ORGANIC CHEMISTRY

A machine-learning tool to predict substrate-adaptive conditions for Pd-catalyzed C–N couplings

N. Ian Rinehart[1], Rakesh K. Saunthwal[1], Joël Wellauer[2], Andrew F. Zahrt[1], Lukas Schlemper[2], Alexander S. Shved[1], Raphael Bigler[2]*, Serena Fantasia[2]*, Scott E. Denmark[1]*

Machine-learning methods have great potential to accelerate the identification of reaction conditions for chemical transformations. A tool that gives substrate-adaptive conditions for palladium (Pd)–catalyzed carbon-nitrogen (C–N) couplings is presented. The design and construction of this tool required the generation of an experimental dataset that explores a diverse network of reactant pairings across a set of reaction conditions. A large scope of C–N couplings was actively learned by neural network models by using a systematic process to design experiments. The models showed good performance in experimental validation: Ten products were isolated in more than 85% yield from a range of couplings with out-of-sample reactants designed to challenge the models. Importantly, the developed workflow continually improves the prediction capability of the tool as the corpus of data grows.

The strategic value of carbon-nitrogen couplings makes them important transformations in many domains of the chemical enterprise. In particular, Buchwald-Hartwig (B-H) couplings (*1*, *2*) are among the most important C–N bond-forming reactions and have revolutionized the practice of modern synthetic organic chemistry (*3*). In this process, palladium complexes catalyze the cross-coupling of (hetero)aryl electrophiles with various nitrogen nucleophiles. Experimentalists routinely identify substrate-specific conditions for new B-H couplings. The extensive scope of electrophiles and nucleophiles competent in this transformation required the development of many catalysts and conditions to enable successful couplings of diverse reaction partners (*3*, *4*). Selection of the appropriate palladium ligand is particularly important because B-H couplings are exceptionally sensitive to changes in ligand structure (*5*).

Empirical guides with selected examples from the literature and heuristics on the basis of reported couplings are available to help experimentalists select appropriate ligands and conditions for a given coupling ("prior experience" in Fig. 1) (*6*–*8*). Because these recommendations are derived from literature data, they are limited to previous experience (i.e., retrospective). Specifically, Wuitschik's guide (*7*) recommends the same conditions for all heteroaryl bromides, though more granular, heteroarene-specific conditions are often required. Buchwald's original user guide recommends conditions for a limited range of heteroaryl bromides (pyridines, pyrazines, pyrimidines, thiophenes, oxazoles, thiazoles, and pyrazoles) on the basis of work from just two references (*9*, *10*) and otherwise focuses on recommending ligands for specific nucleophile types. For example, 2-amino oxazoles are a challenging class of nucleophile that required a specific publication from Buchwald's group (*11*). More recently, high-throughput experimentation has been used to evaluate a range of five-membered heteroaryl bromides in B-H couplings, and that work highlights the difficulty of couplings between five-membered heteroaryl bromides and aliphatic heterocycles (*12*). Even with these published reports, the chemical literature does not come close to describing the enormous scope of possible B-H reactant pairings; thus, when a new (hetero)aryl halide is used, experimentalists must rely on intuition.

The use of B-H couplings often creates a bottleneck in routine synthesis campaigns in both academia and industry. An experimentalist begins with a specific chemistry problem: coupling of a new pair of reactants (Fig. 1). They then identify a subset of conditions on the basis of prior knowledge, amalgamating recommendations from the B-H user guides and B-H cheat sheets described above (when recommendations are available), as well as personal experience, intuition, and specific literature precedent. Those prior knowledge–based recommendations serve as a starting point for an experimental campaign to survey catalyst-solvent-base combinations for experimental hits. For many applications, a broad range of compounds must be synthesized in a timely manner, and hits are an end point. In practice, an experimentalist will invest time and resources into an optimization campaign to fine-tune those conditions only if necessary. Of note, the Doyle group recently published a machine learning (ML)–based tool that accelerates the optimization of yield from a user-defined reaction space ("reaction optimizer" in Fig. 1) (*13*). Our goal was to create an ML-guided tool that immediately provides predicted hits for a new proposed coupling (which could then be optimized if necessary), offering more than empirical guides and avoiding an experimental campaign from an empirical approach, thereby accelerating the routine application of B-H couplings (Fig. 1). This goal is complementary to optimization, and we foresee that the combination of the two could potentially create an end-to-end artificial intelligence–driven process like that shown in Fig. 1.

From an ML standpoint, there are crucial differences between an optimizer tool and the tool proposed in this work. A visual illustration contrasting reaction optimization to a tool based on substrate-adaptive models is shown in Fig. 2. A three-dimensional plot represents a hypothetical reaction space where any specific combination of reactant(s) and condition(s) produce an unknown yield, and the goal of using ML is to use relatively few measured yields to predict the rest. After selecting a specific coupling from all those possible [a slice of a hypothetical reaction space along the reactant dimension(s); top of Fig. 2], an optimizer directs the selection of experiments within that slice of reaction space to increase yield. Sophisticated optimizers are still being developed to make iterative rounds of experimentation as efficient as possible (*14*). An optimizer could, in principle, be used on multiple reactants at once, but the reactants must have related reactivity trends (i.e., the slices in reaction space are close enough that similar conditions have similar reactivity). Because B-H couplings are so sensitive to reactant structure (vide supra), a new slice—even close in the reactant dimension(s)—frequently requires a fresh start. A complementary approach to the optimizer, shown on the right in Fig. 2,

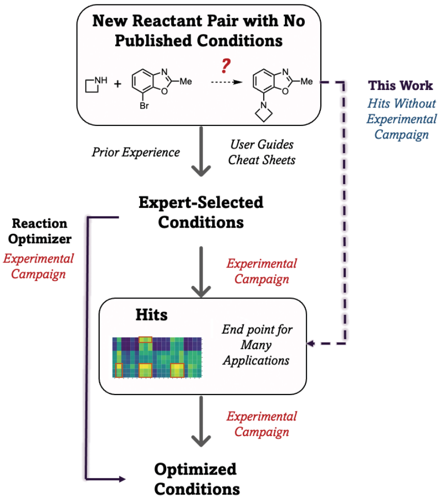

Fig. 1. The goal of this work. Identifying conditions that furnish synthetically useful yields for new couplings without experimental campaigns.

[1]Roger Adams Laboratory, Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [2]Pharmaceutical Division, Synthetic Molecules Technical Development, Process Chemistry and Catalysis, F. Hoffmann–La Roche, Ltd., Basel, Switzerland.
*Corresponding author. Email: sdenmark@illinois.edu (S.E.D.); raphael.bigler@roche.com (R.B.); serena_maria.fantasia@roche.com (S.F.)

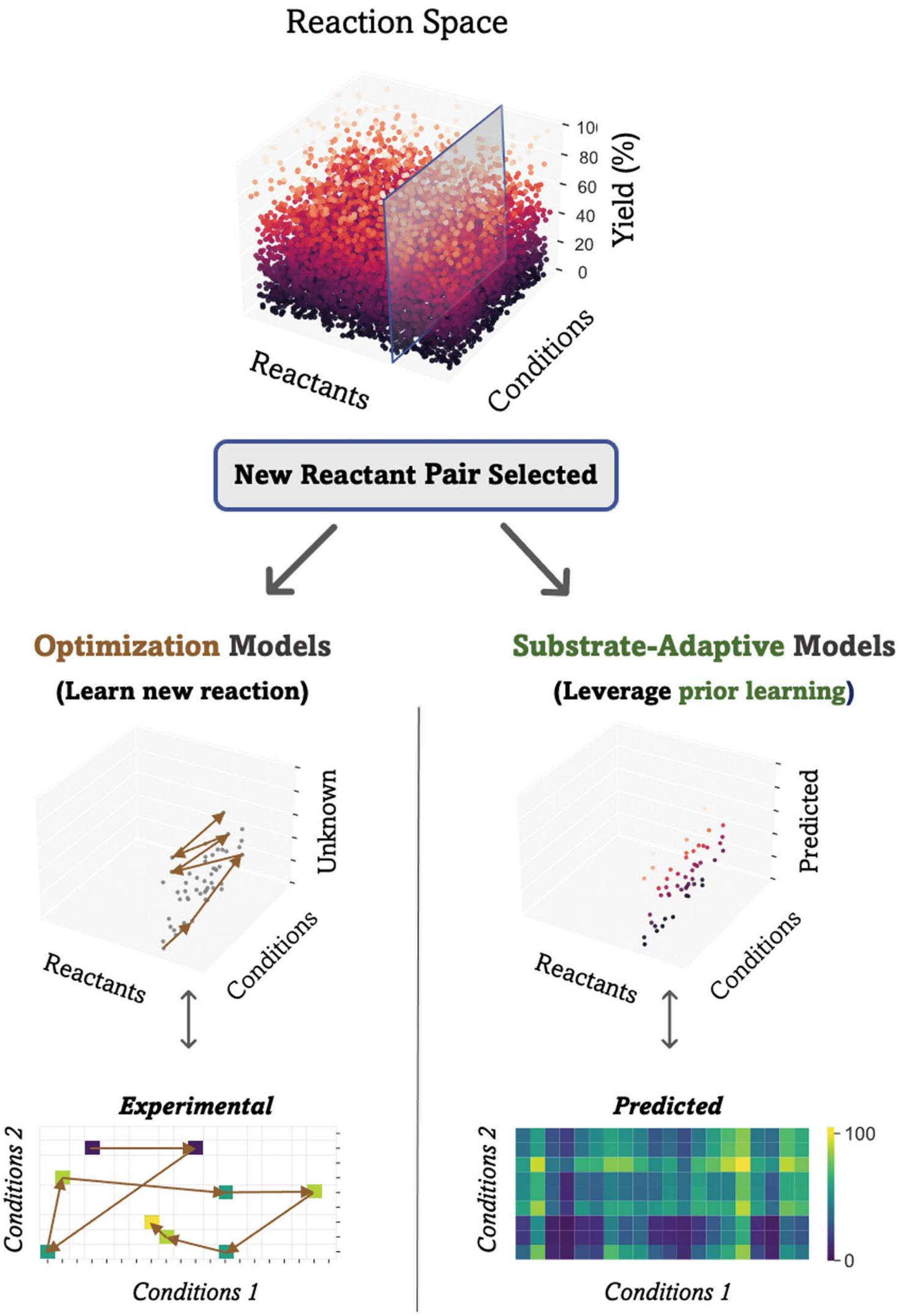

Fig. 2. Defining substrate-adaptive models and contrasting them with ML-assisted optimization models.

addresses this limitation because it involves training models on the entire reaction space beforehand. Thus, when a new specific coupling is selected, those models can immediately leverage prior learning to predict the yield patterns of that new reaction without additional experiments. This approach produces predicted hits, which may still need validation, but circumvents an experimental campaign. That approach requires an appropriate experimental dataset that provides sufficient prior experience to the substrate-adaptive models. Before embarking on this enterprise, it was prudent to evaluate state-of-the-art applications of ML to predict the outcome of new B-H coupling reactions and related work.

In 2018, Doyle and co-workers trained ML models to predict the yields of B-H couplings between one nucleophile and several similar electrophiles (*15*). By the introduction of isoxazole additives (catalyst poisons) (*16*), the authors simulated the process of gathering data for many different reactant pairs while avoiding the analytical challenge of having many more distinct products. Models effectively predicted the extent of catalyst poisoning by the additives and were tested on out-of-sample additives. Importantly, their models were not applied to predicting couplings with new nucleophiles or electrophiles and thus do not address the problem outlined in Fig. 1. Recent follow-up work using that same dataset is subject to the same inherent limitations and does not address the problem outlined above (*17*, *18*).

A year later, Li and Eastgate combined proprietary and literature B-H data to create classification models to predict which ligand to use for a specific B-H coupling for the desired synthetic route (*19*). Although the models successfully leveraged prior experiments to make predictions, experimental validation is limited to reactants that furnish one specific diarylamine product. Wuitschik and co-workers very recently reported their efforts using a similar approach to that of Li and Eastgate but with a robust experimental validation that demonstrated poor model performance (*20*). The authors attribute that to the limitations of the dataset and suggest that new approaches such as the workflow presented in this work are needed for B-H couplings, citing the associated preprint of this work (*21*).

In 2022, Zimmerman and co-workers published an interesting approach to transfer learning on B-H couplings, wherein models were trained on one type of reactant (i.e., benzamide) and used to predict effective conditions for a new substrate (i.e., pyrazole or an aniline) (*22*). Their goal was to maximize the value of limited data from one reaction to assist in the development of a new reaction, effectively learning one slice of a reaction space from Fig. 2 and applying that knowledge to another. When two slices show positive transfer, this approach avoids training models on the new reaction. In Zimmerman *et al.*'s work, positive transfer was observed between aniline and benzamide, whereas pyrazole and benzamide couplings showed poor transfer. That observation supports the premise that general conditions are unlikely to exist for B-H couplings.

In a very recent disclosure, Burke, Grzybowski, and co-workers reported general conditions for Suzuki-Miyaura (S-M) cross-couplings (*23*). Their approach bypasses using predictive models that respond to substrate structure, as described in Fig. 2. Instead, the goal of general conditions is to work across the many slices of the reaction space of S-M couplings, and their work relies on the assumption that those exist. Unfortunately, general conditions are unlikely to exist for B-H couplings (vide supra). The structural variation of diverse nitrogen nucleophiles and resulting changes in reactivity necessitate domain-specific conditions, and decades of research have not solved all those challenges. Zimmerman *et al.*'s work demonstrating poor transfer within B-H couplings illustrates the distinct reactivity domains of B-H coupling reaction space. As a case in point, Burke, Grzybowski, and co-workers used their workflow on a dataset of B-H couplings (*24*); however, no model predictions or recommended general conditions are presented.

Although also unrelated to B-H coupling reactions, Doyle and co-workers reported successfully modeling of a noncatalytic functional group interconversion of carbinols to alkyl fluorides (*25*). The authors collected a dataset capable of training ML models to learn the inherent reactivity patterns of many substrates for that reaction. Those models then successfully predict substrate-specific conditions for new reactants, as illustrated in Fig. 2. Doyle *et al.*'s approach is proof of concept for designing a prediction tool for substrate-adaptive conditions. However, by virtue of having a single reactant and no catalyst, that transformation has a substantially lower complexity.

Research strategy

The reactant dimension for B-H coupling reaction space like that in Fig. 2 actually comprises multiple subdimensions, and the conditions dimension also captures solvents, bases, and catalysts. All those dimensions are independent (nucleophile, electrophile, catalyst, solvent, and base), and all affect yield. Importantly, each reactant can show different preferences with regard to catalysts, solvents, and bases (*6*, *8*). As a result, models must learn the preferences of each reactant and the interaction terms between various combinations of them and then correctly weigh those to be useful. The data used to train such models must explore these complex relationships, and no such dataset of appropriate complexity exists. As a case in point, Schwaller *et al.* show that modeling on the US Patent and Trademark Office (USPTO) dataset of B-H couplings, which explores a broad range of reactants in a noncombinatorial fashion, did not produce predictive models (*17*). A dataset with a similar reactant diversity to that of the USPTO dataset is ideal, but that exact dataset cannot support models complex enough to address the problem outlined in Fig. 1.

To build such a dataset, the reactant dimensions must be unbounded so that it is possible to continue expanding to new reactant domains without starting over. Zimmerman *et al.*'s observation of poor transfer and good transfer between different reactant domains of B-H reaction space indicates that there may be types of couplings that can be grouped and learned together and others that must be separately addressed. A new strategy for dataset design that is founded on separating reactant domains is proposed. By combining expert knowledge, new chemical descriptors, and well-established clustering techniques, representative neighborhoods (subspaces) of the multidimensional B-H coupling reaction space could be identified. Then, in a subsequent experimental campaign, new data in new subspaces could be iteratively generated, and the applicability domain of models expanded when models are updated with that new data (vide infra).

Before running experiments, we had to define a starting point for exploring reactant dimensions in B-H couplings. The reactivity patterns and pitfalls of this important transformation are immensely complex: Indoles form constitutional isomers (*8*, *26*), benzylic and cyclic aliphatic amines are prone to an unproductive electrophile reduction pathway (*8*), and many heterocycles undergo unproductive, palladium-catalyzed C–H activation pathways, to name a few (*27*). The relative rates of these unproductive pathways, catalyst deactivation, and productive coupling all affect the yield and vary from substrate to substrate. For a model to predict useful conditions for new B-H reactions, it must learn under which circumstances these side reactions appear as well as the inherent structure-reactivity patterns for each reactant, catalyst, and condition. Accordingly, to make this work as more than just a proof of concept, problematic substrates like those described above were included.

Figure 3 shows a representative list of 19 of the 50 nitrogen nucleophiles used in this work. In a similar manner, 50 (hetero)aryl bromides were selected to broadly represent many building blocks of interest to pharmaceutical development as well as a range of electronic and steric properties [see supplementary materials (SM) for the full list]. The scope of both Zimmerman *et al.*'s and Doyle *et al.*'s work on B-H couplings is subsumed within and greatly expanded upon by the scope of this study. The boxed structures represent the good and poor transfer learning, represented by colored arrows, that was demonstrated by Zimmerman and co-workers (*22*).

To realize effective couplings across such a broad range of reactants, many catalysts were necessary. Buchwald- and Beller-type phosphine ligands were identified because (i) they share a similar backbone structure, and (ii) more than 50 such ligands are commercially available and were developed to address the broad scope of C–N couplings (*3*, *28*). To capture that range, 20 ligands were selected to represent the crucial ligand dimension of the reaction space with a combination of algorithmic selection and expert knowledge (see SM for details). In the interest of practical applicability, models were trained to make predictions for solvents and bases used at the bench. Thus, two inorganic bases—potassium carbonate and sodium *tert*-butoxide—and one organic base—1,8-diazabicyclo[5.4.0]undec-7-ene (DBU)—were selected. Finally, 1,4-dioxane, toluene, and *tert*-amyl alcohol were selected as representative solvents because they broadly represent cyclic ethers, arenes, and alcoholic solvents regularly used in B-H couplings (see SM for selection of catalyst and conditions). Thus, the reaction space for this study—with nucleophile, electrophile, ligand, solvent, and base dimensions—

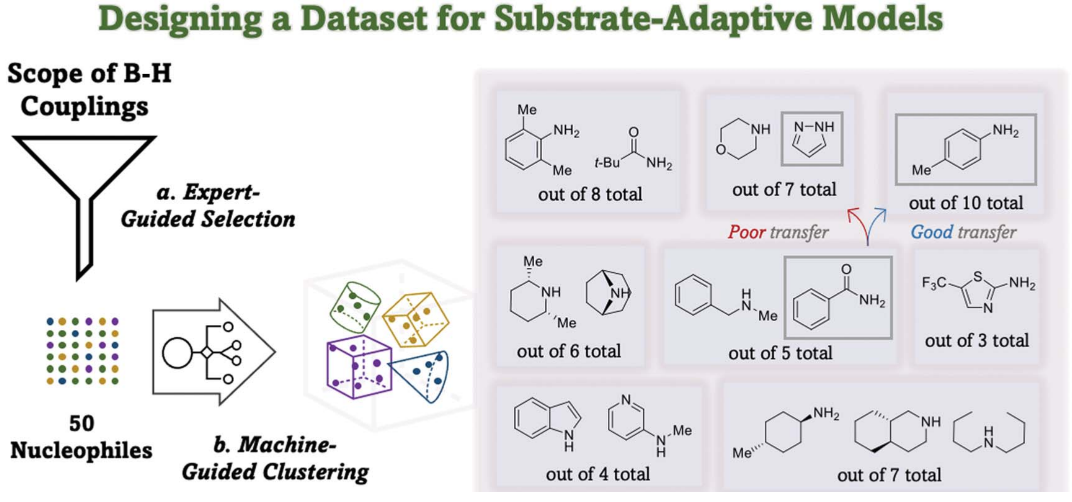

Fig. 3. Representative scope of nitrogen nucleophiles for the B-H coupling reaction used in this work and comparison to other validated ML studies on B-H couplings. The process for selection is depicted: (a) a representative scope was curated, and then (b) an algorithmic method clustered them using the new chemical descriptors developed in this work. The structures shown are some exemplars from the eight clusters that were identified. Me, methyl; *t*-Bu, *tert*-butyl.

contains 450,000 possible reactions from 180 conditions (3 bases times 3 solvents times 20 ligands) and 2500 reactant pairs (50 amines times 50 bromides). As mentioned, this space already represents an intractable number of experiments for combinatorial experimentation.

The first experiments were designed to represent the reaction space broadly with 23 different algorithmically selected reactant pairs and a systematically varied set of conditions for each (see SM). Extensive experimental development identified reproducible conditions on the 0.5-mmol scale in a 24-tube parallel reactor. Twenty-four conditions were evaluated for each reactant pair out of the 180 possible (20 catalysts times 3 solvents times 3 bases). This approach balanced the need to rapidly explore new reactant pairings (i.e., slices from Fig. 2) with the need to explore enough conditions to learn that slice of the reaction space (i.e., points on each slice from Fig. 2). The data showed 63% of experiments with 0% yield, and 82% with less than 20% yield. The paucity of hits from which models needed to learn was a problem. To generate a higher fraction of hits in the data, a new strategy was needed for evaluating the condition component of the B-H reaction space.

To increase the number of positive hits, ML models were trained to recognize zero and nonzero yield patterns. Deep feed-forward neural networks were trained to classify reactions as zero- or nonzero-yielding using that first dataset and showed an average accuracy of 87%. Although unable to distinguish between a 1%-yielding and a 99%-yielding reaction, such a classifier could still increase the number of nonzero-yielding reactions that are run. Eighteen reactant pairs from the first 23 gave significantly more nonzero data (the rest were hypothesized to be chemically challenging using expert knowledge). The initial results from those 18 reactant pairs showed 56% zero-yielding reactions, and only 12% with a >80% yield. Twenty-four new conditions were selected from the 156 remaining conditions (180 total, with 24 already evaluated) using the classifier to predict nonzero-yielding conditions. The new results from those 24 classifier-selected conditions contained just 22% zero-yielding reactions, and 29% of the data showed >80% yield across all 18 reactant pairs (see SM for details). Thus, even a limited binary classifier could be used to improve the yield distribution of new data.

Those first models allowed us to connect steps 1 to 4 in a new workflow depicted in Fig. 4. This workflow begins and ends with the experimentalist. Thus, (i) a new reactant pair is selected by the experimentalist, (ii) the tool calculates the corresponding chemical descriptors, (iii) the tool then uses models to predict the yield of all 180 conditions, and (iv) the experimentalist can decide which conditions to evaluate on the basis of both the predictions and their expert knowledge. The second half of the workflow, steps 5 to 8, describes the process of domain expansion by (v) including new data, (vi) retraining models with that new data, (vii) testing those models in control experiments, and (viii) having the experimentalist evaluate model performance. At this point in the cycle, the experimentalist again intervenes with expert knowledge and their evaluation of model performance to select the next reactant pair to target.

To build the desired dataset, the experimentalist directs what the model learns by their selection of the next reactant pairs. This process relies on defined reaction subspaces to target for domain expansion (vide supra). Figure 5A depicts the map of reaction subspaces used to select the next reactant pairs, which was constructed from the output of the process in Fig. 3. The B-H reactant space map is an array with amines organized by cluster on the vertical axis and bromides organized in the same manner on the horizontal axis. Each small square represents a coupling between a specific bromide and a specific amine; each rectangular subdomain therefore represents a set of couplings between a cluster of amines and a cluster of bromides. The varying size of the clusters is a consequence of the reactants selected in Fig. 3A. The entire array is organized through the lens of new chemical descriptors. The new descriptors combine a radial distribution of atoms within each reactant around the reaction center with various atomic properties relevant to reactivity, producing what we call a radial distribution function (see SM). The initial algorithmically selected reactant pairings that were used to build the first classifier model are shown in light green. The workflow in Fig. 4 was then used, and reactant selections were made in an approach similar to the transfer learning conducted by Zimmerman and co-workers but on a larger scale; many such transfer steps were made to expand the domain of models (by moving to a new subspace). Then, for each new subspace, multiple reactant pairs were evaluated to enable learning of substrate-level trends, making models robust in that subspace. The results of those many selections are shown in dark green.

The dataset can be described as a network, and the goal of this work is to explore enough connections (reactant pairings) to make inferences about the missing connections that are possible. To visualize this, a structured chord diagram is depicted in Fig. 5B, showing amine and bromide nodes on either side with edges connecting reactants that were coupled in the dataset. Similar to the map in Fig. 5A, reactants are organized by cluster on either side, and the edge bundling in the figure illustrates that the data are distributed across exemplars from each cluster. The first visualization emphasizes that 121 out of the possible 2500 combinations of substrates were evaluated in this work. The sparsity of these data is a feature, not a flaw; identifying 121 out of 2500 couplings and 24 out of 180 conditions to evaluate reduced the experimental burden from 450,000 experiments (180 conditions and 2500 substrate pairs) to about 3300 experiments. The diversity of the dataset is best represented by the reactants evaluated (see SM for the full list). Using randomly partitioned data, models achieved a mean absolute error (MAE) of 9% in an external test set. However, the problem outlined in Fig. 1 is best represented by a test set of out-of-sample reactant pairs, which is a more difficult test.

After evaluating multiple couplings in each subspace, models predicted reactivity trends (but not necessarily exact yields) for new reactants

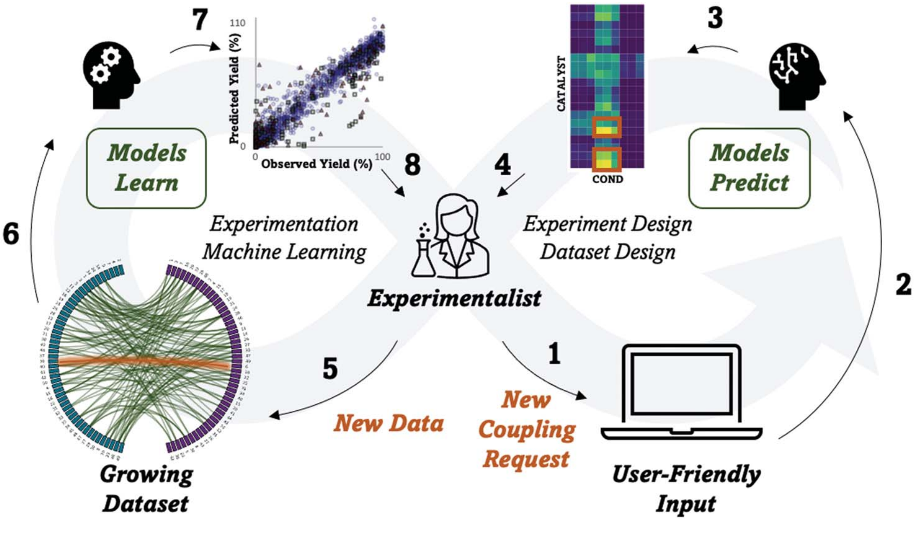

Fig. 4. New, experimentalist-driven, active-learning workflow for exploration of reaction space.

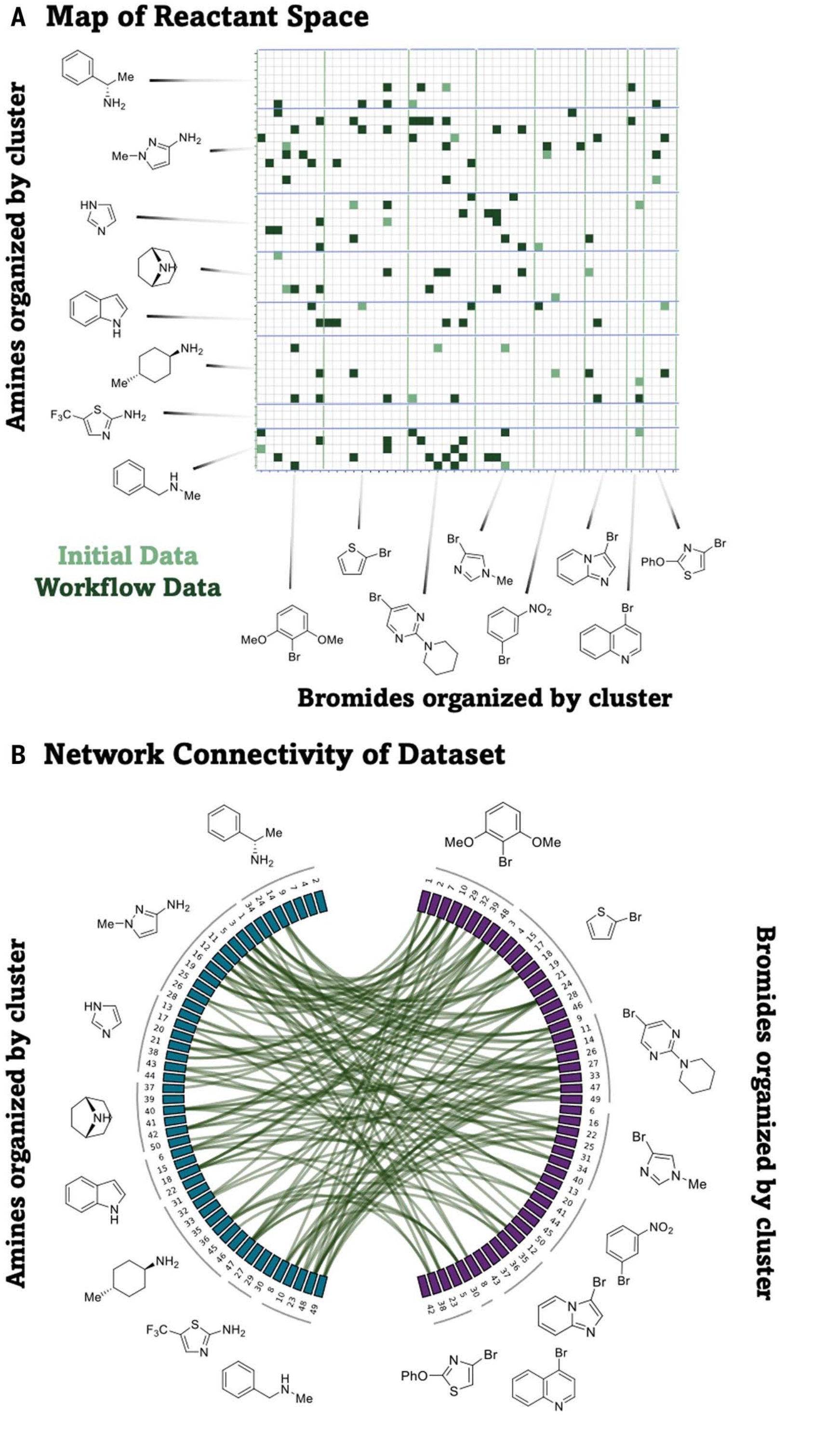
A
Map of Reactant Space
Amines organized by cluster
Initial Data
Workflow Data
Bromides organized by cluster
B
Network Connectivity of Dataset
Amines organized by cluster
Bromides organized by cluster

Fig. 5. Visualizations of B-H reaction space reactant component. (**A**) Dataset distributed across reactant clusters in a two-dimensional grid. (**B**) A structured chord diagram showing the network connectivity of the reactant pairs sampled in the dataset. One exemplar is shown from each cluster.

with the 180 conditions. For example, the highest-yielding predictions may be about 50% and the lowest about 0%, whereas the experimental data might range from 85 to 0%. However, the linear correlation of predicted and observed yields would be high, with a slope in the range of 0.5 to 0.75. Once models could predict reactivity trends, much more data was required to achieve accurate yields. Ultimately, the trade-off between accurate predictions and solving the problem outlined in Fig. 1 required that we accept this as an experimental limitation and continue domain expansion before reaching yield accuracy in every subspace. Although models sometimes showed limited accuracy, they were still able to address the key problem because they could identify the yield trends for the various reaction conditions out of the 180 available. Accepting this limitation, a pragmatic, ordinal ranking of predictions was used. Predictions were divided into quartiles (25% of the total predicted range), and these quartiles were labeled as “high,” “moderate,” “low,” and “poor.” In practice, multiclass classification models (instead of regression models with this ordinal ranking of predictions) were slightly less performant. To test this system as a tool, models trained on the resulting dataset were experimentally evaluated with new B-H couplings.

Experimental validation

To evaluate the performance of these models, the tool was tested in a typical use case: New couplings were selected in which one or both reactants were not seen by the model (out-of-sample) and were tested using an experimentally tractable number of conditions (one to five in most cases) with the highest predicted yields. To ensure that this test was rigorous, the number of catalysts evaluated for each coupling was limited to 3 out of the 20 available. This precaution minimized the possibility that a selected catalyst would work well by chance instead of models learning meaningful chemical structure and reactivity relationships. When the experiments were carried out, predictions were ranked into quartiles, and, as they were available, a few conditions in the top quartile(s) were evaluated. A complementary set of conditions predicted to give low yields were also evaluated to test whether the models were correctly learning reactivity trends as described above. Together, conditions with high and low predicted yields bookend a representative experimental yield range for each coupling (see SM for details). Figure 6 presents the results of experimental validation, including (i) a chemical structure indicating out-of-sample reactants in red, in-sample reactants in blue, and the coupling locus highlighted by the bold bond; (ii) the number of experiments evaluating different conditions; (iii) the subset of conditions with a high predicted yield (hits) that were tested, represented as circles; (iv) an indication of success by the coloring of the circle (green, experimental yield in the top quartile), failure (red, experimental yield not in the top quartile), or a special case (gray); (v) the highest isolated yield obtained from those predicted hits; and (vi) a heatmap showing all 180 predicted yields for use as a visual fingerprint of the response of the models to changing reactant structure.

To interpret experimental validation results, success and failure must be defined. The

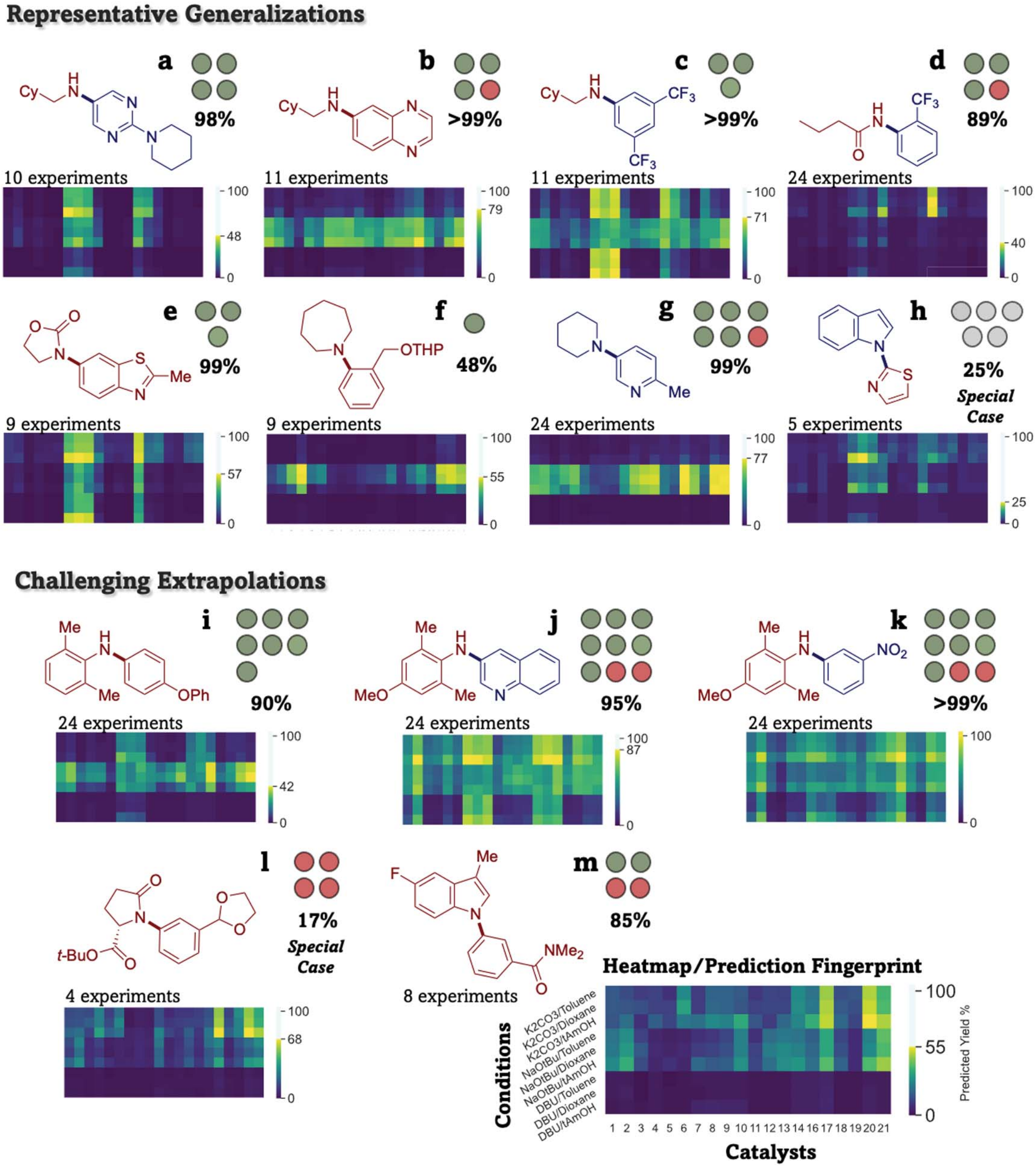

Fig. 6. Experimental validation of substrate-adaptive models as condition recommenders. Out-of-sample (red) and in-sample (blue) reactant fragments are indicated for all products. Circular iconography indicates the number of predicted hits that were tested; green indicates a success, and red indicates a failure. Gray color is explained in the text. The highest isolated yield is indicated below the circular icons. The prediction heatmaps share the legend shown for compound **m**. The scale indicates the highest yield prediction for each coupling (middle number) and color scale. Cy, cyclohexyl; Ph, phenyl; THP, tetrahydropyranyl.

hypothesis behind the experimental validation is that models can correctly recommend synthetically useful hits for couplings with one or more new reactants by predicting reactivity trends. A successful experiment was defined as one in which the predicted hit furnished a yield in the top 25% of those observed for that coupling (a top quartile yield). For practical reasons, not all 2340 experiments (13 reactant pairs in Fig. 6 times 180 conditions) were evaluated; instead, 187 experiments (both hits and zero-yield predictions) were performed across all 13 reactant pairs. Note that although this experimental design does not test whether all high-yielding conditions were identified, in 10 out of 13 cases, the highest observed yields were >85%, and all three of the remaining cases were expected to furnish a lower yield (vide infra). The full details of predicted and observed yields, correlations between them, and common error metrics are provided in the SM.

Our claim that yield trends are sufficient to gauge performance is illustrated by the first five validation experiments in Fig. 6. The couplings to form the products **a** to **e** showed very good predicted-observed correlations [0.93, 0.89, 0.91, 0.77, and 0.75 coefficient of determination (R^2); see SM]. Poor residual errors [>25% MAE, >34% root-mean-square error (RMSE)] were observed for **a**, **c**, and **e** because the highest yield predictions of only 33, 49, and 57% led to the highest observed yields of 98 to 99%. Moderate-to-good residual errors were observed for **b** (14% MAE, 18% RMSE) and **d** (9% MAE, 18% RMSE). Couplings **a** to **c** share the same nucleophile, and the best conditions changed among those couplings, demonstrating that the models adapted to electrophile structure.

Next, the coupling of azepane with a protected 2-bromobenzyl alcohol derivative (product **f**) used two out-of-sample reactants. The results showed very good accuracy, with 4% MAE and 7% RMSE and a good correlation of 0.80 R^2 between predicted and observed yields. By contrast, the coupling of piperidine with 3-bromo-5-methylpyridine (product **g**) involves two in-sample reactants (the only example here of coupling two in-sample reactants) but the combination was never tested in the dataset. The results were satisfying, with an accuracy of 12% MAE and 16% RMSE and a correlation of 0.86 R^2. Comparing prediction fingerprints of **f** and **g** provides some insight into the models' ability to adapt predictions to substrate structures: For the more sterically hindered bromide in coupling **f**, the less bulky diarylphosphino-substituted ligand 4 (CPhos hybrid) is predicted to be the best. For the unhindered bromide in coupling **g**, bulkier tricyclohexylphosphine-substituted ligands 17, 20, and 21 (CPhos, RuPhos, and SPhos) are predicted to be superior.

In the coupling of indole and 2-bromothiazole (product **h**), the bromide was new to the model. This simple coupling is reported using copper catalysis only (*29–32*) and is a difficult coupling for palladium catalysis. As a result, this experiment tests the ability of the model to identify couplings that will not work. Indeed, the models predicted a maximum 23% yield, with mostly near- and zero-yield predictions. Experimental evaluation of the five-highest predictions showed that one condition provided a 25% yield, and the rest furnished no product. Within the confines of this experimental design, it appears that the model successfully identified a challenging coupling; however, to fully substantiate that claim, all 180 conditions would need to be evaluated. An experimentalist with those predictions should pursue another synthetic method if it is available.

By contrast, the coupling of *tert*-butyl pyroglutamate with 2-(3-bromophenyl)-1,3-dioxolane (both of which are out-of-sample, product **l**) represents an extrapolation because the pyroglutamate (i.e., contains an ester group) is substantially different from anything in the dataset. As was seen with product **h**, the four top predictions ranging from 59 to 68% yield were evaluated, yet the highest observed yield was 17%, a clear failure on the part of the models to recognize a challenging, extrapolative coupling. We observed that the nucleophile was converted to the carboxylate in situ by ^{1}H quantitative nuclear magnetic resonance (qNMR) and hypothesized that this by-product prevented coupling by coordinating Pd. Those complications are excellent examples of real challenges in generalizing this reaction that prevent the tool from accurately identifying a low-yielding reaction because it had not learned to predict such complicated behavior.

2,6-Dimethylaniline derivatives represent a class of nucleophile that is new to the models because all other primary (hetero)aromatic amines included in the dataset have no ortho substitution. Twenty-four conditions were evaluated for three separate pairings of 2,6-dimethylanilines with one out-of-sample bromide (4-phenoxybromobenzene, product **i**) and two in-sample bromides (3-bromoquinoline and 3-bromonitrobenzene, products **j** and **k**). The fractions of top-quartile predictions that furnished yields in the top quartile of those observed were 7/7, 7/9, and 7/9, respectively. That **i** to **k** represent extrapolations is more evident in their range of poor to moderate R^2 of 0.70, 0.05, and 0.40, respectively. The observed yield ranges for the best predictions evaluated were 76 to 99%, 83 to 99%, and 77 to 89% yields, respectively. These results suggest that for a subclass of anilines not represented in the dataset, models were still able to identify high-yielding conditions. Thus, a tractable number of experiments (7 to 9 out of the 24 evaluated) were predicted to furnish good yields and seven did for each of the three couplings.

Finally, 5-fluoroskatole was included as a nucleophile for its similarity to indole but its inability to undergo C(3) arylation (product **m**). While acquiring the dataset, we regularly observed the C(3) arylation products—often as dominant species—effectively forcing models to predict C–N couplings that are in competition with C(3) aryl coupling. As chemists with an understanding of reaction mechanisms, it is intuitive that a minor structural perturbation such as adding a methyl group to the C(3) position of indole will prevent competitive C(3)-arylation. The models being evaluated do not learn mechanisms or the implications of such a minor structural change (and correspondingly minor change in the descriptors) as adding a methyl group to that position of an indole. The experimental validation results show a poor correlation between predicted and observed yields and stochastic errors of predicting zero- and nonzero-yielding conditions, suggesting a different observed reactivity pattern than what the models had learned. Despite this, out of the four conditions with the highest predicted yield (40 to 55%), two provided good yields of 84 and 85%. This example illustrates that the tool can still be useful, even on a challenging coupling that represents a substantial mechanistic extrapolation.

Taken together, the results of the validation experiments demonstrate that the performance of the model exists on a gradient. For new reactants from a reaction subspace that is well represented in the dataset, predictions are robust (i.e., **a** to **c** and **g** to **f**). For cases in which reactants represent structural permutations from those in the dataset (i.e., **d** and **e**), the models correctly learned reactivity trends and could predict hits. For new types of structures that may have different reactivity patterns than those in the dataset, performance ranged from moderate (**i** to **k**) to poor (**l** and **m**). However, even the lowest model performance demonstrated here provided good yields (depending on the chemical limitations of the coupling in question, e.g., **l**). Most importantly, poor model performance can be rescued by domain expansion of the dataset (see fig. S22). This approach to network exploration using active learning has enabled us to create models that are useful over a broad applicability domain by evaluating only 0.7% of the reaction space. Although we will continue to explore this space, more importantly, we have now created a transportable blueprint for domain expansion of a dataset.

Outlook

The dataset described herein comprises >120 reactant pairs that systematically explore a microcosm of B-H coupling space. Models trained on these data simultaneously learned nonlinear reactivity trends for many different classes of reactants. These models could predict the yield of reactions with a mean absolute error of 9% using randomly partitioned data and were performant at reactant generalization, as demonstrated by out-of-sample substrate validation.

Key to achieving this goal was an informatics-guided strategy that reduced the experimental impossibility of exploring a 450,000-member reaction space to an experimentally tractable problem of acquiring a dataset comprising only 3300 experiments. We present both this validated tool for Pd-catalyzed C-N couplings as well as an active-learning workflow, which, unlike prior work, was used to build an expansive dataset for the chemical community. The chemistry community can engage with this work on four different levels. An experimentalist with no interest in ML can use the snapshot of the tool presented in this work without expertise in ML or programming and expect performance similar to experimental validation. We invite any practitioner with an interest in ML to take the tool and resume the workflow in Fig. 4, honing the tool to new reactant domains of interest or steadily improving prediction accuracy on existing dataset domains. Furthermore, we invite any practitioner with expertise in ML to use the new active-learning framework on other important reactions with expansive multireactant spaces. Finally, we offer this dataset for focused development on modeling noncombinatorial, diverse datasets, which are rare in the chemistry domain.

REFERENCES AND NOTES

1. A. S. Guram, R. A. Rennels, S. L. Buchwald, *Angew. Chem. Int. Ed.* **34**, 1348–1350 (1995).
2. J. Louie, J. F. Hartwig, *Tetrahedron Lett.* **36**, 3609–3612 (1995).
3. P. Ruiz-Castillo, S. L. Buchwald, *Chem. Rev.* **116**, 12564–12649 (2016).
4. J. F. Hartwig, K. H. Shaughnessy, S. Shekhar, R. A. Green, in vol. 100 of *Organic Reactions*, S. E. Denmark, Ed. (Wiley, 2019), pp. 853–958.
5. R. J. Lundgren, M. Stradiotto, *Chemistry* **18**, 9758–9769 (2012).
6. B. T. Ingoglia, C. C. Wagen, S. L. Buchwald, *Tetrahedron* **75**, 4199–4211 (2019).
7. M. Fitzner *et al*., *Chem. Sci.* **11**, 13085–13093 (2020).
8. D. S. Surry, S. L. Buchwald, *Chem. Sci.* **2**, 27–50 (2011).
9. M. D. Charles, P. Schultz, S. L. Buchwald, *Org. Lett.* **7**, 3965–3968 (2005).
10. D. Maiti, B. P. Fors, J. L. Henderson, Y. Nakamura, S. L. Buchwald, *Chem. Sci.* **2**, 57–68 (2011).
11. E. P. K. Olsen, P. L. Arrechea, S. L. Buchwald, *Angew. Chem. Int. Ed.* **56**, 10569–10572 (2017).

12. A. C. Sather, T. A. Martinot, *Org. Process Res. Dev.* **23**, 1725–1739 (2019).
13. B. J. Shields *et al.*, *Nature* **590**, 89–96 (2021).
14. F. Häse, L. M. Roch, C. Kreisbeck, A. Aspuru-Guzik, *ACS Cent. Sci.* **4**, 1134–1145 (2018).
15. D. T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher, A. G. Doyle, *Science* **360**, 186–190 (2018).
16. K. D. Collins, F. Glorius, *Acc. Chem. Res.* **48**, 619–627 (2015).
17. P. Schwaller, A. C. Vaucher, T. Laino, J.-L. Reymond, *Mach. Learn. Sci. Technol.* **2**, 015016 (2021).
18. T. R. Gimadiev *et al.*, *Mol. Inform.* **40**, e2100119 (2021).
19. J. Li, M. D. Eastgate, *React. Chem. Eng.* **4**, 1595–1607 (2019).
20. M. Fitzner, G. Wuitschik, R. Koller, J.-M. Adam, T. Schindler, *ACS Omega* **8**, 3017–3025 (2023).
21. N. I. Rinehart *et al.*, chemRxiv chemrixiv-2022-hspvw-v2 [Preprint] (2023); https://doi.org/10.26434/chemrxiv-2022-hspwv-v2.
22. E. Shim *et al.*, *Chem. Sci.* **13**, 6655–6668 (2022).
23. N. H. Angello *et al.*, *Science* **378**, 399–405 (2022).
24. A. Buitrago Santanilla *et al.*, *Science* **347**, 49–53 (2015).
25. M. K. Nielsen, D. T. Ahneman, O. Riera, A. G. Doyle, *J. Am. Chem. Soc.* **140**, 5004–5008 (2018).
26. M. Yamaguchi, K. Suzuki, Y. Sato, K. Manabe, *Org. Lett.* **19**, 5388–5391 (2017).
27. S. Basak, S. Dutta, D. Maiti, *Chemistry* **27**, 10533–10557 (2021).
28. D. S. Surry, S. L. Buchwald, *Angew. Chem. Int. Ed.* **47**, 6338–6361 (2008).
29. H. Chen, M. Lei, L. Hu, *Tetrahedron* **70**, 5626–5631 (2014).
30. S. A. Iqbal, K. Yuan, J. Cid, J. Pahl, M. J. Ingleson, *Org. Biomol. Chem.* **19**, 2949–2958 (2021).
31. J.-K. Kwon, J.-H. Lee, T. S. Kim, E. K. Yum, H. J. Park, *Bull. Korean Chem. Soc.* **37**, 1927–1933 (2016).
32. A. Shoberu, C.-K. Li, H.-F. Qian, J.-P. Zou, *Org. Chem. Front.* **8**, 5821–5830 (2021).
33. N. I. Rinehart, Substrate-adaptive C–N coupling yield prediction. Zenodo (2023); https://doi.org/10.5281/zenodo.8185014.

ACKNOWLEDGMENTS

We acknowledge the support services of the NMR, mass spectrometry, and microanalytical laboratories of the University of Illinois at Urbana-Champaign. **Funding:** We acknowledge generous financial support from Hoffmann La-Roche. A.S.S. thanks the National Science Foundation (NSF) (grant no. CHE 1900617) for financial support. This work was also supported in part by the Molecule Maker Lab Institute, an AI Research Institutes program supported by the NSF under grant no. CHE 2019897. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect those of the NSF. **Author contributions:** N.I.R. contributed to the dataset, wrote the main code package, designed new descriptors, tested neural network architectures for modeling, conceptualized the informatics-guided strategy to exploring the reaction space as a network, designed and contributed to the validation experiments, contributed to composing the manuscript, and wrote the Computational and informatics methods, Experimental methods, Workflow development, General references, Experimental validation, and Master list of dataset products sections of the SM; R.K.S. contributed experimentally to the dataset, contributed to the validation experiments, made authentic products for zero-yielding plates, and composed most of the Characterization data, Characterization references, and NMR spectra sections of the SM; J.W. contributed experimental results to the dataset and prepared reference materials; A.F.Z. helped conceptualize the scope of the project and design descriptors as well as initial classification modeling efforts; L.S. prepared palladium complexes and reference materials; A.S.S. shared code that facilitated the calculation and testing of new descriptors; R.B. conceived of and supervised the project, designed the experimental part of the workflow, contributed experimental results to the dataset, and contributed to composing the manuscript; and both S.F. and S.E.D. conceptualized and supervised the project and contributed to composing the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Full experimental procedures, including validation runs, characterization data, experimental apparatus, qHPLC analytical methodology, and copies of ^{1}H, ^{13}C, ^{31}P, and ^{19}F spectra as well as feature engineering, modeling details and model validation, structures of each product made in the dataset, and predictions for each condition with every reactant pair in the dataset can be found in the SM. Code developed in this work is available at Zenodo (*33*), and ongoing development through our laboratory's Github organization will be available at github.com/SEDenmarkLab/Lucid_Somnambulist. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg2114
Materials and Methods
Supplementary Text
Figs. S1 to S22
Tables S1 to S4
Scheme S1
Charts S1 and S2
NMR Spectra
References (*34*–*96*)

Submitted 8 December 2022; resubmitted 26 April 2023
Accepted 1 August 2023
10.1126/science.adg2114

Science
AAAS

A machine-learning tool to predict substrate-adaptive conditions for Pd-catalyzed C–N couplings

N. Ian Rinehart, Rakesh K. Saunthwal, Joël Wellauer, Andrew F. Zahrt, Lukas Schlemper, Alexander S. Shved, Raphael Bigler, Serena Fantasia, and Scott E. Denmark

Science, **381** (6661), .
DOI: 10.1126/science.adg2114

Editor's summary

The palladium-catalyzed coupling of amines with aryl halides is one of the most widely used reactions in pharmaceutical research and manufacturing. Nonetheless, it depends sensitively on the structure of the two coupling partners and therefore often requires a trial-and-error process to identify pertinent optimal conditions. Rinehart *et al*. trained and validated a machine learning model to predict appropriate ligand, solvent, and base for coupling of particular reactant pairs. Ten products were isolated in more than 85% yield under the individualized conditions predicted by the model. —Jake S. Yeston

Science (ISSN) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.