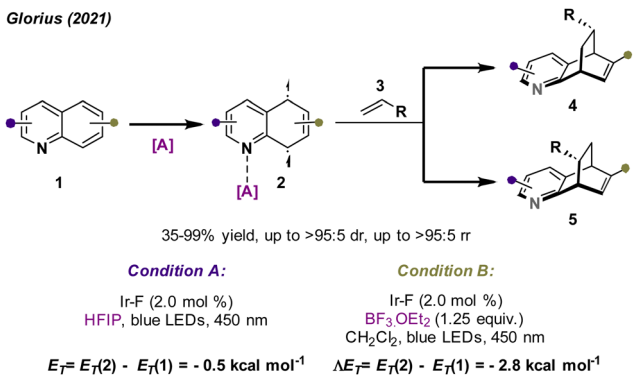
Fig. 7 Schematic representation of photocycloaddition *via* EnT.

possess a limited scope for exploring chemical space. On the contrary, intermolecular cycloaddition allow the expansion of substrate diversity on both units. However, the latter requires two sequential bimolecular reactions of excited states, posing a kinetic challenge to reactivity. A breakthrough example shown by Brown, Houk, and our group presents the EnT-facilitated intermolecular dearomative [4+2] cycloaddition of quinolines and alkenes with high regio- and *endo*-diastereoselectivity (Scheme 1).[44] This methodology transforms 2D units to 3D architectures, overcoming the ground state aromaticity (81.0 kcal mol^{-1}), which was otherwise inaccessible starting from feedstock materials in a single step. In this work, Lewis acids (HFIP and $BF_3{\cdot}OEt_2$, respectively) were applied as additives for substituted and unsubstituted quinolines, delivering conformationally rigid hetero-polycycles in good yields. Both HFIP and $BF_3{\cdot}OEt_2$ served as quinoline activators by lowering its triplet energy, where the former is used as a reaction solvent. Moreover, the authors reported a second role of the additives in lowering the transition state energy of the first radical addition step. Several subsequent works were reported exploiting a similar principle of exciting the core organic molecule followed by olefin trapping, and *vice versa*.[60–65] Parallelly, You and coworkers also reported a double dearomative intramolecular [4+2]-cycloaddition engaging an indole and a pendant naphthalene moiety.[66]

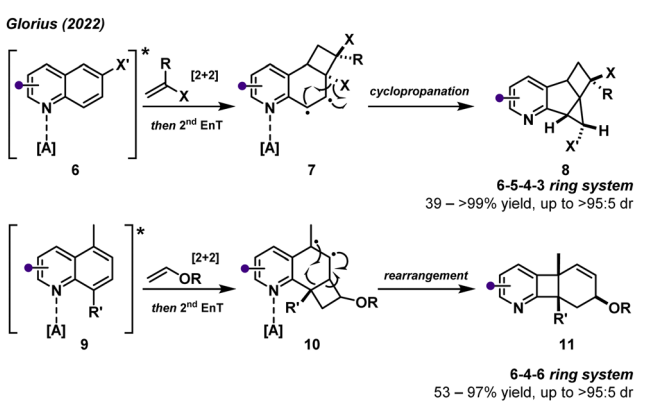
In all of these examples, *para*-cycloaddition products were formed, especially if olefins were employed as the ground-state reaction partner. This stems from the photolability of the *ortho*-cycloadduct bearing a strained four-membered ring next to the sensitizable styrene- or diene-type double bond.[62,67] Our group's recent study, in collaboration with the Houk group, on the cascade synthesis of fused 2D/3D rings *via* two consecutive energy transfer processes endorses the former statement (Scheme 2).[68] Here, chloro- and ester-substituted quinolines guide the second EnT-enabled cyclopropanation and vinylcyclobutane rearrangement, respectively, thereby enabling the construction of sp^3-carbon-rich ring systems. Another interesting study from our group exercises the triplet state activity of azaarenes in *peri*-(3+2) photocycloadditions with alkynes.[69]

Non-dearomatization reactions. Excited-state reactions of open-chain olefins tethered to a photo-excitable scaffold often result in attractive 3D molecules, thereby upgrading the saturated sp^3-rich chemical space to target pharmacologically relevant structural motifs. This can either take place *via* straight or crossed intramolecular [2+2]-cycloadditions producing cyclobutane building blocks and related bridged frameworks. Following this concept, a range of intriguing methodologies towards both types of product motifs have been developed. The first visible light EnT-mediated example was demonstrated by Yoon and coworkers, disclosing an intramolecular [2+2] photocycloaddition of styrenes[56] and dienes[70] bearing an additional tethered alkene unit. Similarly, on the intermolecular level, Xiao and coworkers displayed a [2+2] photodimerization of 3-ylideneoxindoles.[71]

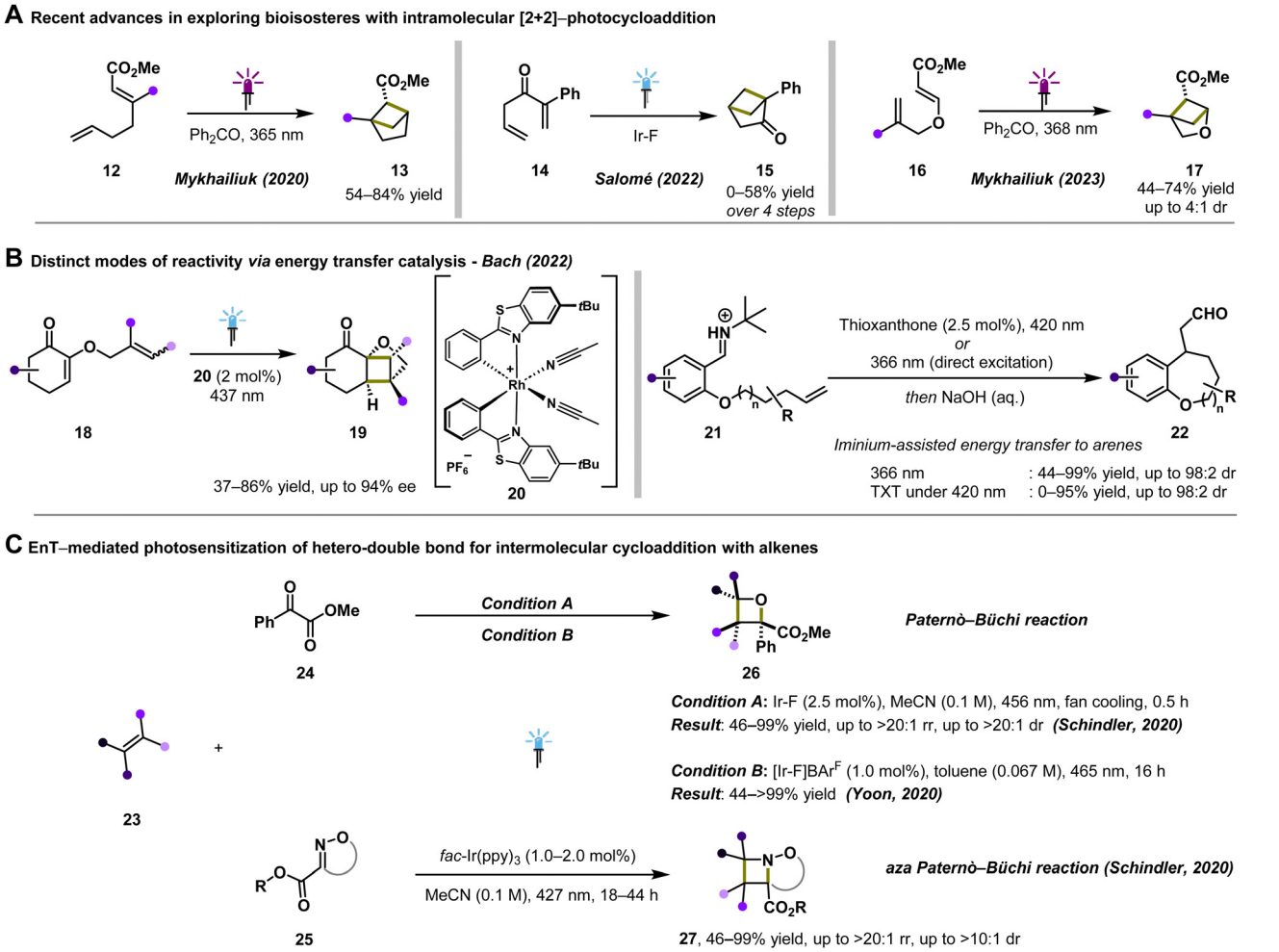
Following a similar intramolecular pattern, Mykhailiuk and coworkers developed a crossed [2+2][72] photocycloaddition of 1,5-dienes with a α,β-unsaturated esters **12** as the sensitizable unit (Scheme 3A).[73] The authors employed benzophenone as triplet photosensitizer, affording the crossed products in 54–82% yield. This generates bicyclo[2.1.1]hexanes **13** as bioisosteres of *ortho*-disubstituted benzenes. The presence of a two-carbon tether between two olefin units drives the selectivity towards crossed products, which would have otherwise generated the strained 4-membered ring in a straight product. Similar transformations have been developed recently, featuring an extra heteroatom in the motif, thereby allowing access to further attractive bioisostere candidates.[74–77] Inspired by the ability to enforce cross-selectivity, Bach and coworkers reported the first enantioselective crossed [2 + 2] photocycloaddition of 2-(alkenyloxy)cyclohex-2-enones **18** using chiral Rh-complex **20**



Scheme 1 Dearomative photocycloaddition of bicyclic azaarenes with olefins.



Scheme 2 EnT-enabled construction of diverse fused rings.



Scheme 3 (A) Targeting bioisosteres with intramolecular [2+2] photocycloaddition, (B) origin of crossed selectivity and iminium-assisted arene activation *via* energy transfer, and (C) EnT-mediated sensitization of benzoylformates and oximes.

giving rise to the unprecedented fused skeleton **19** in moderate to decent enantioselectivities (Scheme 3B).[78] The endocyclic olefin component prevents the undesired *E*/*Z* isomerization, which in turn, enables the generation of four stereocentres with exclusive selectivity. Another remarkable work by the same group utilized the formation of transient iminium ions to activate benzaldehydes, yielding benzoxacyclic moieties, tetracyclic units, and azetidines based on substitution patterns in the molecule. The authors majorly focussed on the formation of benzoxepanes **22** that are produced *via* arene-alkene photocycloaddition followed by multiple rearrangement steps.[79] A combination of the previous two concepts has also been reported by the same group.[80]

Although there are numerous exciting reports on the intermolecular version of EnT-mediated photocycloaddition, two case studies from Schindler's and Yoon's groups introduced diversity in the selection of photosensitizable substrates (Scheme 3C).

In 2020, they presented the EnT-assisted Paternò-Büchi[81,82] and aza-Paternò-Büchi[83] reactions with benzoylformates ($E_T \sim 60$ kcal mol^{-1}) and oximes ($E_T \sim 45$–72 kcal mol^{-1}) as substrates, respectively. This mild and easy operating method allows the synthesis of a variety of pharmaceutically relevant azetidines and oxetanes in synthetically useful yield and good selectivity. Further noteworthy advances of EnT-mediated intermolecular photocycloadditions include, for instance, the ability to use water as solvent,[84] substrate-substrate coordination,[85] and the use of photoenzymes.[86]

Merger of strain-release and cycloaddition

Expanding the reactivity of photocycloadditions from conventional [2π+2π]-cycloadditions to other type of acceptors enables access to a further 3D-enriched chemical space. To this end, the concepts of 'conformational restriction' and 'escape from flatland' in the context of biosiosterism have been actively investigated in the past few decades for optimizing pharmacokinetic activities. This concept holds a high value in drug discovery and structure-based molecular replacement of lead compounds. In this context, synthetic chemists have aimed for structural modifications in terms of size, flexibility, distance, or angles between substituents, eventually tuning the metabolic response, toxicity, and bioactivity. A study by Lovering indicates that the increase in saturation and the a number of chiral

 This journal is © The Royal Society of Chemistry 2024

centres directly influences the success of a drug candidate.[87] They formulated Fsp^3 as an index for saturation.

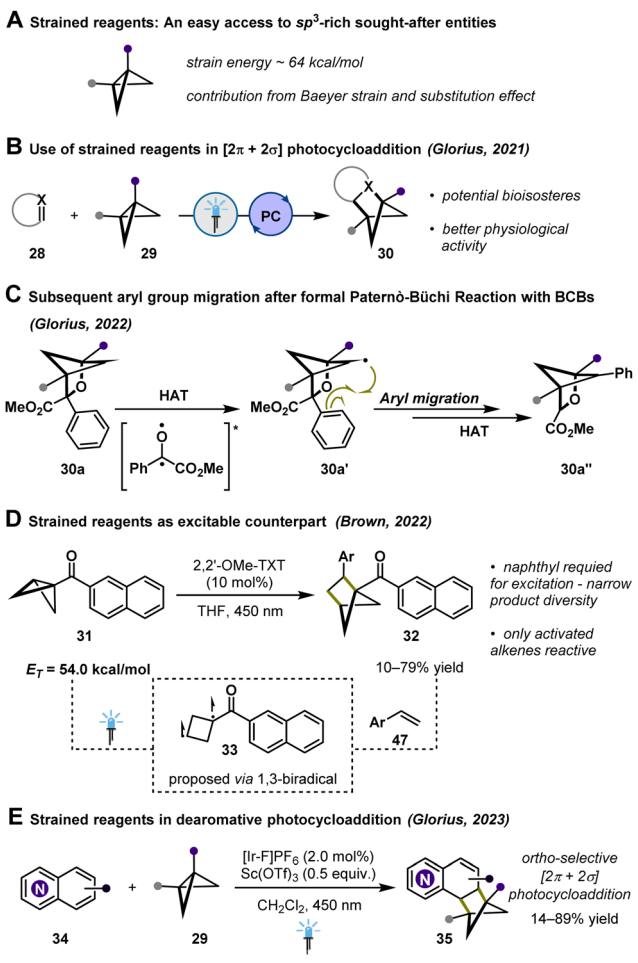
$$Fsp^3 = \frac{\text{number of } sp^3 \text{ hybridized carbons centres}}{\text{total carbon count}}$$

With this in mind, scientist around the globe have attempted to find bioisosteres for the basic *ortho*-, *meta*- and *para*-substituted benzene. Bicyclo[1.1.0]butanes (BCBs) **29** can be viewed as a suitable building block for installing a conformationally rigid sp^3-rich skeleton (Scheme 4A).[88,89]

With a relatively high strain energy (~64 kcal mol^{-1}), leveraging this thermodynamic driving force for cycloadditions has been a key aspect in this rapidly expanding field.[90–92] In this regard, our group discovered the first intermolecular [2π+2σ]-photocycloaddition mediated by energy transfer (Scheme 4B).[93] A series of commercially available endocyclic and aromatic double bonds, including coumarins, indoles, and flavones, were activated through energy transfer, readily undergoing cycloaddition with differently substituted bicyclo[1.1.0]butanes **29**. This single-step reaction offered bicyclo[2.1.1]hexanes (BCHs) **30** with high functional group density and excellent diastereoselectivity.

The scope also revealed the tolerance towards the formation of an aza-BCH moiety, a potential bioisostere of pyrrolidine. Next, intending to incorporate further hetero-atoms into the BCH motif, our group used benzoylformate esters as the biradical source, leading to the formation of an oxa-BCH framework.[94] Interestingly, the mechanism involves subsequent C–H abstraction and aryl group migration after the formal [2π+2σ]-photocycloaddition (Scheme 4C). This transferred an aryl group to the bridge position of the BCH, which had previously been inaccessible. Independently, Brown and coworkers disclosed the synthesis of BCHs **32** with reversed roles of the alkene and BCB units, where the strained moiety is taking part in the initial formation of the biradical species (Scheme 4D).[95] The authors constructed a BCB ketone unit **31** with a naphthyl group attached to it. Sensitization of the keto-naphthyl group enabled the formation of biradical **33** from the strained C–C bond, which reacted further with alkenes. The authors claimed the possibility of direct excitation of **31** to the ketone's (n → π*) excited state followed by ring opening to the biradical intermediate **33**.

Inspired by the ability to use (aza)arenes as the sensitized component in [4+2]-photocycloadditions,[44] our group, in collaboration with the Houk lab, set on developing a similar dearomative cycloaddition process with BCB partners. Intriguingly, we discovered the *ortho*-selective [2π+2σ]-photocycloaddition of BCBs **29** and azaarenes **34** (Scheme 4E).[40] In contrast to our previous report,[68] this reaction terminates at the *ortho* product rather than undergoing a downstream sensitization-rearrangement cascade. Computational analysis revealed a highly complex mechanistic picture, initiated by Lewis acid activation of the azaarene, which lowers both the triplet energy (62.6 kcal mol^{-1} → 52.1 kcal mol^{-1}) and the activation barrier of the rate-limiting addition into the BCB; and points towards the involvement of single-electron transfer (SET) steps. Contrary to our earlier report, the computations suggest that the [2+2]-cycloadduct is thermodynamically favoured over the [4+2] (−53.9 kcal mol^{-1} *vs.* −63.9 kcal mol^{-1}).



Scheme 4 (A) Bicyclo[1.1.0]butanes as building blocks, (B) [2π+2σ]-photocycloaddition with double bonds, (C) mechanistic proposal for aryl shift in oxa-BCH units, (D) EnT-induced sensitization of BCBs, and (E) conservation of *ortho*-selectivity in [2+2]-photocycloaddition.

Homolytic σ-bond cleavage

Organic molecules containing (photo-)labile bonds have often represented versatile reagents to synthetic chemists since they offer simple operational techniques to access reactive, synthetically usable fragments. There are mainly three known methods of executing such bond cleavage, (1) transition-metal catalysis through oxidative addition, (2) single electron reduction, and (3) single bond homolysis. Whereas single-electron transfer approaches usually require an external sacrificial additive, and often lead to the "loss" of one fragment, this poor atom economy can be circumvented by homolytic bond cleavage (Scheme 5A).

However, thermal cleavage or direct excitation of substrates are usually accompanied by harsh reaction conditions, unsafe environment, and poor functional group tolerance. As an alternative, energy transfer catalysis can proceed in an atom-economic and sustainable fashion where bifunctional reagents[96] serve as the potential carrier of reactive intermediates. After energy

transfer, these reagents cleave into two distinct radicals, one transient and highly reactive, another one with a persistent nature.

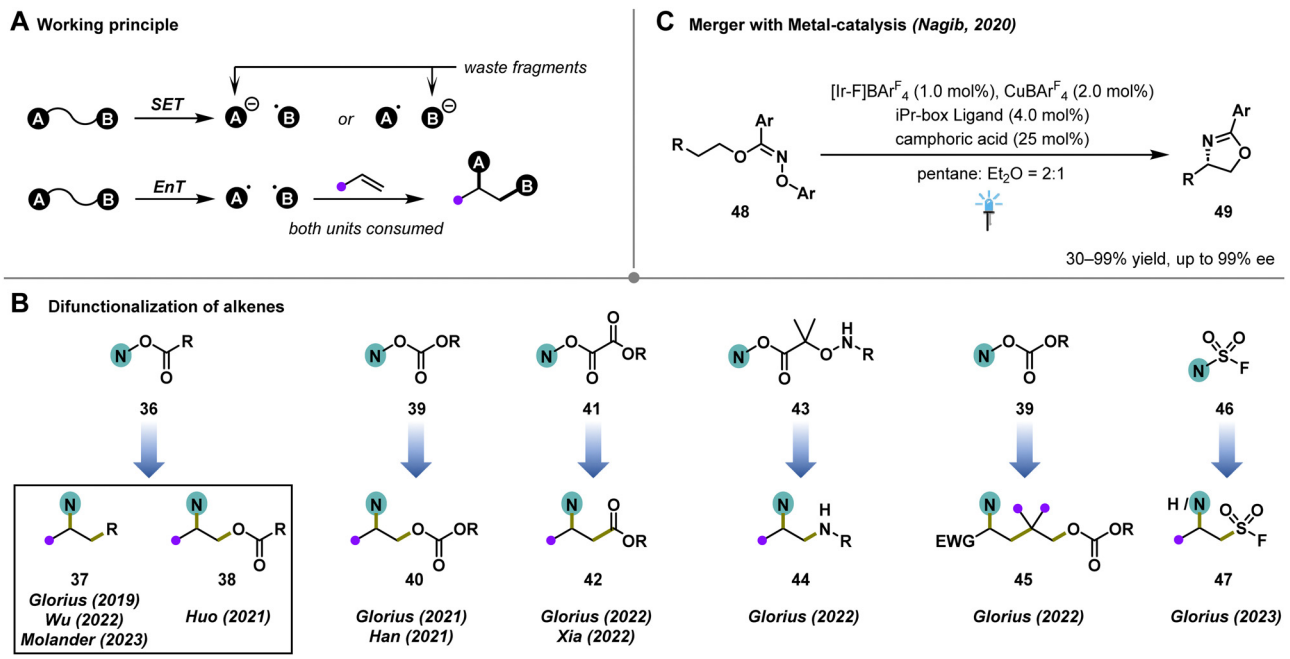
In this vein, many researchers implemented EnT-catalysis for the direct difunctionalization of cheap and abundant alkene feedstocks to generate diverse target molecules (Scheme 5B), usually in a racemic fashion. These transformations require the careful design of bifunctional reagents (Fig. 8), enabling an EnT-mediated homolytic bond cleavage events to give rise to the desired radical species while suppressing reagent decomposition[97] through *e.g.* single electron transfer or ionic pathways.

There are a few key notes about bifunctional reagents: (a) the design of bifunctional reagents is often guided by the differences in the lifetime of the generated radical fragments to ensure high regioselectivity, (b) there is no general and conclusive trend between the stability of the radical and its lifetime,[100] (c) a different self-reaction rate of two radicals is necessary to ensure cross-selectivity, and (d) the generation of radical species can be favoured by an entropic driving force, incorporating eliminable CO_2, SO_2 or acetone units in the core structure. Basic reactivity principles dictate the reactive (transient) radical to add onto the alkene first, followed by recombination with the long-lived radical (persistent), resulting in high regio- and cross-selectivity.[100] In 2019, our group reported reagent **36** for the EnT-mediated iminoalkylation of unactivated alkenes. Upon energy transfer, this reagent undergoes a sequence of homolytic N–O bond cleavage and decarboxylation, to form an iminyl (persistent) and alkyl (transient) radical.[101] Many other groups also adopted a similar strategy for carboimination of alkenes.[102,103] The same class of reagent was used by Huo group for oxyimination purposes, where they introduced an aryl moiety (R = Ar) to suppress the decarboxylation and generate a transient O-centred radical.[104] Later on, a simple manipulation of one functional entity in the bifunctional source **39** offered a quick route to diverse 1,2-amino alcohols.[105,106] The same class of reagents can also be used for hydro-oxygenation of unactivated alkenes by modifying both the imine moiety to form a less persistent radical and the solvent to $CHCl_3$, which suppressed radical–radical recombination in favour of a hydrogen-atom abstraction step.[106]

The simultaneous installation of a carbonyl (transient radical) and an iminyl unit (persistent radical), formally accessing β-amino acids and β-amino ketones was independently reported by Yang and our group, respectively.[107,108] In both works, the installation of a dicarbonyl unit in radical source **41** led to the extrusion of CO_2 and the generation of a transient acyl radical. In addition to unactivated alkenes, our group also demonstrated the dearomative aminocarboxylation of indoles, benzothiophenes, benzofurans, and phenanthrenes using the same reagent. One step further, our group accomplished the diamination of alkenes and aromatic cores with reagent **43**, generating a transient N-centered radical after the simultaneous elimination of CO_2 and acetone.[98]

Here, we succeeded in installing two distinct N-centred radicals onto the alkenes with perfect regioselectivity, forming 1,2-diaminated products **44**. The use of two electronically distinct olefins (an unactivated olefin and a Michael acceptor) offered a highly selective new class of 1,4-difunctionalized product motifs in a three-component coupling.[109,110]

So far, all the examples deal with N–O bond cleavage, which poses a restriction on the generality of bifunctional reagent design. Against this background, our group recently unveiled the first case of homolytic N–S bond scission for difunctionalization of alkenes



Scheme 5 (A) Outcome of σ-bond cleavage under different conditions, (B) recent reports on alkene functionalizations, and (C) amalgamation of metal catalysis with σ-bond homolysis.

 This journal is © The Royal Society of Chemistry 2024

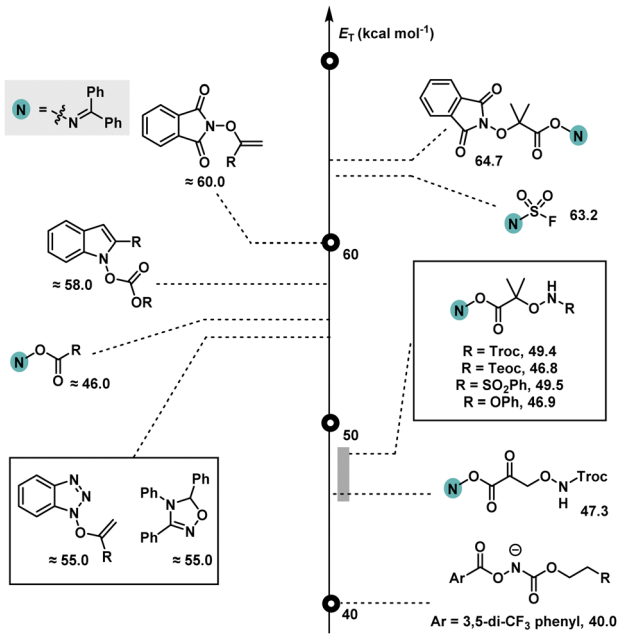


Fig. 8 Triplet energies of organic compounds for σ-bond cleavage.[97–99]

using a bench-stable sulfamoyl fluoride reagent **46**.[99] This work enabled the introduction of the SO_2F moiety, which opens up a variety of down-stream functionalization reactions, owing to its ability to take part in late-stage click chemistry (sulfur(VI) fluoride exchange-SuFEx).[111] As such, this reagent was demonstrated to enable the synthesis of β-ammonium sulfonates, β-sultams, and 1,3-sultones apart from β-amino sulfonyl fluorides. In addition to that, a HAT (hydrogen atom transfer) reagent was also employed to offer aliphatic sulfonyl fluorides. The principle of difunctionalization *via* σ-bond cleavage was also recently extended to strained systems.[112]

Although racemic transformations of alkenes present an easy entry to a diverse chemical space, the asymmetric version would further enable the installation of well-defined stereocentres in a single synthetic operation using cheap and abundant feedstock. However, the design of a special catalyst, the selection of a suitable ligand, and its multistep synthesis pose a tremendous synthetic challenge. Along this line, Nagib and coworkers demonstrated the merger of EnT-mediated N–O bond cleavage, hydrogen atom transfer (HAT), and copper catalysis (Scheme 5C).[113] This dual-catalytic protocol enables a robust enantio- and regioselective β-C–H amination of alcohol-derived bifunctional reagent **48**, affording enantioenriched ozaxoline moieties **49**. Notably, the authors demonstrate the necessity of a BAr^F_4 anion and a bulky acid additive for high degrees of stereo-control and stabilization of the Cu intermediates, respectively. The same group also targeted similar dual catalytic approach for alcohol functionalization.[114,115]

Stereochemical editing

The stereochemical nature of molecules directly translates to their properties and function, offering considerable opportunities if molecules of choice can be constructed in a stereocontrolled fashion.

In an ideal synthetic scenario, bond formation and stereochemical control are amalgamated in a single step, however the chemical toolbox may be limited in certain contexts. In those cases, decoupling of bond forming and stereochemistry-determining steps may help to achieve the equivalent synthetic quest. Following this logic, properties of a parent starting material may be tuned by diversification to multiple isomers, rather than introduction or interconversion of functional groups.

***E*/*Z* Isomerization.** In a sp^2-rich chemical space, control over double bond geometry (for triplet energies, see Fig. 9) is a central target. While the thermodynamically downhill process ($Z \rightarrow E$) toward the more stable isomer can be achieved under thermal conditions (*e.g.* with Brønsted acids[116] or electrophilic halogens[117]), the inverse thermodynamically uphill process ($E \rightarrow Z$) represents a considerable challenge. Inspired by nature's isomerization processes,[29] coupling of the isomerization to an energetically downhill process may compensate for the endergonic step. Harvesting visible light energy in energy transfer catalysis, high-energy triplet species may be obtained, which have forfeited their fixed geometry, allowing for geometric relaxation by rotation around the central σ-bond (Scheme 6A). Since relaxation to the *E*- or *Z*-isomer occurs statistically, selectivity can be controlled through differences in the sensitization rates between *E*- and *Z*-isomers. In most conjugated olefins, steric repulsion in the *Z*-isomer induces a distortion from planarity, which leads to reduced conjugation and, in turn, higher triplet energies. The choice of an appropriate sensitizer that can differentiate between the two isomers can lead to a photostationary state in which the *Z*-isomer accumulates (Scheme 6B).

Alternatively, attractive secondary interactions in the *syn*-conformer can lead to the selective formation of the *Z*-isomer

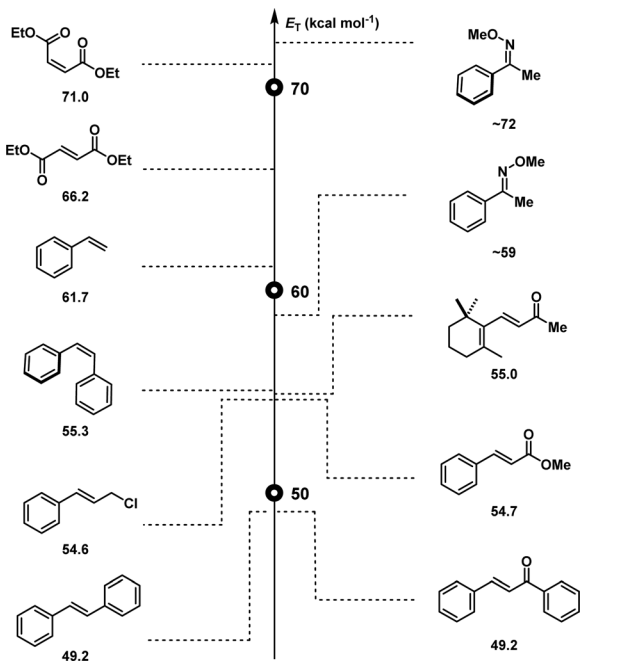
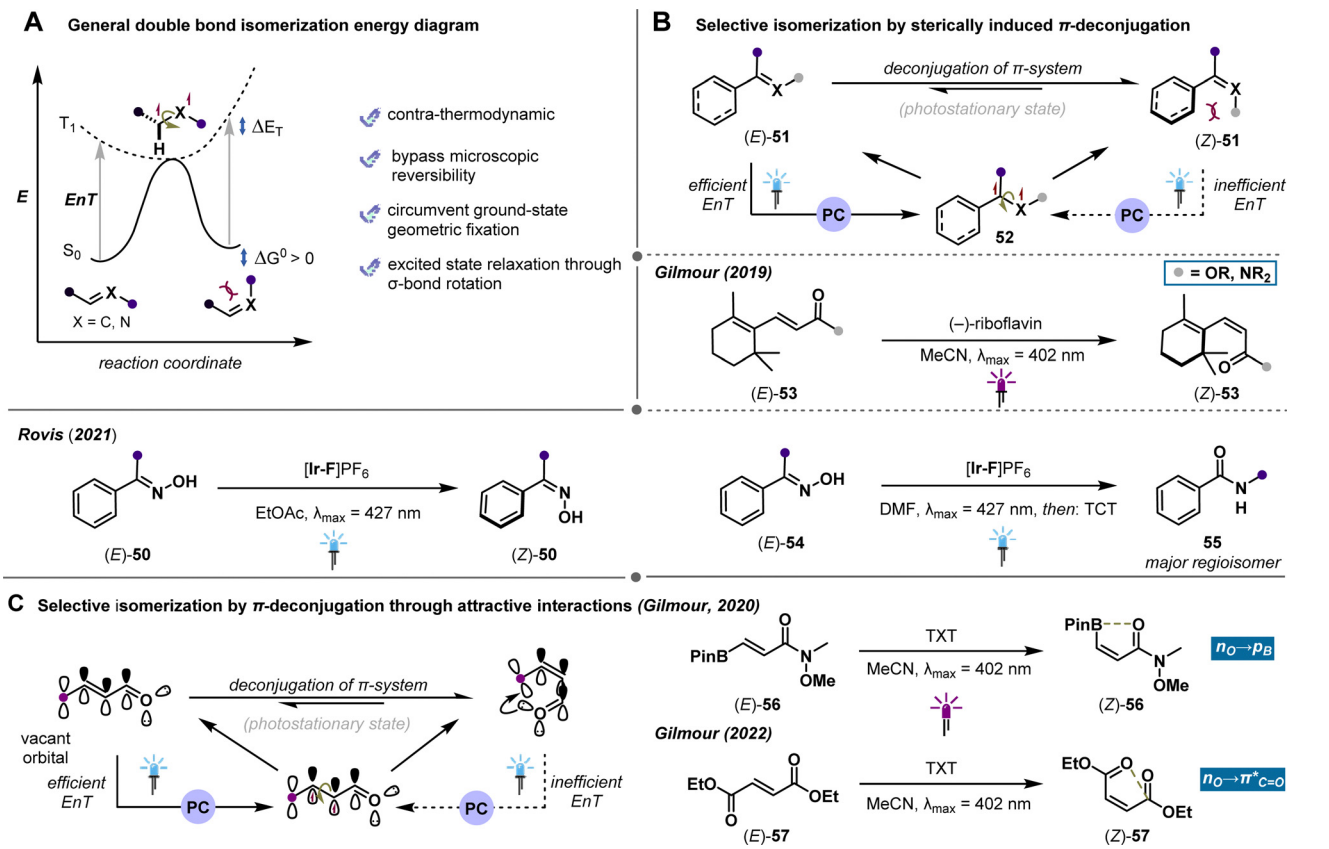


Fig. 9 Assorted triplet energies for photoisomerization substrates.[54,118,119]

(Scheme 6C). Following a rich history of UV light-mediated isomerization processes of styrenes and stilbenes,[29,120] visible light-driven $E \rightarrow Z$ isomerization of cinammyl derivatives[121–123] has shaped the field, highlighting the importance of 1,3-allylic strain to favourably shift the photostationary state by repulsive effects.[122,123] More recently, the above mentioned deconjugation strategy has been showcased for a selected class of dienes, namely β-ionyl derivatives,[124] as well as oximes[125,126] (Scheme 6B). Inspired by the photochemical isomerization of acetophenone-based oxime ethers reported by Padwa and Albrecht,[125] Rovis and coworkers utilized EnT catalysis for the $E \rightarrow Z$ isomerization of aromatic oximes and oxime ethers, which was further expanded to a regioselective one-pot photoisomerization/Beckmann rearrangement sequence.[126] Utilizing attractive interactions to modify the degree of deplanarization and deconjugation in small and highly functionalized alkene hubs, Gilmour and coworkers took advantage of the $n_O \rightarrow p_B$ interaction of β-boryl acrylates upon EnT-mediated $E \rightarrow Z$ isomerization.[127,128] Along the same lines, Gilmour and coworkers utilized $n_O \rightarrow \pi^*_{CO}$ interaction in the photoisomerization of fumarate derivatives.[118]

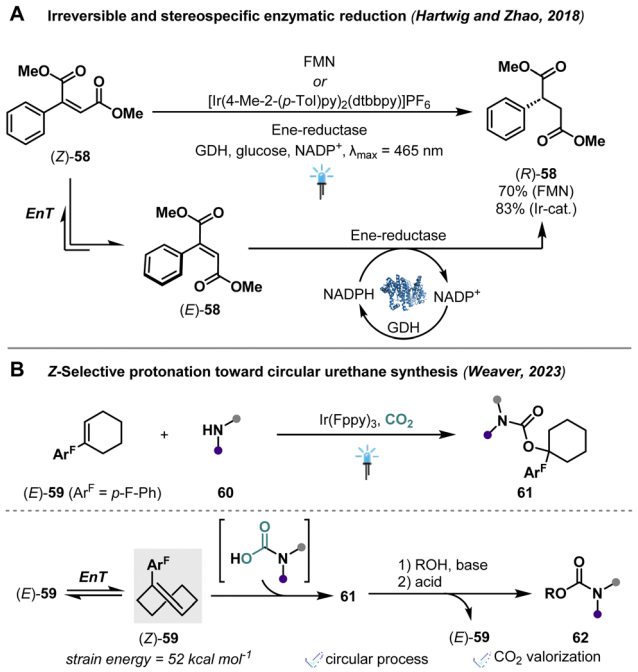
Rather than only shifting the photostationary state in the favour of one isomer, further interruption of the photostationary state by a sequential irreversible and stereospecific reaction may shift the equilibrium in accordance with Le Chatelier's principle. Most prominently, *in situ* ring closure of cinnamyl derivatives and related compounds upon $E \rightarrow Z$ isomerization falls into this category.[29] Translating 2D to 3D stereoinformation, Zhao, Hartwig and coworkers coupled the reversible, EnT-mediated isomerization of readily available *Z*-phenylsuccinate to *E*-phenylsuccinate with the selective reduction of the latter with an NADPH-driven enereductase, achieving dual EnT/enzyme catalytic DYKAT (Scheme 7A).[129] Taking advantage of the high strain energy of *Z*-cyclohexenes,[130] Weaver and coworkers recently showcased their facile protonation/nucleophilic trapping to furnish urethanes, which could be further extended to a circular CO_2 valorization by alcohol transposition and (*E*)-**59** recovery (Scheme 7B).[131]

Deracemization. In 3D-chemical space, enrichment of one enantiomer from a racemic mixture (for triplet energies, see Fig. 10) coincides with contra-entropic behaviour. With thermal deracemization strategies being either capped in yield (kinetic resolution) or substrate scope (requiring a readily epimerizing stereocenter for DYKAT), energy transfer catalysis may offer an alternative strategy for deracemization of specific substrate classes in high yields.[132] While the photonic energy may provide a thermodynamic driving force and compensate for the entropic loss throughout the progressing deracemization, the increased rotational degrees of freedom (*e.g.* around double bonds) in an excited triplet state facilitates a dynamic isomerization upon ground-state relaxation.



Scheme 6 (A) general energetic profile of EnT-mediated 2π-isomerization, (B) repulsive π-deconjugation, and (C) π-deconjugation by secondary interactions.

This journal is © The Royal Society of Chemistry 2024



Scheme 7 (A) Merging photoisomerization with enzyme catalysis, and (B) strain-driven protonation of aromatic (*Z*)-cyclohexenes upon photoisomerization. FMN = flavin mononucleotide, GDH = glucose dehydrogenase.

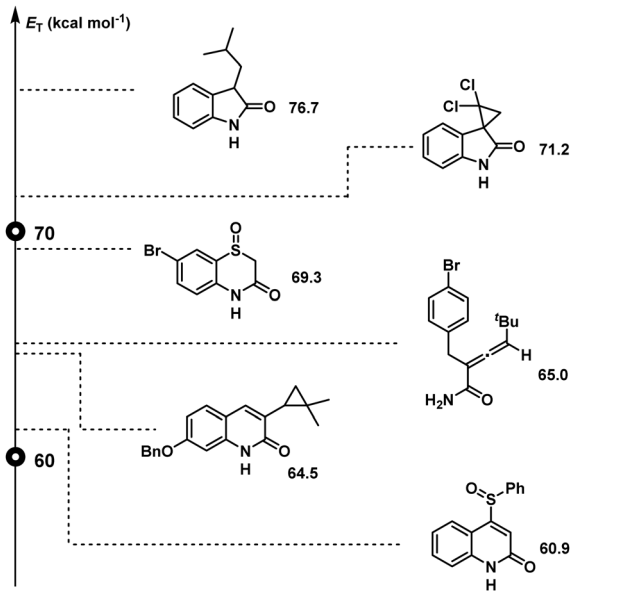


Fig. 10 Selected triplet energies of substrates for deracemization.[133–137]

The discrimination of the two isomers and therefore the selective shifting of the photostationary state is the main challenge to be overcome.[15] As the energy transfer event is highly distance-dependent, photosensitizers with tight substrate binding that may differentiate between subtle deviations in binding affinity and sensitizing activity of both isomers, successfully achieved high *ee*'s. Most notably, Bach and coworkers developed a series of photosensitizers bearing a rigid C(sp^3)-backbone featuring a hydrogen-bonding motif, which may readily position an amide substrate "in parallel" to the photosensitizer (Scheme 8). With two chiral photosensitizers available (**63** and *ent*-**63**), both enantiomers could be enriched starting from a racemate. Related systems with tethered benzophenone photocatalysts were successfully employed in HAT-mediated deracemization[138] of substrates with higher triplet energies.[137]

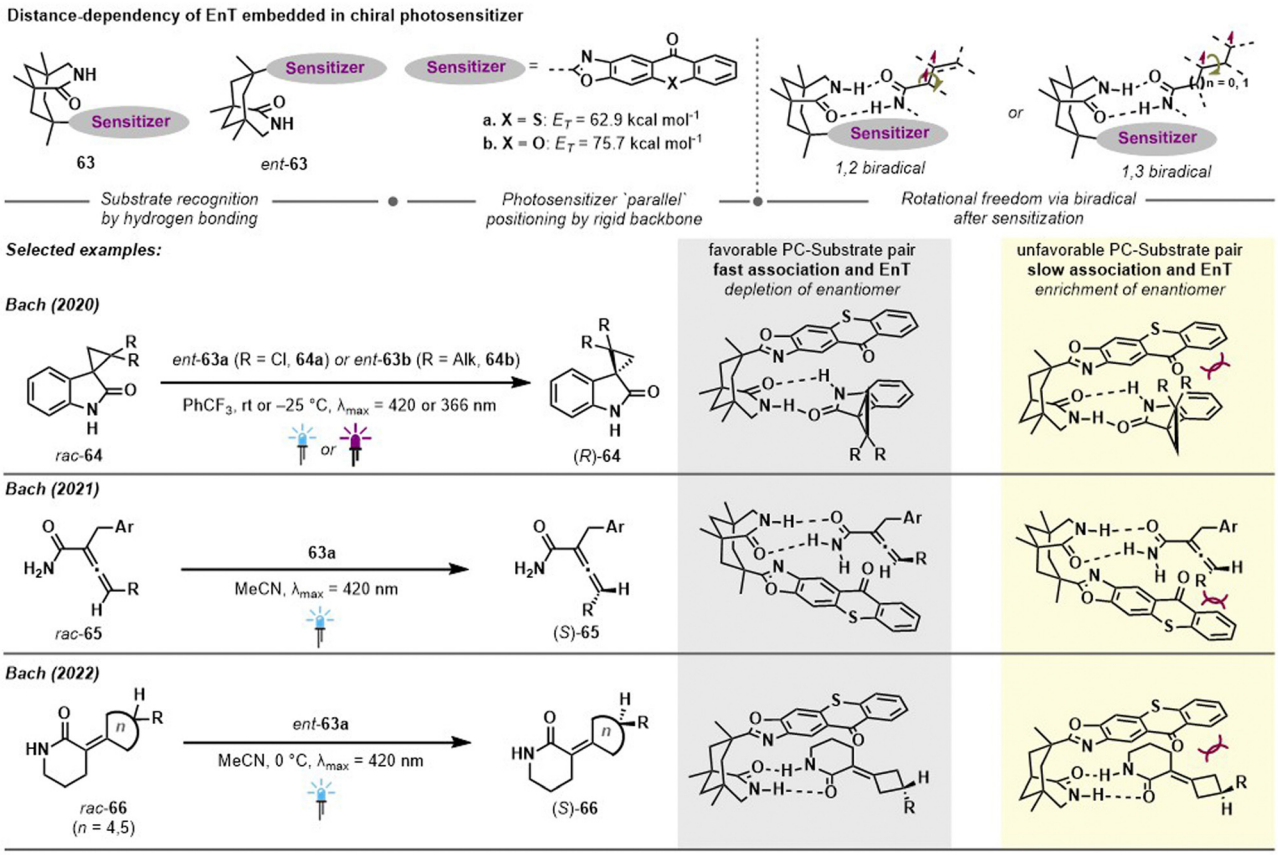
Upon EnT, 1,2- or 1,3-biradical species are formed from allenes or vinyl cyclopropanes, respectively, comprising the freely rotating deracemization intermediate. While other substrate classes, such as sulfoxides, were also shown to undergo deracemizations *via* different intermediates, enantioselectivities remained limited for these examples.[136] Above-mentioned catalysts were successfully employed in the deracemization of axially chiral molecules (*e.g.* cyclic[139] or acyclic[133] allenes and alkylidenecycloalkanes[140]), as well as centrally chiral 3-cyclopropylquinolin-2-one derivatives[134] and spirocyclopropyl oxindoles.[135]

In case of the deracemization of cyclic allenes,[139] the authors proposed the efficient deracemization to be facilitated by the interplay of differences in ground-state association constants on the one hand, and distance-dependent energy transfer rates in excited-state PC-substrate adducts on the other hand (Scheme 8). The latter effect was further supported by DFT studies in the deracemization of spirocyclopropyl oxindoles.[135] In this work, the importance of triplet lifetime was further highlighted, as the substrate triplet lifetime exceeded the lifetime of the hydrogen-bonding complex, favoring a sterically unbiased radical–radical recombination upon PC·substrate adduct dissociation. Finally, configurational stability of **64a** was observed with thioxanthone at low temperatures, which is in line with a thermal contribution due to the triplet energy gap between photosensitizer and **64a**. Besides steric repulsion, the importance of dispersion interactions for enantiomer binding differentiation was highlighted in the deracemization of open-chain allenes.[133]

Recently, Luo and coworkers reported the EnT-mediated deracemization of α-stereogenic aldehydes and ketones using chiral amine additive (*S*)-**68**.[141] Mechanistically, the enamine isomers, formed upon condensation, may be interconverted through a biradical intermediate in the triplet state, which relaxes by rotation around the central σ-bond to give a statistical *E*/*Z* ratio upon ground state relaxation. Selective shifting of the photostationary state favoring $E \rightarrow Z$ isomerization is achieved by differentiation in vertical triplet energies due to π-deconjugation by sterical repulsion (*vide supra*) of aromatic (*Z*)-enamines. Due to the chiral information of the attached ammonium salt, protonation from one face is favored, leading to the overall enrichment of (*R*)-**67**. Notably, the translation of the *E*/*Z* configuration into (*R*)- or (*S*)-enantiomers upon photoisomerization has been widely explored (Scheme 9).[129,142,143]

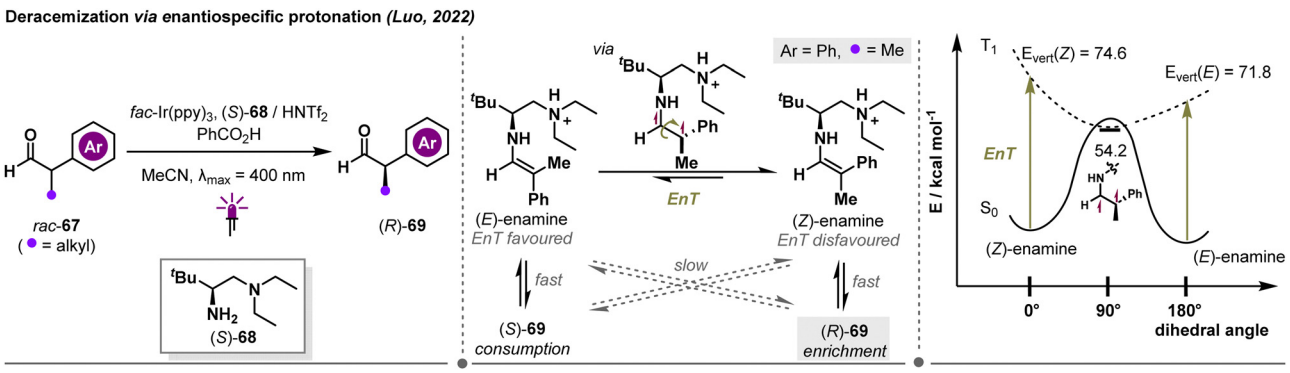
Sensitization of first-row transition metal complexes

Transition metal (TM) catalysis is experiencing a paradigm shift from traditionally employed precious transition metals to first-row



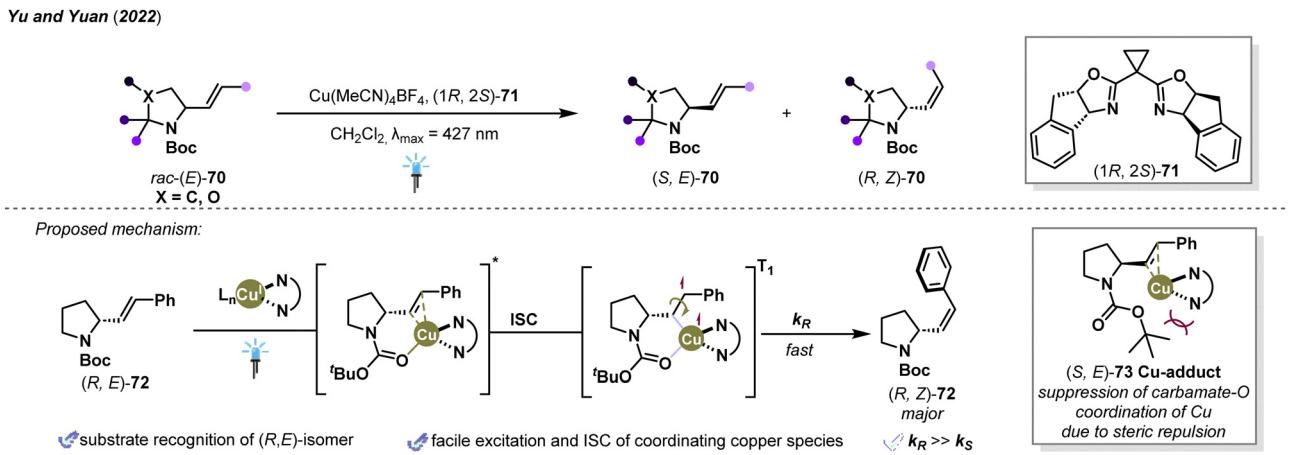
Scheme 8 EnT-mediated deracemization with chiral amide tethered photosensitizers. *ent* = enantiomer; *rac* = racemic.

transition metals, which may offer cheaper and more abundant alternatives to mimic precious transition metals in certain catalytic processes, but also exhibit new catalytic modes. Meeting both demands, early examples of directly excited first-row transition metals provided the proof-of-concept that excited-state reactivity may enable unconventional open-shell operations by the introduction of light energy as an additional driving force compared to the ground state.[144] On a molecular level, excitation may induce a charge transfer event between the metal center and the ligand (or *vice versa*), leading to the population of antibonding orbitals.[13] For instance, recent developments in utilizing ligated copper species in $E \rightarrow Z$ isomerization upon direct photoexcitation[145–147] enabled Yu, Yuan and coworkers to employ a chiral copper catalyst for the kinetic resolution of 2-styrylpyrolidine derivatives.[148] The authors proposed that due to the chiral ligand backbone, the copper catalyst could discriminate between the enantiomers, achieving more rigid substrate binding with (*R*,*E*)-**72** due to interaction with the carbamate-oxygen. DFT studies confirmed a lower isomerization barrier for (*R*,*E*)-**72** compared to (*S*,*E*)-**72**, leading to the observed enantioselectivity (Scheme 10). Based on a similar mode of Cu coordination, kinetic resolution of 2-cinnamylpyrrolines was achieved.[149]



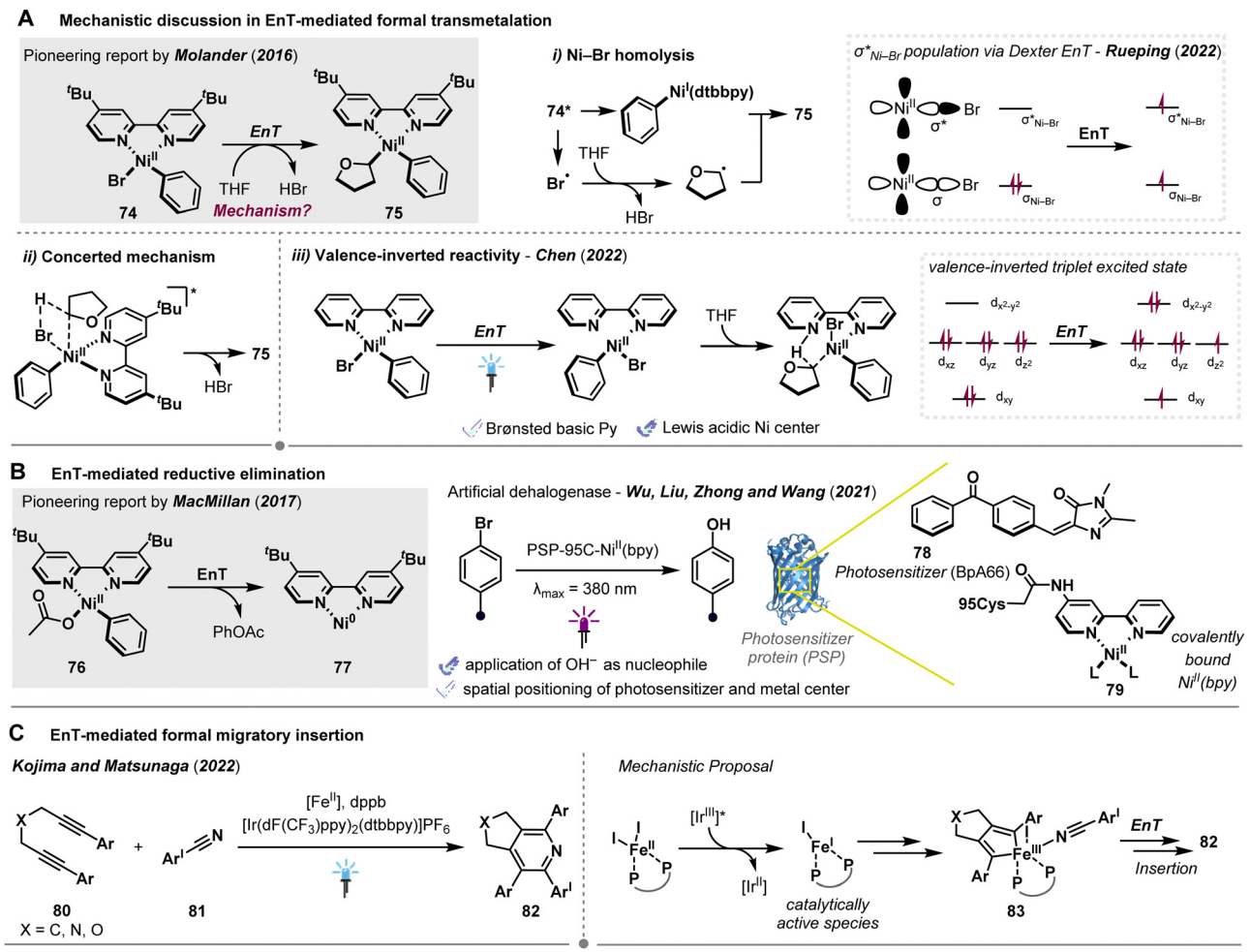
Scheme 9 Deracemization through dynamic $E \rightarrow Z$ isomerization.

 This journal is © The Royal Society of Chemistry 2024

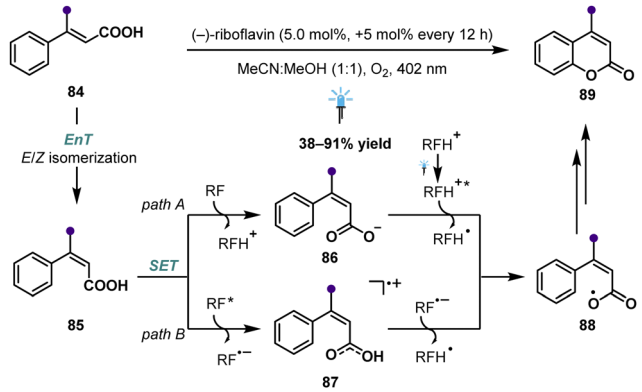


Scheme 10 Sensitized Cu-mediated kinetic resolution through $E \rightarrow Z$ isomerization.

In contrast to direct excitation of a transition-metal catalyst, introduction of a photosensitizer may offer ligand-independent (or ligand-free)[150] and overall more general reaction setups, as longer excitation wavelengths and potentially non-absorbing complexes become accessible.[12] Pioneering work by Molander and coworkers[151] showed the EnT activation of Ni^{II} oxidative addition adducts bearing bipyridine ligands for the formal transmetalation with unfunctionalized coupling partners



Scheme 11 (A) Mechanistic picture for the formal transmetalation of excited state nickel complexes, (B) reductive elimination of excited state nickel complexes, and (C) formal migratory insertion of excited state ferracyclic intermediates.



Scheme 12 Cascade photocatalysis for the conversion of cinnamic acids to coumarins.

bearing α-heteroatoms or benzylic C–H bonds[152] are prone to HAT, the latter being showcased by Rueping and coworkers (Scheme 11A). Notably, benzophenone derivatives as additives or photosensitizers may contribute to the HAT step.[151,153]

Mechanistically, the initially proposed homolysis of the Ni–Br bond through population of antibonding $d_{x^2-y^2}$ orbital in the excited state, followed by HAT by the bromine radical, has been most widely discussed in the literature. Recent in-depth computational investigations by Chen and coworkers[154] point towards a different mode of action. For the triplet nickel species, ''valence-inverted reactivity'' ($d_{xy}^1 d_{xz}^1 d_{x^2-y^2}^2$) with a metal–ligand antibonding $d_{x^2-y^2}$ orbital was proposed, leading to a three-coordinate nickel(II) complex. The authors concluded that the exposed Brønsted basic pyridine motif and Lewis-acidic metal center may rather contribute to the HAT/ligand exchange event than the previously proposed bromine atom generation[155] or concerted mechanism.[151]

MacMillan and coworkers exploited sensitized nickel(II) species for unconventional C–O reductive elimination with carboxylates (Scheme 11B).[156] While an electron transfer pathway could not be excluded, photocatalyst properties and slow reductive elimination in the presence of light without photocatalyst point towards an EnT-mediated mode of action, which was further supported by transient absorption spectroscopy studies.[157] Amongst various advances of achieving C–O,[158] C–N[159–162] and C–S[163] reductive elimination while gaining valuable mechanistic insights into the complex catalytic system, the recent development of an artificial dehalogenase by Wu, Liu, Zhong, Wang and coworkers[164] stands out in highlighting basic principles and the synthetic potential of this strategy. Using a photosensitizing benzophenone-imidazolinone unit **78** and covalently attached bipyridine-bound nickel catalyst **79** in a well-defined proximity, water could be employed as the reductive elimination partner.

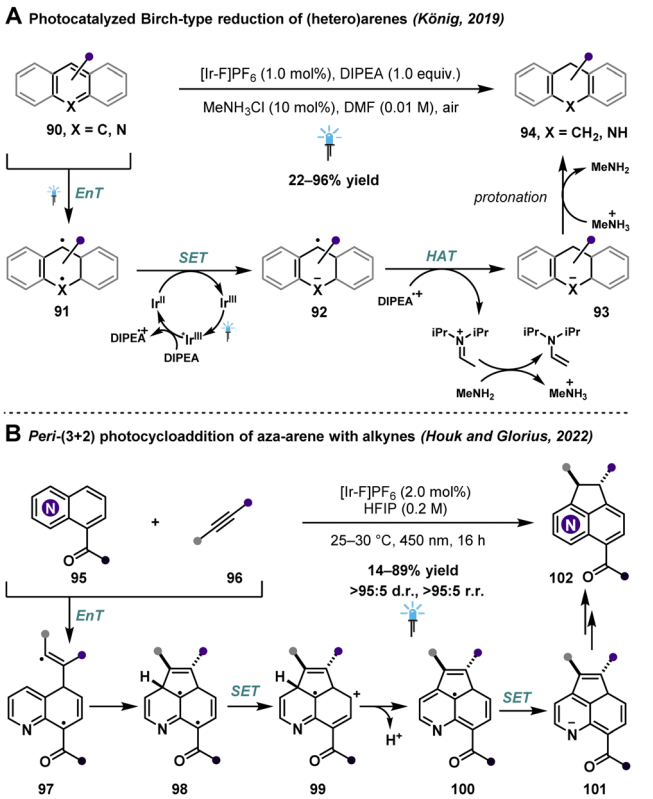
Recently Kojima, Matsunaga and coworkers employed a photosensitized iron-complex for the formal migratory insertion of aromatic nitriles into the Fe–C bond, which is difficult to achieve under thermal conditions (Scheme 11C).[165] Utilizing tethered bis-alkynes **80**, densely substituted pyridine derivatives **82** (and azafluoranthenes) were obtained. Cyclic voltammetry studies suggest that initial SET between excited photocatalyst and Fe^{II} would give catalytically active Fe^{I} species, which may further be converted to ferracyclic intermediate **83**. Rate-limiting insertion into the nitrile through photosensitization may then result in product formation while regenerating the catalytic Fe^{I} species.

Merger of EnT and ET in TTEnT photocatalysis

As discussed in the introductory section, the photocatalysts employed for TTEnT catalysis are inherently active in excited-state single-electron transfer processes too. Whereas these pathways constitute undesired side reactions in most cases, harnessing this reactivity for designing dual EnT-photoredox catalytic processes, mediated by the same photocatalyst, can be highly attractive.

As an early example, in 2016, Gilmour and coworkers leveraged the multiple activation mode of (–)-riboflavin (RF) for the direct conversion of (*E*)-cinnamic acids **84** to coumarins **89** (Scheme 12).[166] The process commences with EnT-catalyzed *E*/*Z*-isomerization to form the (*Z*)-isomer **85**. Path *A* uses RF to deprotonate **85** followed by SET to form radical intermediate **88**. Path *B* uses the excited RF* for SET followed by deprotonation to form the same intermediate **88**. The latter undergoes cyclization with adjacent arene followed by HAT step to form the desired coumarin **89**.

In 2019, König and coworkers explored the dual role of Ir–F for the Birch-type photoreduction of (hetero)arenes



Scheme 13 (A) Two photons-assisted dearomatization of arenes, and (B) dearomatization/rearomatization sequence for *peri*-(3+2) cycloaddition.

This journal is © The Royal Society of Chemistry 2024

(Scheme 13A).[167] After the EnT step, the triplet state of the arene **91** was reduced by Ir^{II} (DIPEA was used as a sacrificial electron donor to quench the excited $*Ir^{III}$) to form intermediate **92**. This is followed by a fast HAT step and protonation to yield the dihydroarene **94**. This proposed two-photon process was corroborated by the quadratic dependence of the product yield on the irradiation intensity. Recently, Houk and our group demonstrated the *peri*-(3+2) photocycloaddition of bicyclic aza-arenes **95** with alkynes **96** to generate otherwise challenging—aza-acenaphthenes **102** (Scheme 13B).[69] This methodology utilizes both terminal and internal alkynes with excellent regio- and diastereoselectivity. The series of dearomatization/rearomatization steps were supported by DFT calculations. Similar studies exercising both the concepts of photocatalysis in a single system have been reported, displaying dual roles of pyrylium salt,[168] $Ir(ppy)_3$,[169] $[Ru(bpy)_3]Cl_2$,[170] acridinium,[171] and riboflavin.[166]

3. Conclusions, perspective, and directions

The rapid growth of EnT catalysis in organic synthesis over the last few years emphasizes the importance of mild, visible-light-mediated reactions for sustainable, environmentally friendly synthesis. Against this background, the field has witnessed commendable conceptual advances in exploiting new modes of triplet excited state reactivity, leading to a versatile range of transformations and synthetically accessible molecular scaffolds. This review does not only attempt to summarize the main progress over the last few years, but, importantly, introduces the fundamentals of energy transfer catalysis, and provides guidelines for effective catalyst and reaction design. Moreover, a selection of substrate classes along with their reported triplet energies has been compiled, to provide the reader with a handle to identify sensitizable motifs and assess their reactivity under energy transfer catalysis.

Noteworthy, despite these advances, the field of EnT catalysis has faced two major operational and synthetic challenges that are shared across the comparatively young discipline of (visible-light-mediated) synthetic chemistry: (a) enabling high irradiation efficiency in solution prevents a straightforward adoption of larger-scale synthesis. (b) The lack of equipment and reporting standards when it comes to the specifications of the light source and photoreactor can hamper reproducibility. While we see a high value in condition-based[172] screening approaches to enable robust reaction reproduction, we foresee major advancements in reactor design and engineering in the next years, motivated by the increasing synthetic potential of photocatalysis.

From a synthetic perspective, the design of suitable high-end photosensitizers has been an underexplored topic for a long time. A recent report from our group identified NHC-bearing *fac*-$[Ir(CF_3\text{-pmb})_3]$ as a photosensitizer with a high triplet energy of 73.3 kcal mol^{-1},[173] addressing the issue of relying on weakly absorbing ketone-based photocatalysts to access high-lying triplet states. Similar structural modifications of existing photocatalysts – for instance the organic chiral photosensitizers[174] – could turn out to be an asset in this field. Recently, the Dexter-type energy transfer also found its application on proximity labelling platforms and delineating protein–protein interactions in a cellular environment.[175] For this to happen, the authors modified Ir–F into a physiologically suitable photosensitizer having a polyethylene glycol, an alkyne, and a carboxylic acid groups attached to it.

In line with this, the next generation of EnT catalyst evolution could focus on exploring heterogenous catalysts, earth-abundant metal-based photocatalysts, water-soluble sensitizers, new ligand designs for metal-based catalysts for enantioselective induction, and substrate-based architectural upgradation to facilitate donor–acceptor interaction for better reactivity control. Examples such as capturing substrates in a photochemically active metal–organic cage,[176] insertion of benzophenone derivatives in photoDNAzyme for enantioselective transformations,[177] or the [4+4] photodimerization of anthracenecarboxylates in a Pt(II)-based cyclodextrin supramolecular cage[178] showcase recent advancement in the field of catalyst design. Along these lines, the incorporation of photosensitizer and catalytically active first-row transition metal into a 3D matrix (*e.g.* artificial protein, polymer), has shown potential in controlling distance-dependent EnT-mediated reductive elimination.[160,164] Here, future directions may include its utilization in different elementary steps or – in the presence of a chiral environment – enantioselective transformations. Furthermore, the concept of valence-inverted reactivity might bear potential for the mechanistic understanding and development of new applications of other sensitized first-row transition metals with d^8-configuration, including Co^I and Cu^{III}.[154]

Visible light compensating for the free energy loss in endergonic reactions represents a unique design principle, complementing widely applied ground-state reactivities. Incorporating additional thermal energy to access rotational and vibrational levels, uphill energy transfer steps can also be feasible.[58,179] Anti-thermodynamic positional alkene isomerization and substrate epimerization is a rapidly evolving research field, yet limited to photoredox and/or HAT catalytic reaction modes to date,[132] hinting toward a complementary role of EnT catalysis in the near future. Furthermore, the isomerization of cinnamyl chlorides to cyclopropanes through EnT-mediated homolytic C–Cl bond scission by Troian-Gautier, Gutierrez, and Tambar[119] is a promising example to go beyond geometrical double bond isomerization. Utilizing substrate recognition in contra-entropic deracemization processes through EnT catalysis,[180] the development of new binding sites (*e.g.* electrostatic interactions with zwitter-ionic character) for sensitizer-tethered catalysts is in high demand, while the use of *in situ* generated chiral intermediates for precise enantiocontrol upon sensitization shows high potential.

The advancement of dual catalysis in combining EnT with different sectors of catalysis could leverage the latent reactivity of unexplored substrates and provides access to attractive (dis)connections. The previously discussed study from Nagib

shows an efficient merger of σ-bond cleavage and metal catalysis.[113] A combined study from Temps, Qi, and our group efficiently merges photoredox catalysis, energy transfer catalysis, and nickel catalysis for carboxylic acid functionalization.[181] Such multi-process combinations could be invaluable in harnessing the intrinsic properties of different catalytic manifolds. Most importantly, conceptual cross-connections of the basic principles of EnT-catalyzed processes are evident, and have significantly contributed to the overall development of the area. We therefore hope this tutorial review to further guide the reader to value and take advantage of recent developments across subtopics, expanding the boundaries of (strain-releasing) cycloadditions, homolytic σ-bond scission, isomerization and deracemization processes, as well as first-row transition metal-sensitized processes.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank the Deutsche Forschungsgemeinschaft (Leibniz Award; GRK 2678-437785492, J. E. E.) and the Verband der Chemischen Industrie (VCI, Fonds der Chemischen Industrie, J. E. E., R. K.) for generous financial support.

References

1 *Green Photocatalysts*, ed. M. Naushad, S. Rajendran and E. Lichtfouse, Springer International Publishing, Cham, 1st edn, 2020, vol. 34.
2 D. H. Sliney, *Eye*, 2016, **30**, 222–229.
3 V. Gold, *The IUPAC Compendium of Chemical Terminology*, International Union of Pure and Applied Chemistry (IUPAC), Research Triangle Park, NC, 2019.
4 M. H. Shaw, J. Twilton and D. W. C. MacMillan, *J. Org. Chem.*, 2016, **81**, 6898–6926.
5 N. A. Romero and D. A. Nicewicz, *Chem. Rev.*, 2016, **116**, 10075–10166.
6 J. M. R. Narayanam and C. R. J. Stephenson, *Chem. Soc. Rev.*, 2011, **40**, 102–113.
7 L. Capaldo and D. Ravelli, *Eur. J. Org. Chem.*, 2017, 2056–2071.
8 J.-H. Shen, M. Shi and Y. Wei, *Chem. – Eur. J.*, 2023, **29**, e202301157.
9 M. N. Hopkinson, B. Sahoo, J.-L. Li and F. Glorius, *Chem. – Eur. J.*, 2014, **20**, 3874–3886.
10 X. Lang, J. Zhao and X. Chen, *Chem. Soc. Rev.*, 2016, **45**, 3026–3038.
11 C. K. Prier, D. A. Rankic and D. W. C. MacMillan, *Chem. Rev.*, 2013, **113**, 5322–5363.
12 F. Strieth-Kalthoff, M. J. James, M. Teders, L. Pitzer and F. Glorius, *Chem. Soc. Rev.*, 2018, **47**, 7190–7202.
13 F. Strieth-Kalthoff and F. Glorius, *Chem*, 2020, **6**, 1888–1903.
14 Q.-Q. Zhou, Y.-Q. Zou, L.-Q. Lu and W.-J. Xiao, *Angew. Chem. Int. Ed.*, 2019, **58**, 1586–1604.
15 J. Großkopf, T. Kratz, T. Rigotti and T. Bach, *Chem. Rev.*, 2022, **122**, 1626–1653.
16 F. H. Quina and G. T. M. Silva, *J. Photochem. Photobiol.*, 2021, **7**, 100042.
17 T. Förster, *Pure Appl. Chem.*, 1973, **34**, 225–234.
18 N. Ben Amor, C. Noûs, G. Trinquier and J.-P. Malrieu, *J. Chem. Phys.*, 2020, **153**, 44118.
19 C. M. Marian, *Annu. Rev. Phys. Chem.*, 2021, **72**, 617–640.
20 P. Anastas and N. Eghbali, *Chem. Soc. Rev.*, 2010, **39**, 301–312.
21 A. L. Andrady, A. M. Heikkilä, K. K. Pandey, L. S. Bruckman, C. C. White, M. Zhu and L. Zhu, *Photochem. Photobiol. Sci.*, 2023, **22**, 1177–1202.
22 A. Kaur, P. Kaur and S. Ahuja, *Anal. Methods*, 2020, **12**, 5532–5550.
23 T. Förster, *Ann. Phys.*, 1948, **437**, 55–75.
24 D. L. Dexter, *J. Chem. Phys.*, 1953, **21**, 836–850.
25 R. A. Marcus, *Rev. Mod. Phys.*, 1993, **65**, 599–610.
26 S. Bai, P. Zhang and D. N. Beratan, *J. Phys. Chem. C*, 2020, **124**, 18956–18960.
27 N. F. Nikitas, P. L. Gkizis and C. G. Kokotos, *Org. Biomol. Chem.*, 2021, **19**, 5237–5253.
28 D. L. Bunbury and T. T. Chuang, *Can. J. Chem.*, 1969, **47**, 2045–2055.
29 T. Neveselý, M. Wienhold, J. J. Molloy and R. Gilmour, *Chem. Rev.*, 2022, **122**, 2650–2694.
30 C. Gao, J. Zeng, X. Zhang, Y. Liu and Z.-P. Zhan, *Org. Lett.*, 2023, **25**, 3146–3151.
31 M. Haga, E. S. Dodsworth, G. Eryavec, P. Seymour and A. B. P. Lever, *Inorg. Chem.*, 1985, **24**, 1901–1906.
32 J. W. Arbogast, A. P. Darmanyan, C. S. Foote, F. N. Diederich, R. L. Whetten, Y. Rubin, M. M. Alvarez and S. J. Anz, *J. Phys. Chem.*, 1991, **95**, 11–12.
33 X. Cui, C. Zhang, K. Xu and J. Zhao, *J. Mater. Chem. C*, 2015, **3**, 8735–8759.
34 J. Zhao, W. Wu, J. Sun and S. Guo, *Chem. Soc. Rev.*, 2013, **42**, 5323–5351.
35 L. Huang, X. Cui, B. Therrien and J. Zhao, *Chem. – Eur. J.*, 2013, **19**, 17472–17482.
36 E. Öztürk Gündüz and E. Okutan, *New J. Chem.*, 2022, **46**, 23088–23095.
37 S. Guo, L. Ma, J. Zhao, B. Küçüköz, A. Karatay, M. Hayvali, H. G. Yaglioglu and A. Elmali, *Chem. Sci.*, 2014, **5**, 489–500.
38 J. B. Birks, *Nature*, 1967, **214**, 1187–1190.
39 T. R. Blum, Z. D. Miller, D. M. Bates, I. A. Guzei and T. P. Yoon, *Science*, 2016, **354**, 1391–1395.
40 R. Kleinmans, S. Dutta, K. Ozols, H. Shao, F. Schäfer, R. E. Thielemann, H. T. Chan, C. G. Daniliuc, K. N. Houk and F. Glorius, *J. Am. Chem. Soc.*, 2023, **145**, 12324–12332.
41 X. Huang, T. R. Quinn, K. Harms, R. D. Webster, L. Zhang, O. Wiest and E. Meggers, *J. Am. Chem. Soc.*, 2017, **139**, 9120–9123.
42 E. M. Sherbrook, H. Jung, D. Cho, M.-H. Baik and T. P. Yoon, *Chem. Sci.*, 2019, **11**, 856–861.

 This journal is © The Royal Society of Chemistry 2024

43 F. M. Hörmann, C. Kerzig, T. S. Chung, A. Bauer, O. S. Wenger and T. Bach, *Angew. Chem., Int. Ed.*, 2020, **59**, 9659–9668.
44 J. Ma, S. Chen, P. Bellotti, R. Guo, F. Schäfer, A. Heusler, X. Zhang, C. Daniliuc, M. K. Brown, K. N. Houk and F. Glorius, *Science*, 2021, **371**, 1338–1345.
45 B. Reiß, Q. Hu, E. Riedle and H.-A. Wagenknecht, *ChemPhotoChem*, 2021, **5**, 1009–1019.
46 F. Wilkinson and A. A. Abdel-Shafi, *J. Phys. Chem. A*, 1997, **101**, 5509–5516.
47 B. S. Jayashree, P. S. Nikhil and S. Paul, *Med. Chem.*, 2022, **18**, 915–925.
48 N. A. Meanwell, *J. Agric. Food Chem.*, 2023, **71**, 18087–18122.
49 L. Candish, K. D. Collins, G. C. Cook, J. J. Douglas, A. Gómez-Suárez, A. Jolit and S. Keess, *Chem. Rev.*, 2022, **122**, 2907–2980.
50 B. R. Shire and E. A. Anderson, *JACS Au*, 2023, **3**, 1539–1553.
51 A. Palai, P. Rai and B. Maji, *Chem. Sci.*, 2023, **14**, 12004–12025.
52 Y.-Z. Cheng, Z. Feng, X. Zhang and S.-L. You, *Chem. Soc. Rev.*, 2022, **51**, 2145–2170.
53 M. Okumura and D. Sarlah, *Eur. J. Org. Chem.*, 2020, 1259–1273.
54 M. Montalti, A. Credi, L. Prodi and M. T. Gandolfi, *Handbook of Photochemistry*, CRC Press, 2006.
55 M. Zamadar and A. Greer, in *Handbook of Synthetic Photochemistry*, ed. A. Albini and M. Fagnoni, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2009, pp. 353–386.
56 Z. Lu and T. P. Yoon, *Angew. Chem., Int. Ed.*, 2012, **51**, 10329–10332.
57 R. Alonso and T. Bach, *Angew. Chem., Int. Ed.*, 2014, **53**, 4368–4371.
58 A. Tröster, R. Alonso, A. Bauer and T. Bach, *J. Am. Chem. Soc.*, 2016, **138**, 7808–7811.
59 M. J. James, J. L. Schwarz, F. Strieth-Kalthoff, B. Wibbeling and F. Glorius, *J. Am. Chem. Soc.*, 2018, **140**, 8624–8628.
60 P. Rai, K. Maji, S. K. Jana and B. Maji, *Chem. Sci.*, 2022, **13**, 12503–12510.
61 T. Morofuji, S. Nagai, Y. Chitose, M. Abe and N. Kano, *Org. Lett.*, 2021, **23**, 6257–6261.
62 R. Guo, S. Adak, P. Bellotti, X. Gao, W. W. Smith, S. N. Le, J. Ma, K. N. Houk, F. Glorius, S. Chen and M. K. Brown, *J. Am. Chem. Soc.*, 2022, **144**, 17680–17691.
63 W. Wang, Y. Cai, R. Guo and M. K. Brown, *Chem. Sci.*, 2022, **13**, 13582–13587.
64 G. Zhen, G. Zeng, F. Wang, X. Cao and B. Yin, *Adv. Synth. Catal.*, 2023, **365**, 43–52.
65 B. K. Gall, A. K. Smith and E. M. Ferreira, *Angew. Chem., Int. Ed.*, 2022, **61**, e202212187.
66 M. Zhu, H. Xu, X. Zhang, C. Zheng and S.-L. You, *Angew. Chem.*, 2021, **133**, 7112–7116.
67 A. Gilbert and T. Bach, *Synlett*, 2023, 1343–1355.
68 J. Ma, S. Chen, P. Bellotti, T. Wagener, C. Daniliuc, K. N. Houk and F. Glorius, *Nat. Catal.*, 2022, **5**, 405–413.
69 P. Bellotti, T. Rogge, F. Paulus, R. Laskar, N. Rendel, J. Ma, K. N. Houk and F. Glorius, *J. Am. Chem. Soc.*, 2022, **144**, 15662–15671.
70 A. E. Hurtley, Z. Lu and T. P. Yoon, *Angew. Chem., Int. Ed.*, 2014, **53**, 8991–8994.
71 Y.-Q. Zou, S.-W. Duan, X.-G. Meng, X.-Q. Hu, S. Gao, J.-R. Chen and W.-J. Xiao, *Tetrahedron*, 2012, **68**, 6914–6919.
72 J. Sivaguru, T. Bach and V. Ramamurthy, *Photochem. Photobiol. Sci.*, 2022, **21**, 1333–1340.
73 A. Denisenko, P. Garbuz, S. V. Shishkina, N. M. Voloshchuk and P. K. Mykhailiuk, *Angew. Chem., Int. Ed.*, 2020, **59**, 20515–20521.
74 B. A. Wright, A. Matviitsuk, M. J. Black, P. García-Reynaga, L. E. Hanna, A. T. Herrmann, M. K. Ameriks, R. Sarpong and T. P. Lebold, *J. Am. Chem. Soc.*, 2023, **145**, 10960–10966.
75 A. Denisenko, P. Garbuz, N. M. Voloshchuk, Y. Holota, G. Al-Maali, P. Borysko and P. K. Mykhailiuk, *Nat. Chem.*, 2023, **15**, 1155–1163.
76 L. Herter, I. Koutsopetras, L. Turelli, T. Fessard and C. Salomé, *Org. Biomol. Chem.*, 2022, **20**, 9108–9111.
77 T. Rigotti and T. Bach, *Org. Lett.*, 2022, **24**, 8821–8825.
78 T. Rigotti, D. P. Schwinger, R. Graßl, C. Jandl and T. Bach, *Chem. Sci.*, 2022, **13**, 2378–2384.
79 J. Proessdorf, C. Jandl, T. Pickl and T. Bach, *Angew. Chem., Int. Ed.*, 2022, **61**, e202208329.
80 F. Pecho, Y. Sempere, J. Gramüller, F. M. Hörmann, R. M. Gschwind and T. Bach, *J. Am. Chem. Soc.*, 2021, **143**, 9350–9354.
81 K. A. Rykaczewski and C. S. Schindler, *Org. Lett.*, 2020, **22**, 6516–6519.
82 J. Zheng, X. Dong and T. P. Yoon, *Org. Lett.*, 2020, **22**, 6520–6525.
83 M. R. Becker, E. R. Wearing and C. S. Schindler, *Nat. Chem.*, 2020, **12**, 898–905.
84 J. C. G. Kürschner, L. Brüss and L. Naesborg, *Chem. – Eur. J.*, 2023, **29**, e202300627.
85 Y. Liu, D. Ni and M. K. Brown, *J. Am. Chem. Soc.*, 2022, **144**, 18790–18796.
86 N. Sun, J. Huang, J. Qian, T.-P. Zhou, J. Guo, L. Tang, W. Zhang, Y. Deng, W. Zhao, G. Wu, R.-Z. Liao, X. Chen, F. Zhong and Y. Wu, *Nature*, 2022, **611**, 715–720.
87 F. Lovering, J. Bikker and C. Humblet, *J. Med. Chem.*, 2009, **52**, 6752–6756.
88 M. Golfmann and J. C. L. Walker, *Commun. Chem.*, 2023, **6**, 9.
89 C. B. Kelly, J. A. Milligan, L. J. Tilley and T. M. Sodano, *Chem. Sci.*, 2022, **13**, 11721–11737.
90 P. Bellotti and F. Glorius, *J. Am. Chem. Soc.*, 2023, **145**, 20716–20732.
91 T. V. T. Nguyen, A. Bossonnet, M. D. Wodrich and J. Waser, *J. Am. Chem. Soc.*, 2023, **145**, 25411–25421.
92 A. S. Harmata, B. J. Roldan and C. R. J. Stephenson, *Angew. Chem., Int. Ed.*, 2023, **62**, e202213003.
93 R. Kleinmans, T. Pinkert, S. Dutta, T. O. Paulisch, H. Keum, C. G. Daniliuc and F. Glorius, *Nature*, 2022, **605**, 477–482.

94 Y. Liang, R. Kleinmans, C. G. Daniliuc and F. Glorius, *J. Am. Chem. Soc.*, 2022, **144**, 20207–20213.
95 R. Guo, Y.-C. Chang, L. Herter, C. Salome, S. E. Braley, T. C. Fessard and M. K. Brown, *J. Am. Chem. Soc.*, 2022, **144**, 7988–7994.
96 H.-M. Huang, P. Bellotti, J. Ma, T. Dalton and F. Glorius, *Nat. Rev. Chem.*, 2021, **5**, 301–321.
97 S. Da Lee, V. K. Soni and E. J. Cho, *Acc. Chem. Res.*, 2022, **55**, 2526–2541.
98 G. Tan, M. Das, R. Kleinmans, F. Katzenburg, C. Daniliuc and F. Glorius, *Nat. Catal.*, 2022, **5**, 1120–1130.
99 J. E. Erchinger, R. Hoogesteger, R. Laskar, S. Dutta, C. Hümpel, D. Rana, C. G. Daniliuc and F. Glorius, *J. Am. Chem. Soc.*, 2023, **145**, 2364–2374.
100 D. Leifert and A. Studer, *Angew. Chem., Int. Ed.*, 2020, **59**, 74–108.
101 T. Patra, P. Bellotti, F. Strieth-Kalthoff and F. Glorius, *Angew. Chem., Int. Ed.*, 2020, **59**, 3172–3177.
102 X. Wang, Y. Chen, P. Liang, J.-Q. Chen and J. Wu, *Org. Chem. Front.*, 2022, **9**, 4328–4333.
103 J. Majhi, R. K. Dhungana, Á. Rentería-Gómez, M. Sharique, L. Li, W. Dong, O. Gutierrez and G. A. Molander, *J. Am. Chem. Soc.*, 2022, **144**, 15871–15878.
104 J. Li, Y. Yuan, X. Bao, T. Sang, J. Yang and C. Huo, *Org. Lett.*, 2021, **23**, 3712–3717.
105 T. Patra, M. Das, C. G. Daniliuc and F. Glorius, *Nat. Catal.*, 2021, **4**, 54–61.
106 S.-Q. Lai, B.-Y. Wei, J.-W. Wang, W. Yu and B. Han, *Angew. Chem., Int. Ed.*, 2021, **60**, 21997–22003.
107 P.-J. Xia, F. Liu, Y.-M. Pan, M.-P. Yang and Y.-Y. Yang, *Org. Chem. Front.*, 2022, **9**, 2522–2528.
108 G. Tan, M. Das, H. Keum, P. Bellotti, C. Daniliuc and F. Glorius, *Nat. Chem.*, 2022, **14**, 1174–1184.
109 G. Tan, F. Paulus, Á. Rentería-Gómez, R. F. Lalisse, C. G. Daniliuc, O. Gutierrez and F. Glorius, *J. Am. Chem. Soc.*, 2022, **144**, 21664–21673.
110 G. Tan, F. Paulus, A. Petti, M.-A. Wiethoff, A. Lauer, C. Daniliuc and F. Glorius, *Chem. Sci.*, 2023, **14**, 2447–2454.
111 J. Dong, L. Krasnova, M. G. Finn and K. B. Sharpless, *Angew. Chem., Int. Ed.*, 2014, **53**, 9430–9448.
112 S. Kim, H. Oh, W. Dong, J. Majhi, M. Sharique, B. Matsuo, S. Keess and G. A. Molander, *ACS Catal.*, 2023, **13**, 9542–9549.
113 K. M. Nakafuku, Z. Zhang, E. A. Wappes, L. M. Stateman, A. D. Chen and D. A. Nagib, *Nat. Chem.*, 2020, **12**, 697–704.
114 Z. Zhang, D. T. Ngo and D. A. Nagib, *ACS Catal.*, 2021, **11**, 3473–3477.
115 A. F. Prusinowski, H. C. Sise, T. N. Bednar and D. A. Nagib, *ACS Catal.*, 2022, **12**, 4327–4332.
116 J. S. Meek, *J. Chem. Educ.*, 1975, **52**, 541.
117 M. Baag, A. Kar and N. P. Argade, *Tetrahedron*, 2003, **59**, 6489–6492.
118 T. Neveselý, J. J. Molloy, C. McLaughlin, L. Brüss, C. G. Daniliuc and R. Gilmour, *Angew. Chem., Int. Ed.*, 2022, **61**, e202113600.
119 B. Xu, L. Troian-Gautier, R. Dykstra, R. T. Martin, O. Gutierrez and U. K. Tambar, *J. Am. Chem. Soc.*, 2020, **142**, 6206–6215.
120 G. S. Hammond, J. Saltiel, A. A. Lamola, N. J. Turro, J. S. Bradshaw, D. O. Cowan, R. C. Counsell, V. Vogt and C. Dalton, *J. Am. Chem. Soc.*, 1964, **86**, 3197–3217.
121 K. Singh, S. J. Staig and J. D. Weaver, *J. Am. Chem. Soc.*, 2014, **136**, 5275–5278.
122 J. B. Metternich and R. Gilmour, *J. Am. Chem. Soc.*, 2015, **137**, 11254–11257.
123 J. B. Metternich, D. G. Artiukhin, M. C. Holland, M. von Bremen-Kühne, J. Neugebauer and R. Gilmour, *J. Org. Chem.*, 2017, **82**, 9955–9977.
124 K. Livingstone, M. Tenberge, F. Pape, C. G. Daniliuc, C. Jamieson and R. Gilmour, *Org. Lett.*, 2019, **21**, 9677–9680.
125 A. Padwa and F. Albrecht, *J. Am. Chem. Soc.*, 1974, **96**, 4849–4857.
126 X. Zhang and T. Rovis, *J. Am. Chem. Soc.*, 2021, **143**, 21211–21217.
127 J. J. Molloy, M. Schäfer, M. Wienhold, T. Morack, C. G. Daniliuc and R. Gilmour, *Science*, 2020, **369**, 302–306.
128 M. Wienhold, B. Kweon, C. McLaughlin, M. Schmitz, T. J. B. Zähringer, C. G. Daniliuc, C. Kerzig and R. Gilmour, *Angew. Chem., Int. Ed.*, 2023, **62**, e202304150.
129 Z. C. Litman, Y. Wang, H. Zhao and J. F. Hartwig, *Nature*, 2018, **560**, 355–359.
130 J. I. Day, K. Singh, W. Trinh and J. D. Weaver, *J. Am. Chem. Soc.*, 2018, **140**, 9934–9941.
131 T. D. Schoch and J. D. Weaver, *J. Am. Chem. Soc.*, 2023, **145**, 14945–14951.
132 P.-Z. Wang, W.-J. Xiao and J.-R. Chen, *Nat. Rev. Chem.*, 2023, **7**, 35–50.
133 M. Plaza, J. Großkopf, S. Breitenlechner, C. Bannwarth and T. Bach, *J. Am. Chem. Soc.*, 2021, **143**, 11209–11217.
134 A. Tröster, A. Bauer, C. Jandl and T. Bach, *Angew. Chem., Int. Ed.*, 2019, **58**, 3538–3541.
135 X. Li, R. J. Kutta, C. Jandl, A. Bauer, P. Nuernberger and T. Bach, *Angew. Chem., Int. Ed.*, 2020, **59**, 21640–21647.
136 L. Wimberger, T. Kratz and T. Bach, *Synthesis*, 2019, 4417.
137 J. Großkopf, A. A. Heidecker and T. Bach, *Angew. Chem., Int. Ed.*, 2023, **62**, e202305274.
138 J. Großkopf, M. Plaza, A. Seitz, S. Breitenlechner, G. Storch and T. Bach, *J. Am. Chem. Soc.*, 2021, **143**, 21241–21245.
139 A. Hölzl-Hobmeier, A. Bauer, A. V. Silva, S. M. Huber, C. Bannwarth and T. Bach, *Nature*, 2018, **564**, 240–243.
140 T. Kratz, P. Steinbach, S. Breitenlechner, G. Storch, C. Bannwarth and T. Bach, *J. Am. Chem. Soc.*, 2022, **144**, 10133–10138.
141 M. Huang, L. Zhang, T. Pan and S. Luo, *Science*, 2022, **375**, 869–874.
142 T. Hostmann, J. J. Molloy, K. Bussmann and R. Gilmour, *Org. Lett.*, 2019, **21**, 10164–10168.
143 C. Onneken, K. Bussmann and R. Gilmour, *Angew. Chem., Int. Ed.*, 2020, **59**, 330–334.
144 M. Parasram and V. Gevorgyan, *Chem. Soc. Rev.*, 2017, **46**, 6227–6240.
145 J. Corpas, M. Gomez-Mendoza, J. Ramírez-Cárdenas, V. A. de La Peña O'Shea, P. Mauleón, R. Gómez Arrayás and J. C. Carretero, *J. Am. Chem. Soc.*, 2022, **144**, 13006–13017.

 This journal is © The Royal Society of Chemistry 2024

146 C. Cruché, W. Neiderer and S. K. Collins, *ACS Catal.*, 2021, **11**, 8829–8836.

147 For isomerization through S1, see: T. Brégent, J.-P. Bouillon and T. Poisson, *Org. lett.*, 2020, **22**, 7688–7693 also referred to as inner-sphere photosensitization, classical/outer sphere possible: see ref. 121 and for an overview on first-row transition metal photosensitizers, see: K. Behm and R. D. McIntosh, *ChemPlusChem*, 2020, **85**, 2611–2618.

148 H. Zhang, C. Huang, X.-A. Yuan and S. Yu, *J. Am. Chem. Soc.*, 2022, **144**, 10958–10967.

149 H. Zhang, X. He, X.-A. Yuan and S. Yu, *ACS Catal.*, 2023, **13**, 2857–2866.

150 W.-J. Yoo, T. Tsukamoto and S. Kobayashi, *Org. Lett.*, 2015, **17**, 3640–3642.

151 D. R. Heitz, J. C. Tellis and G. A. Molander, *J. Am. Chem. Soc.*, 2016, **138**, 12715–12718.

152 R. Kancherla, K. Muralirajan, B. Maity, S. Karuthedath, G. S. Kumar, F. Laquai, L. Cavallo and M. Rueping, *Nat. Commun.*, 2022, **13**, 2737.

153 A. Dewanji, P. E. Krach and M. Rueping, *Angew. Chem., Int. Ed.*, 2019, **58**, 3566–3570.

154 S. Wang, P. Ma, S. Shaik and H. Chen, *J. Am. Chem. Soc.*, 2022, **144**, 14607–14613.

155 Doyle and coworkers concluded for similar, directly excited triplet Nickel species with tetrahedral geometry that Ni–Ar homolysis is favored over Ni–halogen homolysis: S. I. Ting, S. Garakyaraghi, C. M. Taliaferro, B. J. Shields, G. D. Scholes, F. N. Castellano and A. G. Doyle, *J. Am. Chem. Soc.*, 2020, **142**, 5800–5810.

156 E. R. Welin, C. Le, D. M. Arias-Rotondo, J. K. McCusker and D. W. C. MacMillan, *Science*, 2017, **355**, 380–385.

157 L. Tian, N. A. Till, B. Kudisch, D. W. C. MacMillan and G. D. Scholes, *J. Am. Chem. Soc.*, 2020, **142**, 4555–4559.

158 D.-L. Zhu, S. Jiang, Q. Wu, H. Wang, H.-Y. Li and H.-X. Li, *Org. Lett.*, 2021, **23**, 8327–8332.

159 T. Kim, S. J. McCarver, C. Lee and D. W. C. MacMillan, *Angew. Chem.*, 2018, **130**, 3546–3550.

160 Y. Pan, N. Zhang, C.-H. Liu, S. Fan, S. Guo, Z.-M. Zhang and Y.-Y. Zhu, *ACS Catal.*, 2020, **10**, 11758–11767.

161 M. Kudisch, C.-H. Lim, P. Thordarson and G. M. Miyake, *J. Am. Chem. Soc.*, 2019, **141**, 19479–19486.

162 L. R. Reddy, S. Kotturi, Y. Waman, V. Ravinder Reddy, C. Patel, A. Kobarne and S. Kuttappan, *J. Org. Chem.*, 2018, **83**, 13854–13860.

163 R. A. Escobar and J. W. Johannes, *Chem. – Eur. J.*, 2020, **26**, 5168–5173.

164 Y. Fu, J. Huang, Y. Wu, X. Liu, F. Zhong and J. Wang, *J. Am. Chem. Soc.*, 2021, **143**, 617–622.

165 Y. Kato, T. Yoshino, M. Gao, J.-Y. Hasegawa, M. Kojima and S. Matsunaga, *J. Am. Chem. Soc.*, 2022, **144**, 18450–18458.

166 J. B. Metternich and R. Gilmour, *J. Am. Chem. Soc.*, 2016, **138**, 1040–1045.

167 A. Chatterjee and B. König, *Angew. Chem., Int. Ed.*, 2019, **58**, 14289–14294.

168 L. Bao, J.-T. Cheng, Z.-X. Wang and X.-Y. Chen, *Org. Chem. Front.*, 2022, **9**, 973–978.

169 Q.-Y. Lin, X.-H. Xu and F.-L. Qing, *J. Org. Chem.*, 2014, **79**, 10434–10446.

170 A. Hossain, S. Pagire and O. Reiser, *Synlett*, 2017, 1707–1714.

171 J. Xuan, X.-D. Xia, T.-T. Zeng, Z.-J. Feng, J.-R. Chen, L.-Q. Lu and W.-J. Xiao, *Angew. Chem., Int. Ed.*, 2014, **53**, 5653–5656.

172 L. Pitzer, F. Schäfers and F. Glorius, *Angew. Chem., Int. Ed.*, 2019, **58**, 8572–8576.

173 T. O. Paulisch, L. A. Mai, F. Strieth-Kalthoff, M. J. James, C. Henkel, D. M. Guldi and F. Glorius, *Angew. Chem., Int. Ed.*, 2022, **61**, e202112695.

174 M. J. Genzink, J. B. Kidd, W. B. Swords and T. P. Yoon, *Chem. Rev.*, 2022, **122**, 1654–1716.

175 J. B. Geri, J. V. Oakley, T. Reyes-Robles, T. Wang, S. J. McCarver, C. H. White, F. P. Rodriguez-Rivera, D. L. Parker, E. C. Hett, O. O. Fadeyi, R. C. Oslund and D. W. C. MacMillan, *Science*, 2020, **367**, 1091–1097.

176 J. Guo, Y.-Z. Fan, Y.-L. Lu, S.-P. Zheng and C.-Y. Su, *Angew. Chem., Int. Ed.*, 2020, **59**, 8661–8669.

177 N. Gaß and H.-A. Wagenknecht, *Eur. J. Org. Chem.*, 2015, 6661–6668.

178 M. Rao, K. Kanagaraj, C. Fan, J. Ji, C. Xiao, X. Wei, W. Wu and C. Yang, *Org. Lett.*, 2018, **20**, 1680–1683.

179 F. Strieth-Kalthoff, C. Henkel, M. Teders, A. Kahnt, W. Knolle, A. Gómez-Suárez, K. Dirian, W. Alex, K. Bergander, C. G. Daniliuc, B. Abel, D. M. Guldi and F. Glorius, *Chem*, 2019, **5**, 2183–2194.

180 J. Großkopf and T. Bach, *Angew. Chem., Int. Ed.*, 2023, e202308241.

181 H. Wang, Z. Liu, A. Das, P. Bellotti, S. Megow, F. Temps, X. Qi and F. Glorius, *Nat. Synth.*, 2023, **2**, 1116–1126.



Glorius Late-Stage Saturation

Chapter | First Online: 19 May 2026

pp 282–284 | Cite this chapter

Access provided by Universitäts- und Landesbibliothek Münster

Download book PDF | Save chapter



Name Reactions

Jie Jack Li

Sections | References

Abstract


4. OG
Organisch-Chemisches Institut
GLORIUS BASTERDS
Hunting Science