Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Digital Discovery*, 2025, **4**, 384

Calibration-free quantification and automated data analysis for high-throughput reaction screening†

Felix Katzenburg, ‡ Florian Boser, ‡ Felix R. Schäfer, Philipp M. Pflüger and Frank Glorius *

The accelerated generation of reaction data through high-throughput experimentation and automation has the potential to boost organic synthesis. However, efforts to generate diverse reaction datasets or identify generally applicable reaction conditions are still hampered by limitations in reaction yield quantification. In this work, we present an automatable screening workflow that facilitates the analysis of reaction arrays with distinct products without relying on the isolation of product references for external calibrations. The workflow is enabled by a flexible liquid handler and parallel GC-MS and GC-Polyarc-FID analysis while we introduce pyGecko, an open-source Python library for processing GC raw data. pyGecko offers comprehensive analysis tools allowing for the determination of reaction outcomes of a 96-reaction array in under a minute. Our workflow's utility is showcased for the scope evaluation of a site-selective thiolation of halogenated heteroarenes and the comparison of four cross-coupling protocols for challenging C–N bond formations.

Received 30th October 2024
Accepted 16th December 2024

DOI: 10.1039/d4dd00347k

rsc.li/digitaldiscovery

Introduction

In organic chemistry,[1,2] high-throughput experimentation (HTE) has established itself as a powerful tool for reaction discovery,[3–8] optimization[9–13] and the collection of data for statistical modeling.[14–18] The miniaturization of synthetic procedures[19] accompanied by engineering advancements enables the generation of standardized reaction arrays for a growing variety of chemical transformations, including transition metal-catalyzed cross-couplings,[14,20] photoredox catalysis[3,21] and asymmetric hydrogenations.[6,7] In the field of molecular machine learning, it has further been shown that consistent metadata-enriched HTE datasets are suitable training data for the prediction of reaction outcomes.[22,23]

While automatization and digitalization gradually enter organic chemistry labs, access to automated or semi-automated HTE equipment has yet to be democratized.[24,25] An underlying reason for this lack of adaption lies in the design of automated systems. Instead of being generally applicable, they are typically specialized for a specific synthetic task, and therefore, workflows require careful development. Reaction analysis is often among the most challenging aspects to integrate as a widely applicable method enabling the desired throughput for various analytes has to be identified.[26]

When switching from batch chemistry to reaction arrays with hundreds to thousands of parallelized reactions, the analytical technique and interpretation of analytical data typically become the bottleneck in the reaction workflow.[27] This is not solely a consequence of the time required for measurements and data interpretation but is also caused by the prerequisites for quantitative reaction analysis: almost all analytical techniques routinely employed in HTE necessitate the tedious isolation of a reference for each target compound. However, this approach becomes impractical in multi-substrate screenings due to the substantial costs and labor associated with the isolation of reference materials. This limitation has impeded the application of HTE in scenarios where the quantification of diverse products formed from combinatorial substrate sets is required (Fig. 1a).

This challenge, however, is becoming ever more relevant, as such experimental designs have recently gained attention in studies aiming for reaction generality to facilitate library syntheses and automation.[28–30] In this context, optimizations seeking robust and substrate-independent conditions are performed by evaluating reaction outcomes on multiple substrates simultaneously. In addition, HTE has been employed to discover reactions *via* multi-substrate screening[4,31,32] and to rapidly investigate the scope of synthetic methods.[8] To avoid product isolation, however, quantification is typically performed assuming uniform UV-vis absorption coefficients for high-pressure liquid chromatography-diode-array detection (HPLC-DAD)[33] or uniform ionizabilities for gas chromatography-mass spectrometry (GC-MS).[34] As these simplifications introduce a non-

Organisch-Chemisches Institut, Universität Münster, Corrensstraße 40, 48149 Münster, Germany. E-mail: glorius@uni-muenster.de

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4dd00347k

‡ Both authors contributed equally.

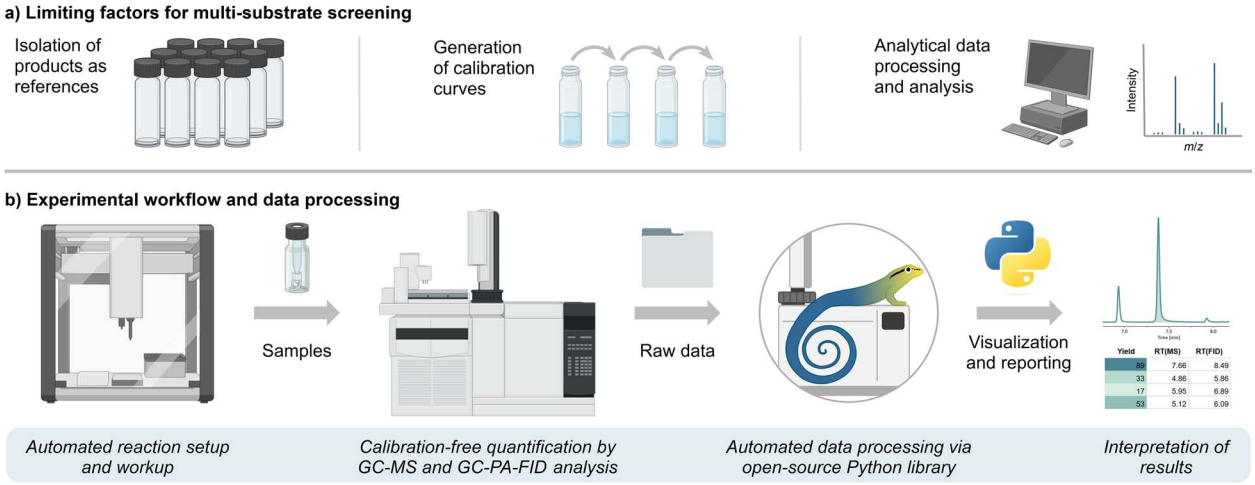

Fig. 1 (a) Challenges in HTE for multi-substrate screenings. (b) Experimental workflow comprising automated reaction setup and workup, GC-MS and GC-PA-FID analysis, and data processing *via* the open-source pyGecko library.

negligible error and strong molecule-dependent bias into the analysis, Jensen and coworkers recently developed a machine learning model for molar extinction coefficient prediction, which they used to estimate analyte UV-vis absorbance, trying to circumvent the need to record calibration curves.[35] However, with a median error of 18% associated with the method, this does not yet present a universal method for quantification. Other methods like matrix-assisted laser desorption/ionization–time-of-flight,[36] desorption electrospray ionization,[37] acoustic ejection[38] and acoustic mist ionization mass spectrometry[39] have also been applied for high-throughput analysis; however, their wider adaption has been limited by instrument availability, analyte compatibility and engineering challenges.[27,40] Another technique that has been used in this capacity is HPLC-coupled charged aerosol detection (CAD).[41] Although CAD gives a more uniform detector response for non-volatile compounds with a molecular weight over 300 Da, the inter-analyte response in the order of >11% still depends on an analyte's molar volume, basicity, density and charge. As these dependencies can only be partially factored into the analysis by considering an analyte's surface area, density and basic pK_a, a generally applicable approach for calibration-free yield quantification in high-throughput reaction screening is still highly sought after.[42]

A subsequent challenge arising downstream of methodical analysis is the processing of the extensive data output generated by HTE.[27] All relevant information (*i.e.*, peak retention times and areas) must be retrieved from the raw data and correlated with the experimental metadata (*i.e.*, identity of products and internal standards). Automating the evaluation of this interconnected data represents a great opportunity for software development, enabling chemists to concentrate on pertinent results and promoting the integration of analytical methods into closed-loop systems.[43–47] The development of such software is, however, thwarted by proprietary vendor formats, tying laboratories to commercial software that mostly lacks the flexibility and interoperability required for automation. These constraints have already spurred the development of open-source programs for handling HPLC-DAD[45] and -MS[48,49] raw data.

Considering the discussed challenges, three key problems must be addressed to leverage the full potential of HTE for combinatorial substrate sets: (i) quantification avoiding reference materials, (ii) identification and quantification of structurally diverse products, and (iii) elimination of manual data analysis through automated data processing.

Herein, we report a semi-automated workflow to run and quantitatively analyze reaction arrays with combinatorial product spaces (Fig. 1b). Automatic reaction setup and workup are performed using a Python programmable liquid handler. Accurate quantification is achieved by integrating GC-MS for product identification with GC-coupled Polyarc-flame ionization detection (PA-FID), ensuring a uniform response across organic analytes. We further developed an open-source Python library—pyGecko—for the processing of reaction metadata and analytical raw data. Our method's utility is demonstrated for the scope evaluation of a site-selective thiolation of heteroarenes[50] and the comparison of a recent metallaphotoredox to classical palladium-catalyzed cross-couplings for challenging C–N bond formations.[34]

Methods

Automated reaction setup and workup

For the setup and workup of 96-reaction arrays, standardized and traceable protocols were developed utilizing an OT-2 liquid handler. Stock solutions of substrates, reagents and catalysts were loaded onto the OT-2's deck and transferred into 96-position reaction blocks using custom protocols defined *via* an open-source Python application programming interface (API) (details in ESI section S1.2†). The resulting flexibility enabled the careful optimization of the protocols (*e.g.*, to minimize dead volumes) and the integration of custom laboratory equipment such as reaction blocks and vial holders. After irradiation or

heating (details in ESI section S1.1†) of the reaction mixtures, a sample from each well was taken by the liquid handler, filtered, diluted, and transferred to GC vials.

Analysis

Analytical techniques enabling the parallel identification and quantification of reaction components were required to determine the reaction outcomes for diverse combinations of substrates. We reasoned that combining GC-MS and GC-FID allowed us to pair structural information about analytes with a robust quantification method. However, due to the structure-dependent response factor of FID, as observed in MS or UV-vis detection, the isolation and calibration of each potential product would be necessary for a quantitative analysis.[51] To circumvent these limitations and avoid high errors introduced by approximated response factors,[52,53] we performed all quantifications on a GC system equipped with a Polyarc (PA) microreactor.[54] This system converts organic compounds to methane with a conversion of >99.9% prior to their detection by FID, ensuring a uniform detector response that is only dependent on a compound's number of carbon atoms (details in ESI section S1.4†).[55] By this modification, quantitative carbon detection is achievable *via* FID, except for samples in sulfur-containing solvents (due to catalyst poisoning)[56] and fully fluorinated analytes such as tetrafluoromethane or hexafluoroethane (due to their exceptional stability). The Polyarc microreactor is commercially available, compatible with most GC systems and easy to retrofit to existing instruments. Mapping between peaks in GC-MS and GC-PA-FID chromatograms was performed by assigning all peaks Kováts retention indices (RIs)[57] (details in ESI section S1.4†). With this procedure, only two additional calibration measurements with commercially available alkane standards are necessary to identify and quantify the target compounds. Alternatively, parallel MS and FID detection could also be employed to further streamline semi-quantitative analysis.

Data processing

The processing of analytical raw data to determine experimental outcomes in the proposed workflow is performed using the developed pyGecko package, a fast, flexible and freely available Python library for extracting and handling GC-MS and GC-FID data (Fig. 2a), which requires only basic programming skills to use. As proprietary formats typically impede the reading of vendor files, pyGecko includes parsing capabilities powered by the msConvert tool from the ProteoWizard software suite.[58] Open mzML and mzXML files can be accessed directly, and exporting data into these open formats is highly encouraged to follow FAIR data principles[59] and achieve the best software performance.

After the information in the raw output files is parsed, requiring only the specification of the file paths, measurements are processed automatically. Following smoothing, background subtraction, peak detection and integration, pyGecko offers a variety of analysis tools for the automated or semi-automated handling of gas chromatographic measurements and sequences (details in ESI section S2†). This includes the interpretation of measurements in the context of the experiment by reading in substrates and potential products, as well as the automatic identification of internal standards based on their retention time or mass spectrum. Information on reaction outcomes is extracted by identifying compounds based on the mass of their molecular ion or by comparison with a reference spectrum in the GC-MS chromatogram and by quantifying them relative to an internal standard in the GC-FID chromatogram. Prior to initiating a new screening campaign, the suitability of gas chromatography analysis for novel analyte classes should be evaluated, with particular attention to molecular ion detection or characteristic molecular fragments, to minimize false negative results (details in ESI section S2.3†). Results of an analysis, as well as chromatograms and spectra, can be compared, visualized, and reported in standardized formats such as the Open Reaction Database (ORD) schema.[60]

For the downstream analysis of reaction arrays, fully automated procedures are implemented to calculate reaction yields and conversions. To enable this analysis, the target compounds formed must be identified in the GC-MS chromatogram, the RIs of the corresponding peaks calculated, the peaks corresponding to this RI found in the GC-PA-FID chromatogram and the yields determined using the carbon number of the compound relative to the internal standard (Fig. 2b). The compounds are assigned to peaks in a total ion chromatogram (TIC) based on the mass trace of their molecular ion. The calculation of the expected isotopic pattern for the molecular ion and comparison to the pattern measured in the mass spectrum allows confident identification of the compounds.[61] The calculation of RIs is performed automatically, only requiring the specification of the calibration measurement's raw data files and retention time of one alkane (details in ESI section S2.4†). Automated quantitative analysis is performed on FID measurements relative to dodecane as an internal standard. Results are reported in the form of heatmaps and report tables (CSV) with the option to generate a comprehensive PDF report. Most importantly, results can be directly compiled in the ORD schema if a JSON file specifying reaction and analysis conditions is given (details in ESI section S2.4†). In practice, the entire evaluation is carried out using a short Python script, which simply requires the specification of the paths to the GC-MS and GC-PA-FID raw data and RI calibration measurements, as well as the reaction scheme and plate design, specifying only substrate SMILES, in the form of a CSV file. The only other user input required is the retention time of the internal standard used and of an alkane in the RI calibration measurements.

The modularity and flexibility of pyGecko also facilitate its adaption to new workflows and enable the seamless introduction of new functionalities. The entire workflow encompassing reaction setup and workup, analysis, and data processing was validated, choosing a site-selective thiolation of heteroarenes previously reported by our group as a model reaction.[50] This case study was designed to verify our analytical method and data processing pipeline for a set of structurally diverse compounds within a method development setting.

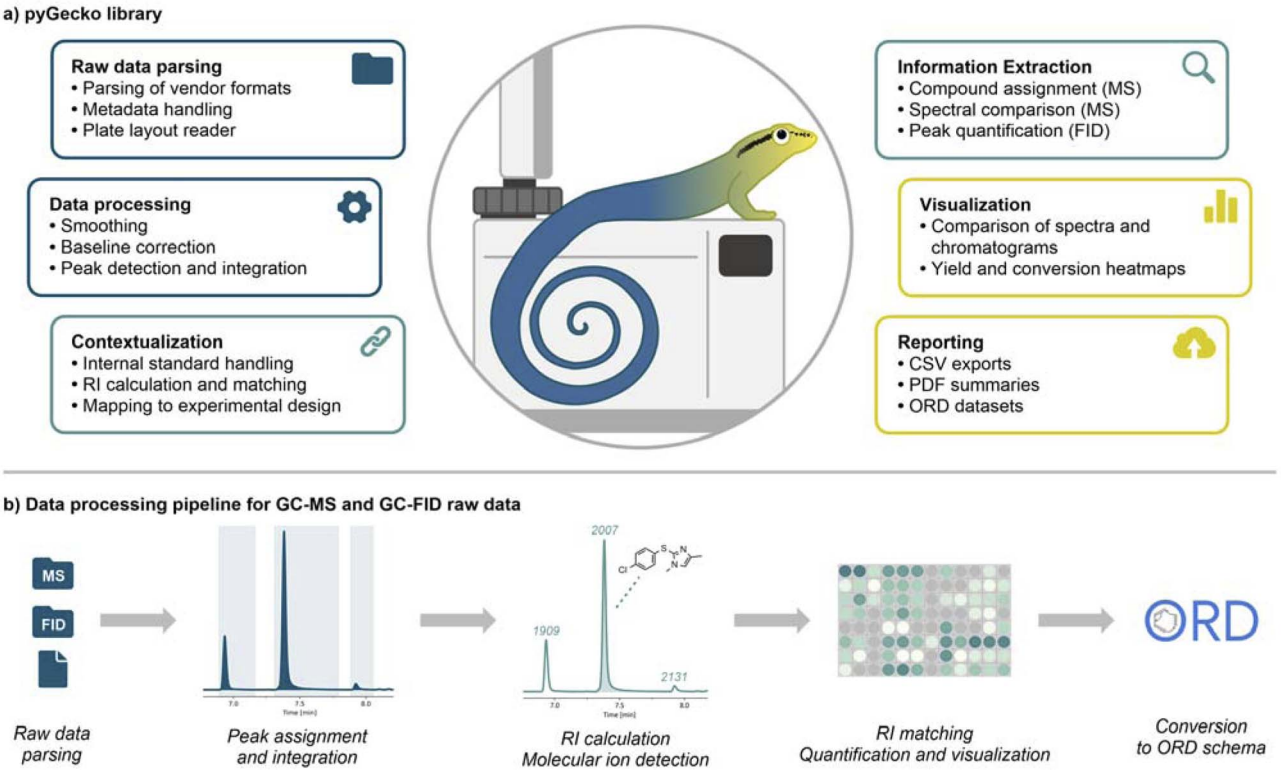

Fig. 2 (a) Overview of features and modules implemented in the pyGecko library. (b) Overview of the automated data processing pipeline for reaction arrays. Raw data files are exported and parsed in Python. Peaks are detected and integrated after background subtraction and smoothing. Product ion-to-charge ratios (m/z) are calculated based on the plate layout and detected in the MS chromatogram. After retention index (RI) calculation, the product MS peaks are matched to the corresponding FID peaks, and yields are calculated as relative peak areas compared to an internal standard. Results can be visualized as heatmaps and exported in the Open Reaction Database (ORD) schema.

Thiolation of heteroarenes

Heteroarenes are among the most relevant structural motifs in drug and materials discovery, making them key targets in synthetic method development.[62] Given the structural diversity of five-membered heterocycles, we chose a site-selective thiolation of halogenated heteroarenes[50] to validate the screening workflow and highlight its utility in combinatorial experimental designs. The functionalization of eight heteroarenes featuring four distinct cyclic cores was evaluated using twelve thiols bearing diverse functionalities like ester, ether, halo and different heterocyclic groups (Fig. 3a).

Before moving to a fully combinatorial 96-reaction array, control experiments were conducted to validate the different steps of the reaction workflow. Six thiolation products were synthesized in batch and isolated to evaluate the suitability of the analytical technique and the quality of the quantification. Triplicate measurements of samples containing the products and dodecane as an internal standard at 1 : 1, 1 : 2 and 1 : 3 ratios resulted in a mean absolute error of 2.2% and an average correlation coefficient of 0.998 using the GC-Polyarc-FID approach (details in ESI section S3.3†). These results show that the method enables the quantitative evaluation of reaction outcomes requiring minimal effort without compromising accuracy. The reproducibility of the reaction array setup and workup was validated by running the thiolation of **1** by **9** in all wells of a 96-reaction array. Thereby, an average yield of 78% with a standard deviation of 2.6% was recorded (details in ESI section S3.3†). To validate the data processing and analysis pipeline's accuracy, we compared pyGecko's quantification results with those obtained by manual analysis and employing commercial software, demonstrating excellent concordance (details in ESI section S3.3†).

With a validated reaction and analytical workflow in hand, we set out to investigate the reaction's performance in a combinatorial space defined by heteroarenes **1–8** and thiols **9–20**. Overall, higher reactivity was observed for more electron-rich imidazole (**1–3**), 2-bromo-5-phenylthiophene (**6**) and furan (**8**) substrates. As already illustrated in the original study, higher yields are observed for imidazoles substituted in the more electron-rich C5-position (**1**, **3**) compared to the less electrophilic C2-position (**2**) (Fig. 3b). Electron-poorer thiazoles (**5**) tend to perform worse for most thiols.

Focusing on the thiol scope, reactions featuring thiophenol derivatives generally (**12–14**) lead to higher yields. In terms of functional group tolerance, chlorides (**12**), bromides (**13**), ethers (**14**), and esters (**9**) are compatible, while primary amines (**15**) are not tolerated in most cases. The thiolation product of *N*-oxide **18** could not be detected for any heteroarene partner. However, the corresponding pyridine reduction product was identified in six cases (**2–4**, **6–8**). To confirm the product

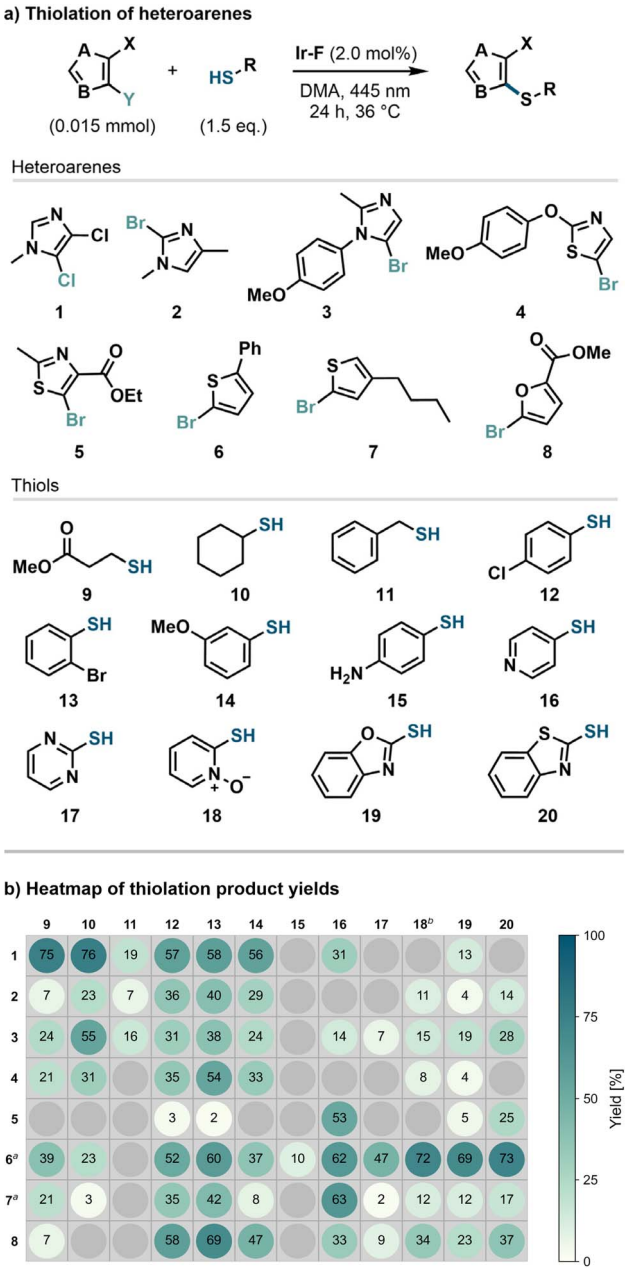

Fig. 3 (a) Reaction scheme and scope for the site-selective thiolation of (multi)halogenated heteroarenes. Standard conditions: heteroarene (0.015 mmol, 1.0 equiv), thiol (1.5 equiv.), [Ir{dF(CF$_3$)ppy}$_2$(dtbpy)]PF$_6$ (IrF) (2.0 mol%), DMA, 445 nm, rt, 24 h. [a]Pyridine (1.5 equiv.) added. [b]Yields are given for the corresponding pyridine reduction product. (b) Heatmap showing the yields of the thiolation product. Grey circles indicate that the expected product peak was not identified in the MS chromatogram.

structure, the thiolation product of **6** and **18** was isolated and analyzed by single crystal X-ray diffraction (details in ESI section S3.5†). One of the strengths of rapid scope evaluation in a 96-reaction array is that the higher throughput and lower experimental effort allow more diverse sets of substrates to be tested and interaction effects to be observed. Interestingly, this led to the finding that heterocyclic thiols **16** and **20** form thiolation products with the electron-deficient thiazole **5**, a substrate class that could not be utilized in the original study.

Overall, our workflow made it possible to evaluate a scope of 96 substrate combinations within two days, requiring 29 hours of measurement time (16 minutes per sample). This increased throughput opened up the possibility to study interaction effects of diverse substrates typically not captured in scope studies. Manual evaluation of these experiments would have required the identification of the product peaks in the MS chromatogram by evaluating the mass spectrum, assigning the corresponding peak in the FID chromatogram and calculating the reaction yield relative to the internal standard. Performing these tasks can easily occupy a trained chemist for multiple days. In contrast, the automated analysis facilitated by pyGecko completes these steps in less than a minute. The integration of visualization and reporting features within the software further streamlines the data analysis workflow, and conclusions can be drawn more quickly, accelerating research progress.

Evaluation of C–N cross-couplings

To further highlight the utility of the developed HTE protocol, our next objective was to demonstrate its applicability to a well-established chemical transformation. C–N cross-couplings are among the most utilized reactions in medicinal and materials chemistry.[63,64] Since their discovery 40 years ago, intensive research and continuous improvements have yielded efficient and easy-to-use precatalysts.[65–68] Since then, and due to the prevalence of C–N bonds in materials, natural products, and drugs, new and specialized methods to form bonds between carbon and nitrogen atoms continue to be an active area of research. In contrast, the development of highly generalized reaction conditions that enable the rapid development of molecular libraries and facilitate automation has recently gained attention.[28–30,41] On this topic, Koenig and coworkers recently published a nickel/photoredox dual-catalytic strategy for general C(sp^2)–(hetero)atom couplings. In this report, they realize nine distinct bond-forming reactions by introducing the concept of adaptive dynamic homogeneous catalysis (ADHoC).[34] Intrigued by the simplicity and generality of this strategy, we wanted to evaluate a set of substrates known to be especially challenging in common Buchwald-Hartwig coupling protocols under two palladium-catalyzed and two nickel ADHoC conditions (Fig. 4a). Such comparisons are especially important when starting screening campaigns or building compound libraries in medicinal chemistry.

In total, eight *N*-nucleophiles and six bromides were investigated, resulting in 48 potential coupling products. To test the limits of the methodologies, coupling partners were selected to contain substrates that are known to coordinate to transition metals (**22**, **32**), inhibit higher N–H acidity (**23**), are base sensitive (**25**, **26**, **34**) or sterically hindered (**24**, **27**, **33**).[63] Interestingly, a certain orthogonality can be observed between the two couplings (Fig. 4b and c). Under Buchwald–Hartwig conditions, electron-rich bromides such as the methoxy-substituted substrate **31** were well tolerated. Likewise, for sterically hindered substrates such as **24** and **33**, product formation in

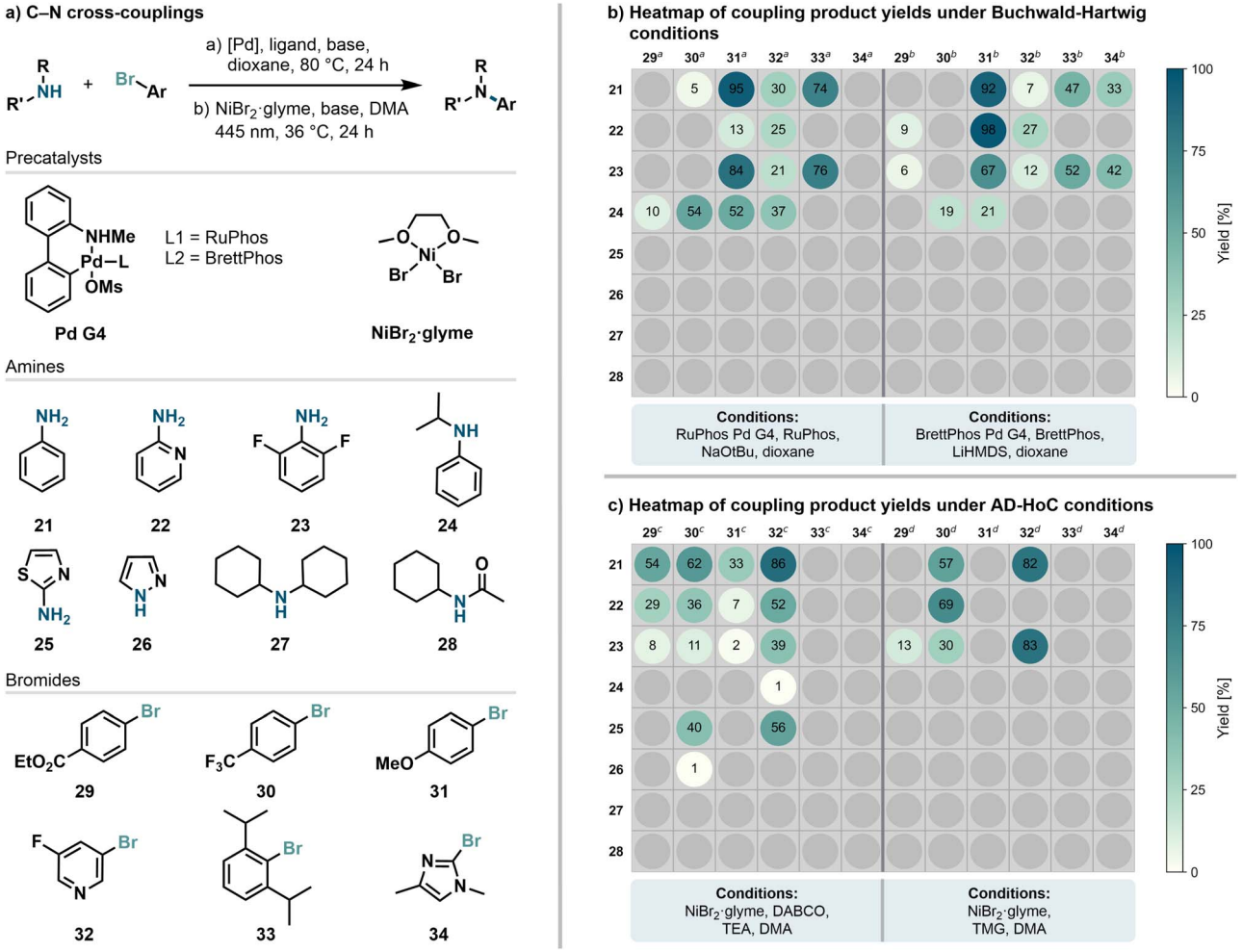

Fig. 4 (a) Reaction scheme, precatalysts and evaluated scope for C–N cross-couplings under Buchwald–Hartwig and adaptive dynamic homogeneous catalysis (AD-HoC) conditions. (b) Heatmap showing the yields of the C–N coupling products formed under Buchwald–Hartwig conditions. Grey circles indicate that the expected product peak was not identified in the MS chromatogram. Conditions: bromide (0.02 mmol, 1.0 equiv.), *N*-electrophile (1.2 equiv.). [a]RuPhos Pd G4 (7.5 mol%), RuPhos (7.5 mol%), NaOtBu (2.4 equiv). [b]BrettPhos Pd G4 (7.5 mol%), BrettPhos (7.5 mol%), LiHMDS (2.4 equiv.). (c) Heatmap showing the yields of the C–N coupling products formed under AD-HoC conditions. Grey circles indicate that the expected product peak was not identified in the MS chromatogram. Conditions: bromide (0.04 mmol, 1.0 eq.), *N*-electrophile (2.0 eq.). [c]$NiBr_2$·glyme (5.0 mol%) DABCO (1.8 equiv.) and *N*,*N*-diethylethanamine (0.25 equiv). [d]$NiBr_2$·glyme (5.0 mol%) TMG (1.2 equiv.).

good yields is observed more frequently under palladium-catalyzed conditions. On the other hand, the AD-HoC conditions were more suited for the conversion of coordinating substrates such as **22** and **32** as well as electron-deficient bromides **29** and **30** when the appropriate base is selected.

Substrates containing five-membered heterocycles (**25**, **26** and **34**) pose a challenge for both methods, as degradation pathways under basic conditions and catalyst deactivation necessitate specialized methods.[69] However, the 2-aminothiazole (**25**) coupling product was observed in moderate yields for electron-poor bromides **30** and **32** under nickel-catalyzed conditions. No product formation was observed for reactions featuring the sterically demanding dicyclohexylamine (**27**) and *N*-cyclohexylacetamide (**28**).

While quantitative method comparison often poses constraints regarding the number of substrates analyzed and the time required for a comprehensive analysis, our workflow enables the rapid examination of a broad spectrum of products. After the automatic or manual preparation and work-up of reaction arrays, a sample only needs to be analyzed by GC-MS and GC-FID to determine reaction yields automatically. pyGecko not only accelerates data processing, minimizing the time and human intervention required, but also offers valuable tools for the detailed analysis of experiments, including spectral matching and conversion monitoring (details in ESI section S4.3†).

Conclusions

In summary, we have developed a semi-automated screening workflow for the quantitative analysis of reaction arrays with diverse products. The method's utility was demonstrated for the photochemical thiolation of heteroarenes and four C–N cross-coupling protocols. Parallel GC-MS and GC-PA-FID analysis,

together with an open-source Python library (pyGecko) for gas chromatographic data processing, enabled the detection and quantification of 70 products in a single 96-reaction array without the need for prior calibration. We, therefore, demonstrate that our methodology is applicable to a broad spectrum of small molecules, provided that they are not subject to constraints in separation by gas chromatography or detection by mass spectrometry, owing to their volatility or differentiability by molecular mass. The evaluation of 288 distinct reactions, encompassing file parsing, peak detection, compound assignment, quantification, and visualization, was executed automatically by the program in a few minutes, allowing organic chemists to focus on the interpretation of experimental results. Due to the high flexibility and modularity of the pyGecko software, we anticipate that it will enable further applications in reaction optimization and automation as it allows reaction scopes to be evaluated in a few days and to identify the best conditions for screening campaigns rapidly. This will not only accelerate the evaluation of synthetic methodologies but also enable the generation of standardized data.

Data availability

The code developed and utilized within this study can be found at **https://github.com/FelixKatz77/pyGecko**. The data analysis scripts of this article are available at **https://github.com/FelixKatz77/pyGecko/tree/main/examples**. The chromatographic raw data, Opentrons protocols and labware definitions are available through Zenodo (**https://zenodo.org/records/10407762**).

Author contributions

Felix Katzenburg: conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, writing – original draft. Florian Boser: data curation, formal analysis, investigation, methodology, validation, writing – review & editing. Felix R. Schäfer: investigation, writing – review & editing. Philipp M. Pflüger: resources, writing – review & editing. Frank Glorius: conceptualization, funding acquisition, supervision, writing – review & editing.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors thank D. Rana, M. Kühnemund and T. Elsbecker for helpful discussions. Fig. 1 was created with BioRender.com. Different artists at flaticon.com are acknowledged for the icons used in Fig. 1 and 2. Generous financial support by the Deutsche Forschungsgemeinschaft (Priority Program "Molecular Machine Learning", SPP 2363) is gratefully acknowledged.

Notes and references

1 E. S. Isbrandt, R. J. Sullivan and S. G. Newman, *Angew. Chem., Int. Ed.*, 2019, **58**, 7180–7191.
2 B. Mahjour, Y. Shen and T. Cernak, *Acc. Chem. Res.*, 2021, **54**, 2337–2346.
3 A. McNally, C. K. Prier and D. W. C. MacMillan, *Science*, 2011, **334**, 1114–1117.
4 D. W. Robbins and J. F. Hartwig, *Science*, 2011, **333**, 1423–1427.
5 K. Troshin and J. F. Hartwig, *Science*, 2017, **357**, 175–181.
6 M. R. Friedfeld, M. Shevlin, J. M. Hoyt, S. W. Krska, M. T. Tudge and P. J. Chirik, *Science*, 2013, **342**, 1076–1080.
7 M. R. Friedfeld, H. Zhong, R. T. Ruck, M. Shevlin and P. J. Chirik, *Science*, 2018, **360**, 888–893.
8 Y. Shen, B. Mahjour and T. Cernak, *Commun. Chem.*, 2022, **5**, 83.
9 J. L. Douthwaite, R. Zhao, E. Shim, B. Mahjour, P. M. Zimmerman and T. Cernak, *J. Am. Chem. Soc.*, 2023, **145**, 10930–10937.
10 E. C. Hansen, D. J. Pedro, A. C. Wotal, N. J. Gower, J. D. Nelson, S. Caron and D. J. Weix, *Nat. Chem.*, 2016, **8**, 1126–1130.
11 Y. E. Lee, T. Cao, C. Torruellas and M. C. Kozlowski, *J. Am. Chem. Soc.*, 2014, **136**, 6782–6785.
12 V. Dimakos, D. P. Canterbury, S. Monfette, P. C. Roosen and S. G. Newman, *ACS Catal.*, 2022, **12**, 11557–11562.
13 K. A. Niederer, P. H. Gilmartin and M. C. Kozlowski, *ACS Catal.*, 2020, **10**, 14615–14623.
14 D. T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher and A. G. Doyle, *Science*, 2018, **360**, 186–190.
15 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.
16 F. Strieth-Kalthoff, F. Sandfort, M. H. S. Segler and F. Glorius, *Chem. Soc. Rev.*, 2020, **49**, 6154–6168.
17 D. F. Nippa, K. Atz, R. Hohler, A. T. Müller, A. Marx, C. Bartelmus, G. Wuitschik, I. Marzuoli, V. Jost, J. Wolfard, M. Binder, A. F. Stepan, D. B. Konrad, U. Grether, R. E. Martin and G. Schneider, *Nat. Chem.*, 2024, **16**, 239–248.
18 E. King-Smith, S. Berritt, L. Bernier, X. Hou, J. L. Klug-McLeod, J. Mustakis, N. W. Sach, J. W. Tucker, Q. Yang, R. M. Howard and A. A. Lee, *Nat. Chem.*, 2024, **16**, 633–643.
19 N. Gesmundo, K. Dykstra, J. L. Douthwaite, Y.-T. Kao, R. Zhao, B. Mahjour, R. Ferguson, S. Dreher, B. Sauvagnat, J. Saurí and T. Cernak, *Nat. Synth.*, 2023, **2**, 1082–1091.
20 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**, 49–53.
21 M. González-Esguevillas, D. F. Fernández, J. A. Rincón, M. Barberis, O. de Frutos, C. Mateos, S. García-Cerrada, J. Agejas and D. W. C. MacMillan, *ACS Cent. Sci.*, 2021, **7**, 1126–1134.

22 M. Saebi, B. Nan, J. E. Herr, J. Wahlers, Z. Guo, A. M. Zurański, T. Kogej, P.-O. Norrby, A. G. Doyle, N. V. Chawla and O. Wiest, *Chem. Sci.*, 2023, **14**, 4997–5005.
23 F. Strieth-Kalthoff, F. Sandfort, M. Kühnemund, F. R. Schäfer, H. Kuchen and F. Glorius, *Angew. Chem., Int. Ed.*, 2022, **61**, e202204647.
24 L. M. Roch, F. Häse, C. Kreisbeck, T. Tamayo-Mendoza, L. P. E. Yunker, J. E. Hein and A. Aspuru-Guzik, *PLoS One*, 2020, **15**, e0229862.
25 X. Caldentey and E. Romero, *Chem. Methods*, 2023, **3**, e202200059.
26 M. Christensen, L. P. E. Yunker, P. Shiri, T. Zepel, P. L. Prieto, S. Grunert, F. Bork and J. E. Hein, *Chem. Sci.*, 2021, **12**, 15473–15490.
27 R. Grainger and S. Whibley, *Org. Process Res. Dev.*, 2021, **25**, 354–364.
28 N. H. Angello, V. Rathore, W. Beker, A. Wołos, E. R. Jira, R. Roszak, T. C. Wu, C. M. Schroeder, A. Aspuru-Guzik, B. A. Grzybowski and M. D. Burke, *Science*, 2022, **378**, 399–405.
29 C. C. Wagen, S. E. McMinn, E. E. Kwan and E. N. Jacobsen, *Nature*, 2022, **610**, 680–686.
30 J. Rein, S. D. Rozema, O. C. Langner, S. B. Zacate, M. A. Hardy, J. C. Siu, B. Q. Mercado, M. S. Sigman, S. J. Miller and S. Lin, *Science*, 2023, **380**, 706–712.
31 H. Kim, G. Gerosa, J. Aronow, P. Kasaplar, J. Ouyang, J. B. Lingnau, P. Guerry, C. Farès and B. List, *Nat. Commun.*, 2019, **10**, 770.
32 X. Gao and H. B. Kagan, *Chirality*, 1998, **10**, 120–124.
33 D. Perera, J. W. Tucker, S. Brahmbhatt, C. J. Helal, A. Chong, W. Farrell, P. Richardson and N. W. Sach, *Science*, 2018, **359**, 429–434.
34 I. Ghosh, N. Shlapakov, T. A. Karl, J. Düker, M. Nikitin, J. V. Burykina, V. P. Ananikov and B. König, *Nature*, 2023, **619**, 87–93.
35 M. A. McDonald, B. A. Koscher, R. B. Canty and K. F. Jensen, *Chem. Sci.*, 2024, **15**, 10092–10100.
36 S. Lin, S. Dikler, W. D. Blincoe, R. D. Ferguson, R. P. Sheridan, Z. Peng, D. V. Conway, K. Zawatzky, H. Wang, T. Cernak, I. W. Davies, D. A. DiRocco, H. Sheng, C. J. Welch and S. D. Dreher, *Science*, 2018, **361**, eaar6236.
37 D. L. Logsdon, Y. Li, T. J. Paschoal Sobreira, C. R. Ferreira, D. H. Thompson and R. G. Cooks, *Org. Process Res. Dev.*, 2020, **24**, 1647–1657.
38 K. J. Dirico, W. Hua, C. Liu, J. W. Tucker, A. S. Ratnayake, M. E. Flanagan, M. D. Troutman, M. C. Noe and H. Zhang, *ACS Med. Chem. Lett.*, 2020, **11**, 1101–1110.
39 I. Sinclair, M. Bachman, D. Addison, M. Rohman, D. C. Murray, G. Davies, E. Mouchet, M. E. Tonge, R. G. Stearns, L. Ghislain, S. S. Datwani, L. Majlof, E. Hall, G. R. Jones, E. Hoyes, J. Olechno, R. N. Ellson, P. E. Barran, S. D. Pringle, M. R. Morris and J. Wingfield, *Anal. Chem.*, 2019, **91**, 3790–3794.
40 W. D. Blincoe, S. Lin, S. D. Dreher and H. Sheng, *Tetrahedron*, 2020, **76**, 131434.
41 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.
42 M. W. Robinson, A. P. Hill, S. A. Readshaw, J. C. Hollerton, R. J. Upton, S. M. Lynn, S. C. Besley and B. J. Boughtflower, *Anal. Chem.*, 2017, **89**, 1772–1777.
43 D. A. Boiko, K. S. Kozlov, J. V. Burykina, V. V. Ilyushenkova and V. P. Ananikov, *J. Am. Chem. Soc.*, 2022, **144**, 14590–14606.
44 A. Howarth, K. Ermanis and J. M. Goodman, *Chem. Sci.*, 2020, **11**, 4351–4359.
45 C. P. Haas, M. Lübbesmeyer, E. H. Jin, M. A. McDonald, B. A. Koscher, N. Guimond, L. Di Rocco, H. Kayser, S. Leweke, S. Niedenführ, R. Nicholls, E. Greeves, D. M. Barber, J. Hillenbrand, G. Volpin and K. F. Jensen, *ACS Cent. Sci.*, 2023, **9**, 307–317.
46 J. Liu and J. E. Hein, *Nat. Synth.*, 2023, **2**, 464–466.
47 G. Wuitschik, V. Jost, T. Schindler and M. Jakubik, *Org. Process Res. Dev.*, 2024, **28**, 2875–2884.
48 J. Mason, H. Wilders, D. J. Fallon, R. P. Thomas, J. T. Bush, N. C. O. Tomkinson and F. Rianjongdee, *Digital Discovery*, 2023, **2**, 1894–1899.
49 R. Schmid, S. Heuckeroth, A. Korf, A. Smirnov, O. Myers, T. S. Dyrlund, R. Bushuiev, K. J. Murray, N. Hoffmann, M. Lu, A. Sarvepalli, Z. Zhang, M. Fleischauer, K. Dührkop, M. Wesner, S. J. Hoogstra, E. Rudt, O. Mokshyna, C. Brungs, K. Ponomarov, L. Mutabdžija, T. Damiani, C. J. Pudney, M. Earll, P. O. Helmer, T. R. Fallon, T. Schulze, A. Rivas-Ubach, A. Bilbao, H. Richter, L.-F. Nothias, M. Wang, M. Orešič, J.-K. Weng, S. Böcker, A. Jeibmann, H. Hayen, U. Karst, P. C. Dorrestein, D. Petras, X. Du and T. Pluskal, *Nat. Biotechnol.*, 2023, **41**, 447–449.
50 F. Sandfort, T. Knecht, T. Pinkert, C. G. Daniliuc and F. Glorius, *J. Am. Chem. Soc.*, 2020, **142**, 6913–6919.
51 E. Cicchetti, P. Merle and A. Chaintreau, *Flavour Fragrance J.*, 2008, **23**, 450–459.
52 J.-Y. de Saint Laumer, S. Leocata, E. Tissot, L. Baroux, D. M. Kampf, P. Merle, A. Boschung, M. Seyfried and A. Chaintreau, *J. Sep. Sci.*, 2015, **38**, 3209–3217.
53 J. T. Scanlon and D. E. Willis, *J. Chromatogr. Sci.*, 1985, **23**, 333–340.
54 S. Maduskar, A. R. Teixeira, A. D. Paulsen, C. Krumm, T. J. Mountziaris, W. Fan and P. J. Dauenhauer, *Lab Chip*, 2015, **15**, 440–447.
55 C. S. Spanjers, C. A. Beach, A. J. Jones and P. J. Dauenhauer, *Anal. Methods*, 2017, **9**, 1928–1934.
56 C. A. Beach, K. E. Joseph, P. J. Dauenhauer, C. S. Spanjers, A. J. Jones and T. J. Mountziaris, *AIChE J.*, 2017, **63**, 5438–5444.
57 E. Kováts, *Helv. Chim. Acta*, 1958, **41**, 1915–1932.
58 M. C. Chambers, B. Maclean, R. Burke, D. Amodei, D. L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egertson, K. Hoff, D. Kessner, N. Tasman, N. Shulman, B. Frewen, T. A. Baker, M.-Y. Brusniak, C. Paulse, D. Creasy, L. Flashner, K. Kani, C. Moulding, S. L. Seymour, L. M. Nuwaysir, B. Lefebvre, F. Kuhlmann, J. Roark, P. Rainer, S. Detlev, T. Hemenway, A. Huhmer, J. Langridge, B. Connolly, T. Chadick, K. Holly, J. Eckels, E. W. Deutsch, R. L. Moritz, J. E. Katz, D. B. Agus,

M. MacCoss, D. L. Tabb and P. Mallick, *Nat. Biotechnol.*, 2012, **30**, 918–920.

59 M. D. Wilkinson, M. Dumontier, I. J. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. Da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao and B. Mons, *Sci. Data*, 2016, **3**, 160018.

60 S. M. Kearnes, M. R. Maser, M. Wleklinski, A. Kast, A. G. Doyle, S. D. Dreher, J. M. Hawkins, K. F. Jensen and C. W. Coley, *J. Am. Chem. Soc.*, 2021, **143**, 18820–18826.

61 P. Dittwald, J. Claesen, T. Burzykowski, D. Valkenborg and A. Gambin, *Anal. Chem.*, 2013, **85**, 1991–1994.

62 E. Vitaku, D. T. Smith and J. T. Njardarson, *J. Med. Chem.*, 2014, **57**, 10257–10274.

63 P. Ruiz-Castillo and S. L. Buchwald, *Chem. Rev.*, 2016, **116**, 12564–12649.

64 D. G. Brown and J. Boström, *J. Med. Chem.*, 2016, **59**, 4443–4458.

65 M. Kosugi, M. Kameyama and T. Migita, *Chem. Lett.*, 1983, **12**, 927–928.

66 D. Maiti, B. P. Fors, J. L. Henderson, Y. Nakamura and S. L. Buchwald, *Chem. Sci.*, 2011, **2**, 57–68.

67 N. C. Bruno, M. T. Tudge and S. L. Buchwald, *Chem. Sci.*, 2013, **4**, 916–920.

68 N. C. Bruno, N. Niljianskul and S. L. Buchwald, *J. Org. Chem.*, 2014, **79**, 4161–4166.

69 E. C. Reichert, K. Feng, A. C. Sather and S. L. Buchwald, *J. Am. Chem. Soc.*, 2023, **145**, 3323–3329.