

# Data Science-Driven Discovery of Optimal Conditions and a Condition-Selection Model for the Chan–Lam Coupling of Primary Sulfonamides

Shivaani S. Gandhi, Giselle Z. Brown, Santeri Aikonen,<sup>=</sup> Jordan S. Compton,<sup>=</sup> Paulo Neves,<sup>=</sup> Jesus I. Martinez Alvarado, Iulia I. Strambeanu,\* Kristi A. Leonard,\* and Abigail G. Doyle\*



Cite This: *ACS Catal.* 2025, 15, 2292–2304



Read Online

ACCESS |

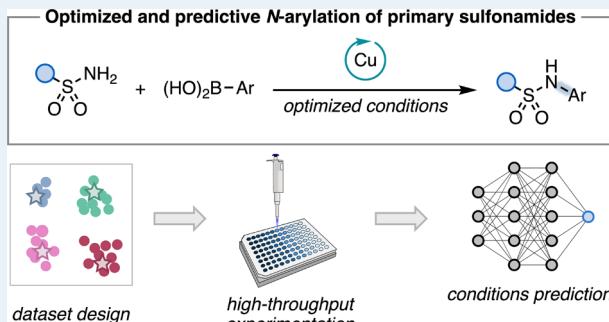
Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** Secondary N-arylsulfonamides are common in pharmaceutical compounds owing to their valuable physicochemical properties. Direct N-arylation of primary sulfonamides presents a modular approach to this scaffold but remains a challenging disconnection for transition metal-catalyzed cross coupling broadly, including the Chan–Lam (CL) coupling of nucleophiles with (hetero)aryl boronic acids. Although the CL coupling reaction typically operates under mild conditions, it is also highly substrate-dependent and prone to overarylation, limiting its generality and predictivity. To address these gaps, we employed data science tools in tandem with high-throughput experimentation to study and model the CL N-arylation of primary sulfonamides. To minimize bias in training set design, we applied unsupervised learning to systematically select a diverse set of primary sulfonamides for high-throughput data collection and modeling, resulting in a novel data set of 3,904 reactions. This workflow enabled us to identify broadly applicable, highly selective conditions for the CL coupling of aliphatic and (hetero)aromatic primary sulfonamides with complex organoboron coupling partners. We also generated a regression model that successfully identifies not only high-yielding conditions for the CL coupling of various sulfonamides but also sulfonamide features that dictate reaction outcome.

**KEYWORDS:** C–N coupling, high-throughput experimentation, machine learning, data science, catalysis



## 1. INTRODUCTION

N-Arylsulfonamides represent an important class of organosulfur compounds, as exemplified by their presence among various FDA-approved drugs spanning multiple disease indications.<sup>1</sup> This motif is prevalent in drug discovery due to its polarity, geometry, hydrogen bonding properties, and tunable NH pK<sub>a</sub> (Figure 1A).<sup>2</sup> Classically, N-arylsulfonamides can be synthesized via addition of anilines to sulfonyl chlorides; however, this approach is not always conducive to synthesis of drug-like molecules due to the water sensitivity of sulfonyl chloride reagents and attenuated nucleophilicity of electron-poor anilines. Furthermore, the potential genotoxicity of both anilines and sulfonyl chlorides has inspired the development of alternate approaches.<sup>3,4</sup> Despite the recent emergence of numerous catalysts and synthetic methods for C–N coupling, transition metal-catalyzed N-arylation of primary sulfonamides remains relatively underdeveloped. To the best of our knowledge, fewer than 40 reports have studied the direct N-arylation of primary sulfonamides, most of which feature high catalyst loadings and/or high temperatures, with relatively limited scope examples.<sup>5–30</sup> This may be, in part, due to the attenuated nucleophilicity of sulfonamides, which

renders them recalcitrant in cross-coupling reactions. Selectivity poses an additional challenge for the coupling of primary sulfonamides, given the possibility that the desired N-arylated product can undergo further reaction to generate the undesired N,N-diarylated product.

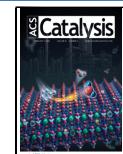
The CL coupling is commonly applauded for its use of inexpensive reagents, mild reaction conditions, greater tolerance of heterocyclic substrates than its Pd-catalyzed counterparts, and orthogonality to traditional cross-coupling reactions of aryl halides. Despite being a cornerstone reaction in organic chemistry, the CL coupling notoriously lacks predictivity and selectivity, limiting its broader applicability for synthesis. In particular, the CL coupling of sulfonamides remains underexplored compared to that of amines (Figure 1B). In fact, of the known examples studying CL arylation of

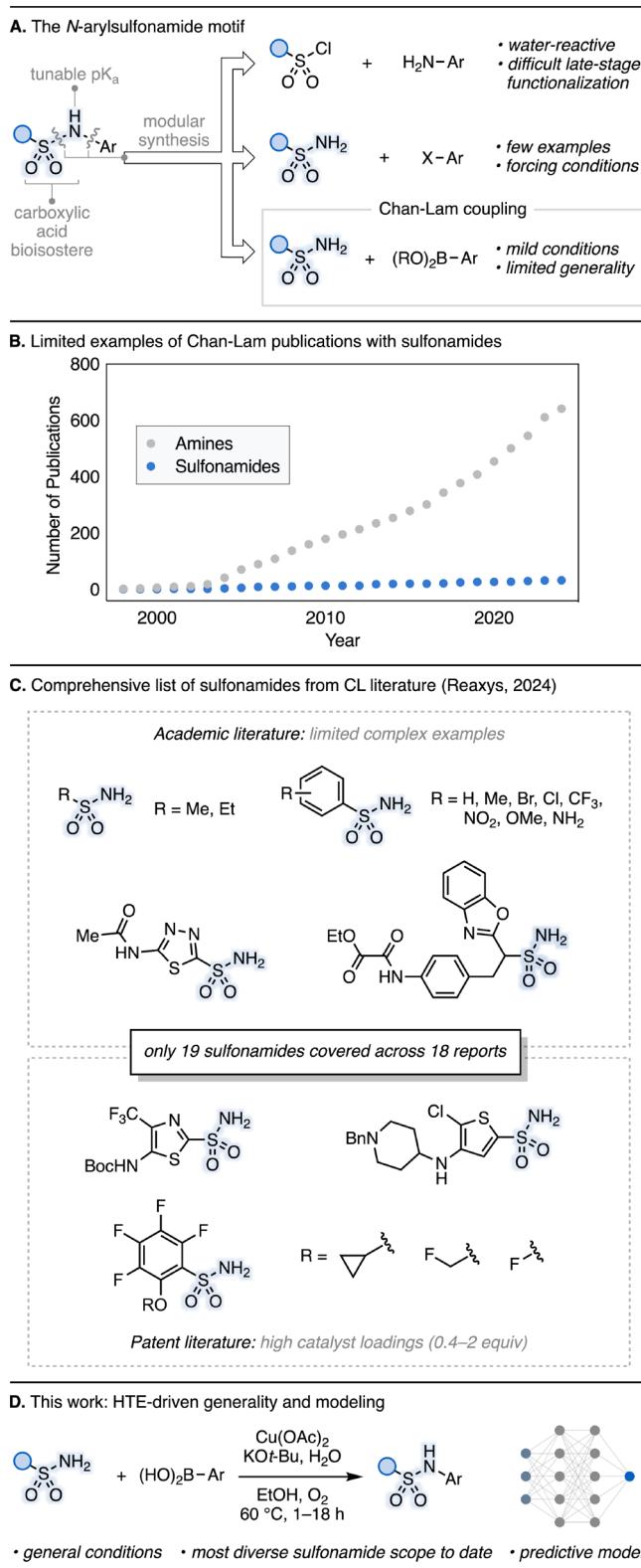
Received: December 23, 2024

Revised: January 6, 2025

Accepted: January 8, 2025

Published: January 24, 2025





**Figure 1.** A) Physicochemical properties and methods for synthesis of the *N*-arylsulfonamide motif. B) Studies of the Chan-Lam (CL) coupling of sulfonamides remain limited (Reaxys, 2024). C) Only 19 sulfonamide examples have been demonstrated in the CL coupling literature across 18 literature and patent reports, which suffer from limited substrate complexity and high catalyst loadings, respectively. D) This work: general conditions, a diverse substrate scope, and predictive model for the *N*-arylation of primary sulfonamides.

primary sulfonamides, the demonstrated scope remains limited to just 19 examples across 18 literature and patent reports (Figure 1C).<sup>23–40</sup> The patent literature features several complex examples, but high or stoichiometric copper loading is often required.

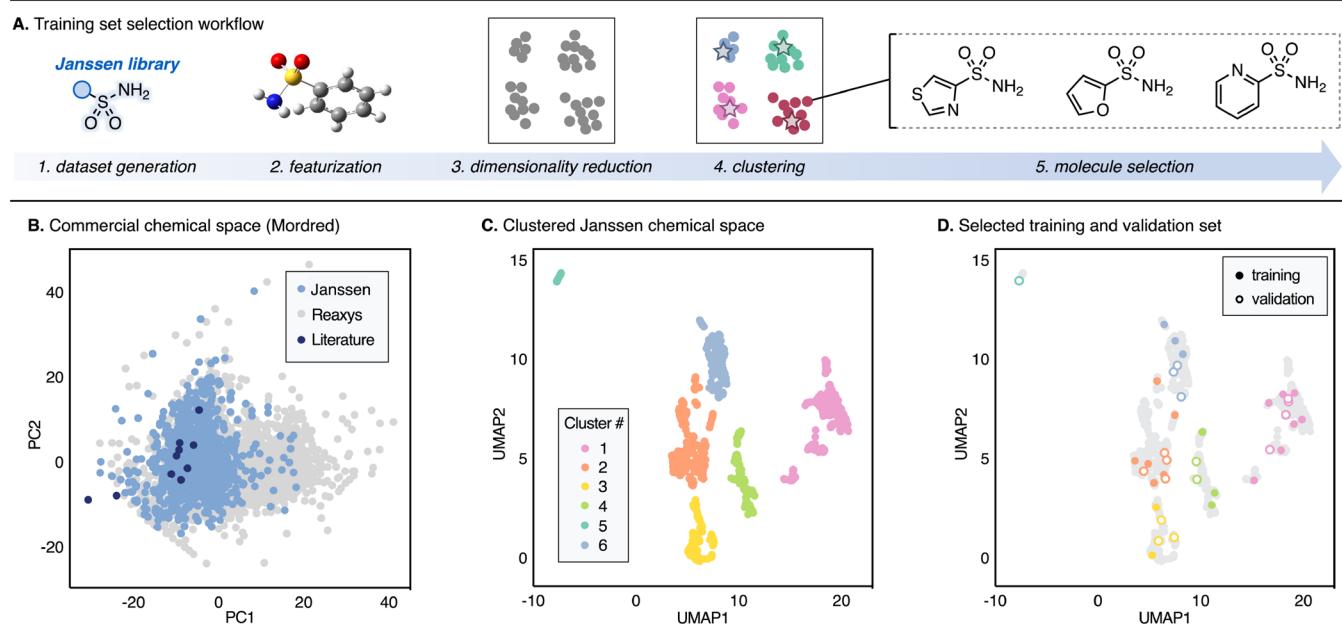
The mechanistic information gained from key experimental<sup>41</sup> and computational<sup>42</sup> studies, while insightful, has highlighted the highly substrate-dependent nature of the reaction: the amine substrate may bind to the catalyst at every step of the catalytic cycle, and even participates in the catalytic turnover step. While significant development has allowed for extension of the CL coupling paradigm to other C–X bond-forming reactions, even substrates within the same class may require highly varied reaction conditions.<sup>43</sup> This characteristic is particularly detrimental for discovery applications in medicinal chemistry, wherein a single set of reaction conditions that can guarantee the generation of “enough” product is highly desirable. It would, therefore, be of great utility to not only identify general reaction conditions for the selective monoarylation of primary sulfonamides, but also to predict conditions suited to unseen substrate pairs.

Machine learning (ML) has emerged as a powerful tool in organic chemistry for optimizing synthetic processes, elucidating reaction mechanisms, and predicting reactivity.<sup>44,45</sup> The ability of ML to analyze obscure patterns and generate predictive models can prove valuable for gaining a deeper understanding of chemical reactivity. Given this, we questioned whether use of ML could address the lack of generality for the CL coupling of primary sulfonamides. Current work in reaction outcome prediction typically features mechanistically well-defined (often catalyst/ligand-controlled) reactions, for which it is straightforward to identify features that govern reactivity. There are few examples applying ML toward poorly understood reactions that may be challenging to study via traditional organometallic and/or physical organic studies, likely due to the large quantities of data or expensive computations needed to elucidate reactivity patterns.<sup>46,47</sup> We sought to leverage an unsupervised learning approach to select a representative subset of substrates from a broader chemical space,<sup>48–51</sup> thereby decreasing the amount of data required to study such systems. We hypothesized that a systematic selection strategy, coupled with high-throughput experimentation (HTE), could allow us to gain a deeper understanding of the CL coupling.

Herein, we report the use of unsupervised learning to select a set of 44 representative primary sulfonamides, which we utilize to generate an HTE data set of 3,904 unique CL coupling reactions. Ultimately, the HTE data enabled the discovery of broadly applicable reaction conditions for the monoselective *N*-arylation of primary sulfonamides, as well as construction of a neural network ML model to predict optimal conditions for both in-domain and out-of-sample substrates (Figure 1D).

## 2. METHODS AND RESULTS

**2.1. Systematic Dataset Design.** In pursuit of general reaction conditions, we sought to select a diverse subset of the Janssen internal library of pharmaceutically relevant primary sulfonamides using our previously developed workflow (Figure 2A). To investigate whether the Janssen library spanned all relevant primary sulfonamide chemical space, we obtained a list of 3,547 commercially available substrates from Reaxys for comparison. To visualize the chemical space in two



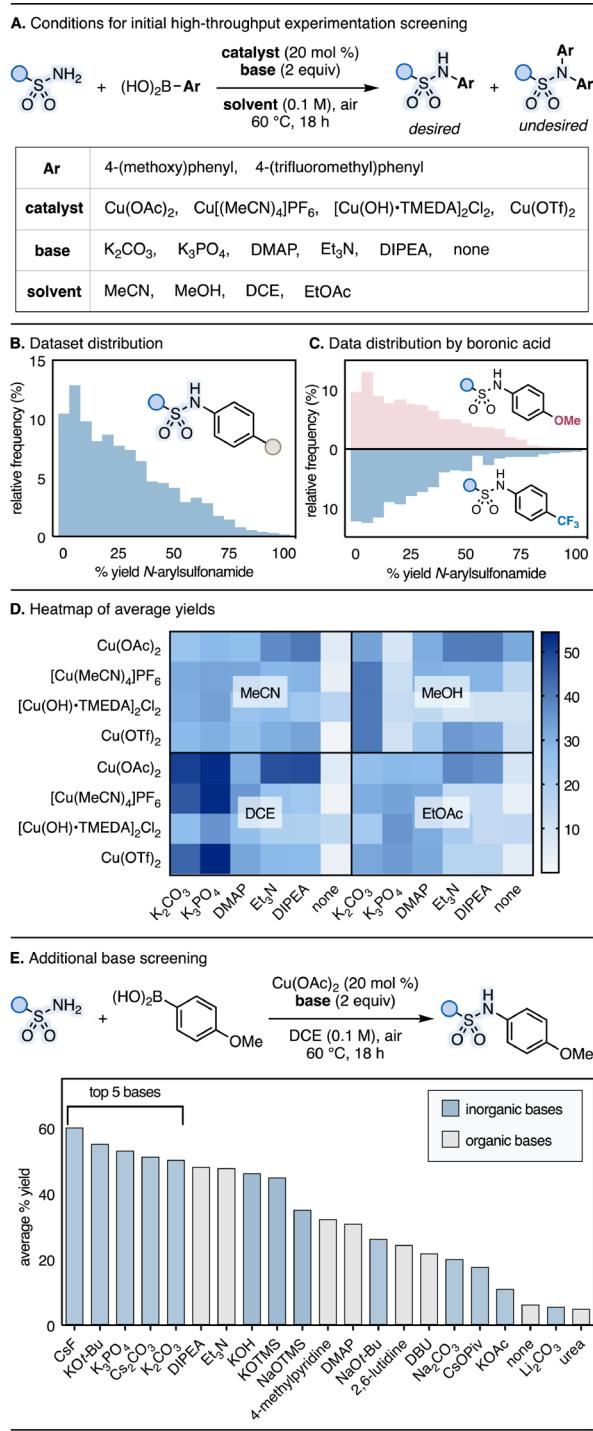
**Figure 2.** A) Summary of selection workflow. B) Chemical space plot showing coverage of the Janssen library (explained variance PC1 = 16%, PC2 = 9%). C) Clustered Janssen chemical space with DFT features. D) Selected data set superimposed on the Janssen space.

dimensions, we undertook featurization via Mordred, an open-source descriptor generator.<sup>53</sup> Reduction of the 1,182 computed features into two dimensions using Principal Component Analysis (PCA) resulted in the chemical space plot shown in Figure 2B. Interestingly, the sulfonamide substrates from the library that have previously been shown in the CL literature—two alkyl and ten simple aryl sulfonamides—are localized in the bottom left quadrant of the chemical space plot (Figure 2B, dark blue; see SI for details).<sup>54</sup> This lack of substrate diversity may contribute to the limited generalizability of known conditions for the CL coupling of sulfonamides. We hypothesized that using HTE to investigate reaction conditions across a diverse and uniformly distributed set of sulfonamides would help us achieve generality in the coupling of this challenging substrate class. Superimposing the Janssen library on the Reaxys chemical space (Figure 2B) revealed generally good coverage of the chemical space. The right quadrant of the plot, where we observe sparser coverage by the Janssen library, largely contains Celecoxib derivatives, from which we planned to select one representative example. Given that the Janssen library would enable quick access and ensure relative substrate diversity, we proceeded with our selection workflow.

Anticipating that DFT-based features would be most appropriate to capture the underlying physical organic properties relevant to reaction outcome, we computed these features for the Janssen data set,<sup>55</sup> then applied Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction<sup>56</sup> to visualize the curated, pharmaceutically relevant chemical space (Figure 2C). Then, we applied a hierarchical clustering algorithm to divide the chemical space into 6 distinct clusters, selected using silhouette score. We selected representatives approximately proportional to the size of each cluster based on commercial availability and ease of access (Figure 2D). Looking toward eventual validation of our modeling efforts, we also selected 22 additional sulfonamides to serve as out-of-sample tests (17 from the Janssen library,

and 5 from the Reaxys data set). This data-science driven approach ultimately afforded the largest and most diverse primary sulfonamide data set to study the CL coupling to date. Our curated scope includes several heterocyclic sulfonamides, as well as multisubstituted substrates for which steric and electronic influences on reactivity may not be straightforward to predict *a priori* (vide infra, Figure 5).

**2.2. HTE Data Collection and Analysis.** With a diverse training set in hand, we turned our attention to HTE data set design. Given the often-contradictory information about the best conditions to use for the CL coupling,<sup>57</sup> we examined a variety of reaction conditions (2 boronic acids, 4 copper catalysts, 5 bases and no base, and 4 solvents), amounting to 3,072 unique experiments (Figure 3A). It is noteworthy that authentic product synthesis, which is a prerequisite for building quantitative UPLC/MS assays and analyzing HTE data, is a significant but often underappreciated challenge of collecting such a designer data set; this presents a formidable obstacle to screening the substrate dimension, as synthesis is required for every unique product. While we expected to observe the desired *N*-aryl sulfonamide as the major product under the chosen reaction conditions, we also anticipated formation of the undesired *N,N*-diaryl sulfonamide. Although the conditions developed over the course of this work would later prove effective for most sulfonamides in our training and validation sets, authentic product synthesis was required at the outset, when we had not yet identified our optimal conditions. Prior to HTE data collection, we modified a protocol reported by Watson and co-workers for the CL arylation of secondary sulfonamides,<sup>58</sup> enabling the synthesis of both the *N*-aryl and *N,N*-diaryl sulfonamide products from the same reaction. While not practical for preparation of large quantities owing to the low selectivity often observed, this protocol was quite efficient for authentic product preparation. Low yields (<20%) were observed in many cases, alluding to the challenge of generality and prompting us to use amine addition into sulfonyl chlorides to prepare 3 especially low-yielding products.



**Figure 3.** A) Conditions evaluated in the initial HTE campaign. B) Histogram showing relative frequency (%) of yields in the HTE data set. C) Histogram showing relative frequency (%) of yields in the HTE data set per boronic acid. D) Heatmap of average yield per reaction condition. E) Average yield per base from follow-up HTE screening.

Upon synthesis of the 64 authentic products and construction of UPLC/MS assays, we collected our training data set in triplicate.

Approximately 28% of the reactions delivered <10% yield of the desired product, with the remaining 72% having yields distributed across the full range from 10–100% (Figure 3B). While most reactions were selective for the desired

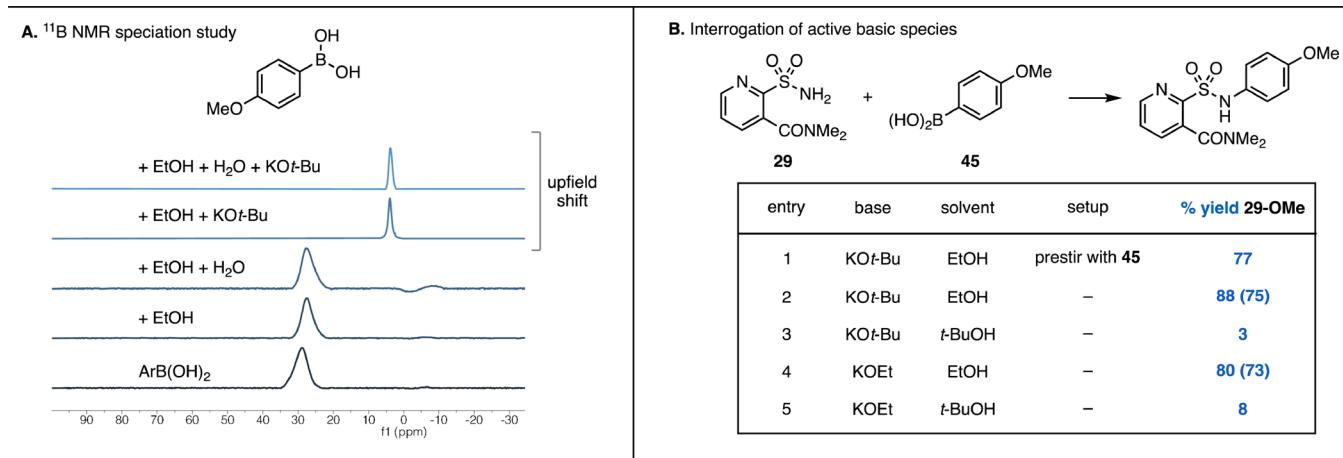
monoarylated product, 10% of the reactions afforded >10% yield of the undesired diarylated product. We also found that the electronic properties of the boronic acid did not seem to impact trends in yield (Figure 3C), DCE proved to be the optimal solvent for the desired coupling reaction (Figure 3D), and the base identity played a significant role in the success of the coupling reaction (Figure 3D). Conditions with DCE and inorganic bases ( $\text{K}_2\text{CO}_3$  or  $\text{K}_3\text{PO}_4$ ) gave high yields of the desired monoarylated product across all 22 sulfonamide substrates. This result is consistent with the few protocols that have been specifically optimized for the CL coupling of sulfonamides, which tend to feature inorganic bases,<sup>23,25,27–29</sup> and in contrast to the majority of CL protocols, which typically employ organic bases such as triethylamine, diisopropylethylamine, and pyridine.<sup>57</sup>

Given the large influence of the base identity on the reaction outcome, we chose to study an additional 15 bases under the optimal conditions identified in our initial study, constituting an additional 480 data points. Visualizing the average yield per base allowed us to evaluate the generality of each reaction condition (Figure 3E). Interestingly, we identified several bases, including CsF and KOt-Bu, that enabled successful and general CL coupling of primary sulfonamides. Previous reports have demonstrated that counterion identity can have a profound effect on the oxidation potential of copper complexes.<sup>59,60</sup> As such, it is possible that a judiciously chosen inorganic base could better facilitate the necessary redox events at the copper center.

In addition to the results of our base screen, we were particularly interested in studying substrate-dependent reactivity trends. Evaluation of the average yield per sulfonamide allowed us to identify two trends: substrates bearing basic heterocyclic motifs such as pyridines and pyrazoles, as well as those bearing electron-withdrawing functional groups, performed better on average than their non-heterocyclic and electron-rich counterparts. These observations eventually informed our featurization and modeling efforts (*vide infra*).

**2.3. Synthetic Application.** Our HTE campaign provided several leads for general reaction conditions, which we next sought to translate to preparative scale. We found that some reoptimization was necessary to ensure mass transfer and reproducibility, particularly due to the heterogeneity of the reaction. While the HTE screens were run under ambient atmosphere, bench-scale reactions were set up under  $\text{O}_2$  atmosphere (see SI for optimization screens).<sup>61</sup> Further screening revealed that while KOt-Bu and CsF both afford high yields of the desired monoarylated product, the former facilitates higher selectivity, particularly when coupled with ethanol as the reaction solvent. We also found that addition of 10 equiv of  $\text{H}_2\text{O}$  affords not only higher yields, but also better reproducibility. The unique combination of KOt-Bu, EtOH, and  $\text{H}_2\text{O}$  was particularly intriguing to us. These results further highlight the enigmatic nature of the CL coupling: although alcohol nucleophiles are typically competent in the reaction,<sup>62</sup> competitive *O*-arylation was not observed under our reaction conditions.

We propose several potential roles of the base, which could account for the unique effectiveness of our optimized conditions. First, our screening revealed that sulfonamide reactivity is dependent on the substrate's relative acidity and base identity. Given the observation in our HTE that CsF and KOt-Bu are both capable of providing generally high yields of the desired product, we wondered whether these bases serve to



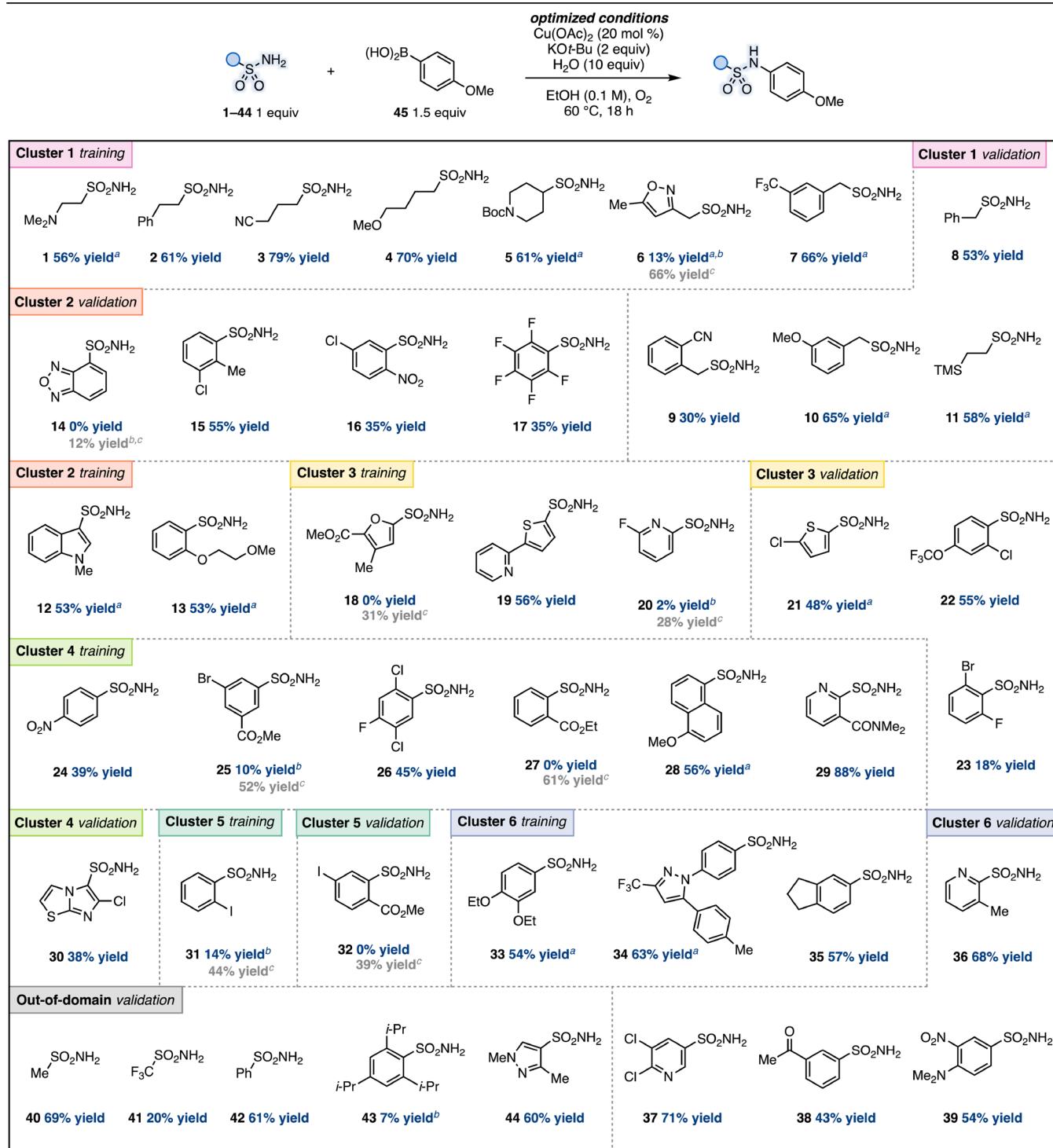
**Figure 4.** A) An upfield shift was observed upon reaction between the boronic acid and  $\text{KOt-Bu}$ . B) Experiments varying the base and solvent identity reveal ethoxide as the active base. Parentheses indicate yield for the reaction in the absence of water.

activate the boronic acid for a more facile transmetalation, in analogy to reports with palladium.<sup>63,64</sup> To interrogate whether this may be operative, we monitored the  $^{11}\text{B}$  NMR shift of 4-methoxyphenylboronic acid in the presence of various reaction components (Figure 4A).<sup>65</sup> In the presence of  $\text{KOt-Bu}$ , an upfield shift of  $\sim 25$  ppm was observed, consistent with formation of a tetracoordinate alkoxyboronate species. Furthermore, performing and subjecting this species to the reaction affords the same yield of the *N*-arylsulfonamide product, supporting its intermediacy in the reaction (Figure 4B, Entry 1). By contrast, other bases either failed to induce ( $\text{Et}_3\text{N}$ ,  $\text{K}_2\text{CO}_3$ ) or only partially induced ( $\text{CsF}$ ) an upfield shift (see SI for details). Given that the reaction mixture contains superstoichiometric quantities of  $\text{H}_2\text{O}$ , EtOH, and  $\text{KOt-Bu}$ , we were curious to investigate the active basic species. Experiments varying the identity of base and solvent revealed that ethoxide generated *in situ* (Figure 4B, entries 2, 4) is the active species, rather than *tert*-butoxide (Figure 4B, entries 3, 5). While exclusion of water still resulted in product formation, slight improvements to the yield are observed on inclusion of water (Figure 4B, parentheses).

With optimized conditions in hand, we set out to evaluate the scope of the coupling reaction using our diverse sulfonamide training and validation sets. Gratifyingly, our conditions afforded good to high yields and selectivities across nearly the entire substrate scope (Figure 5). We were especially pleased to find that heterocyclic sulfonamides underwent successful coupling (5, 6, 12, 18, 19, 21, 29, 30, 33, 36, 38, 44), given that such motifs can function as catalyst poisons in other transition metal-catalyzed cross couplings. We also found that while steric bulk can hinder reactivity in extreme cases (23, 43), *ortho*-substitution was well tolerated under the reaction conditions (13, 15–17, 22, 26, 29, 36). Despite the broad scope of our conditions, couplings of substrates prone to side reactions such as ester hydrolysis (18, 25, 27, 32),  $\text{S}_{\text{N}}\text{Ar}$  (20), and competitive Ullmann coupling (31, 32) were relatively low-yielding. However, we found that simply reverting to the original conditions identified through our HTE campaign enabled the successful coupling of these substrates, albeit with increased formation of the undesired *N,N*-diarylsulfonamide product (Figure 5, gray yields). A notable exception was 14, which formed only trace product, likely due to its propensity to undergo metal-mediated N–O bond cleavage.

To better contextualize our optimized conditions relative to prior work, we compared the yield of the CL coupling of an alkyl (4) and aryl (12) sulfonamide across a suite of previously published conditions (Figure 6A).<sup>23,25,26,32–34,58</sup> Our protocol afforded the highest yield of desired product for both substrates, as well as exquisite selectivity for the desired monoarylated product. Interestingly, we found that the other conditions were not only lower yielding and less selective, but also highly substrate dependent, further highlighting the challenge of generality for the CL coupling of primary sulfonamides. The protocol developed by Watson and co-workers<sup>58</sup> afforded good yield and generality for the CL coupling of 4 and 12, albeit with competitive overarylation. Given the success of the Watson conditions relative to other literature conditions, we surveyed a broader scope of substrates for comparison. In general, we found that our conditions afforded higher (2, 4, 5, 7, 12, 19, 28, 29, 35) or comparable (24, 26) yields of the desired product for the substrates tested (Figure 6B). We also investigated whether introduction of an oxygen atmosphere could improve the yield under the conditions developed by Watson and co-workers; this generally led to a slight increase in the yield of the desired *N*-arylsulfonamide product, but at the expense of selectivity, with a concomitant increase in formation of the *N,N*-diarylsulfonamide byproduct (see SI for details). These observations are a testament to the orthogonality of these methods: Watson and co-workers' protocol works well for formation of *N,N*-diarylsulfonamides (symmetric and nonsymmetric alike), whereas our protocol affords selective monoarylation in high yields.

Pleased with the breadth of sulfonamide substrates that underwent successful *N*-arylation, we sought to explore the scope of boronic acid coupling partners with sulfonamide 29 (Figure 7A). Electron-rich (45), electron-poor (46), and neutral (47) boronic acids all underwent coupling efficiently. A testament to the orthogonality of the CL coupling to other cross-coupling methodologies, chloride- (48) and bromide-containing (49) boronic acids were well-tolerated under the reaction conditions. Sterically hindered boronic acids also underwent successful coupling, albeit with diminished yields (50, 51). Arylation was successful with pyridine boronic acid 52, demonstrating the promise of these conditions for the rapid construction of polyheterocyclic scaffolds. With an eye toward synthetic utility, we sought to evaluate whether pinacol

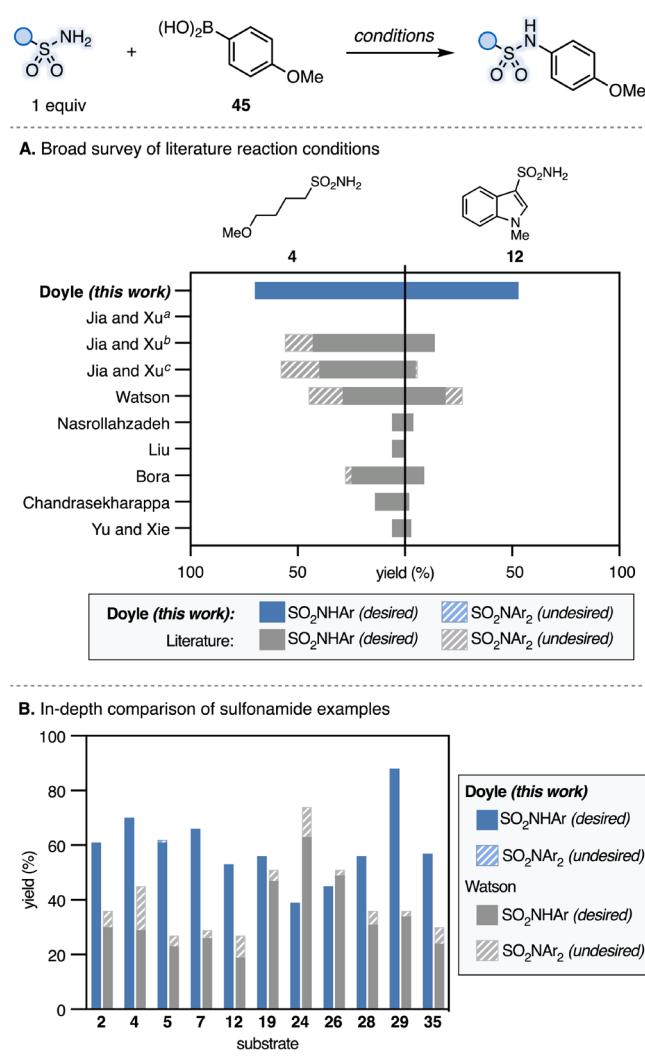


**Figure 5.** Yields of reaction between the training/validation set sulfonamides and 4-methoxyphenylboronic acid under the optimized reaction conditions (dark blue yields). Unless otherwise noted, trace or no diarylation was observed. <sup>a</sup>rt, 1 h. <sup>b</sup><sup>1</sup>H NMR yield. <sup>c</sup>Cu(OAc)<sup>2</sup> (20 mol %), K<sub>2</sub>CO<sub>3</sub> (2 equiv), DCE (0.1 M), O<sub>2</sub>, 60 °C, 1 h. Competitive diarylation observed in 0–28% yield (see SI for details).

boronate esters were competent under our reaction conditions. Excitingly, the coupling of phenylboronic acid pinacol ester **53** proceeded in comparable yield to the parent boronic acid.

Having demonstrated the broad utility of our optimized conditions across a variety of sulfonamide and boronic acid coupling partners, we sought to evaluate their application in more complex settings. In 2016, scientists at Merck developed a chemistry informer library of 24 medium-complexity aryl

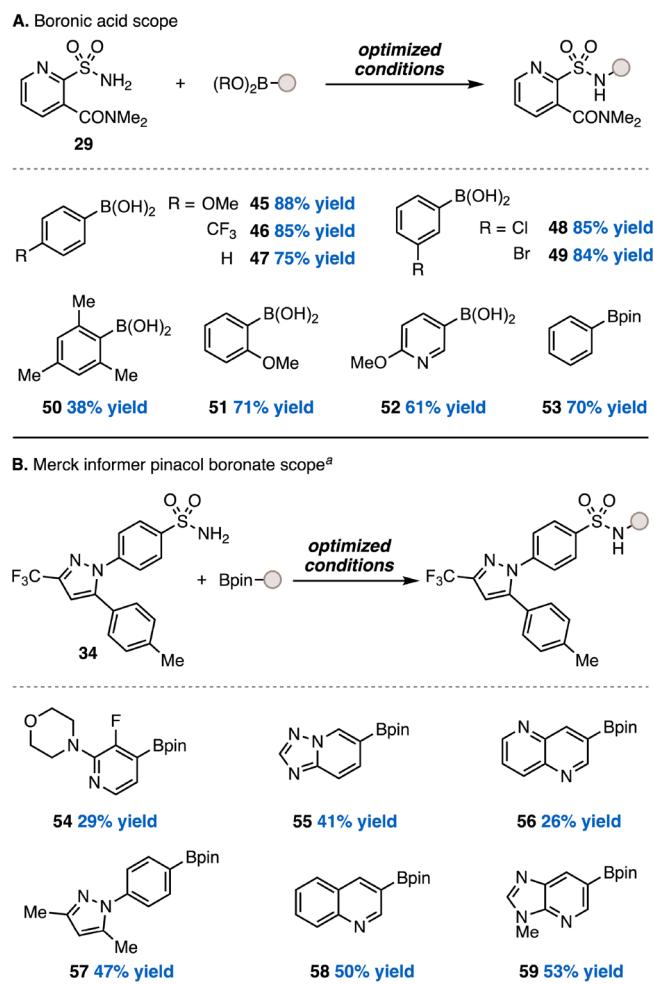
pinacol boronates as part of a study evaluating catalytic methods for the synthesis of pharmaceutically relevant compounds.<sup>66</sup> We evaluated the CL coupling of six of these pinacol boronates with Celecoxib to emulate a late-stage diversification campaign (Figure 7B). While the yields for these couplings were moderate, we were gratified to see that our conditions, which were optimized for boronic acids, could be



**Figure 6.** A) Survey of literature conditions for the CL coupling of substrates **4** and **12**. Jia and Xu conditions for <sup>2</sup>-, <sup>4</sup>-, and <sup>3</sup>-aminobenzenesulfonamides.<sup>26</sup> B) In-depth comparison of this work vs state-of-the-art conditions by Watson and co-workers.

successfully extended to drug-like heterocyclic pinacol boronates (**54–59**) with no changes.

To further highlight the utility of our protocol in the context of medicinal chemistry, we report the use of our CL coupling protocol as a key step in the synthesis of antitumor agent ODM-203.<sup>67,68</sup> The sulfonamide is a key structural feature of ODM-203, and is often incorporated in tyrosine kinase inhibitors.<sup>69</sup> As such, a synthetic route that enables late-stage diversification of the sulfonamide motif is highly desirable for medicinal chemistry campaigns. *N*-Arylsulfonamides are commonly introduced via aniline addition into a sulfonyl chloride, as in the original medicinal chemistry route toward ODM-203.<sup>67</sup> However, this route necessitates reduction of a nitro group to generate the aniline, installation of an acetyl protecting group to render the aniline compatible with subsequent chemistry, and deprotection to unveil the aniline required for the sulfonamide generation. By contrast, our proposed route utilizes robust reactions to build a convergent route for late-stage CL coupling, installing the *N*-arylsulfonamide as the final step (Figure 8). We leveraged our optimized conditions for the installation of cyclopropanesulfonamide in

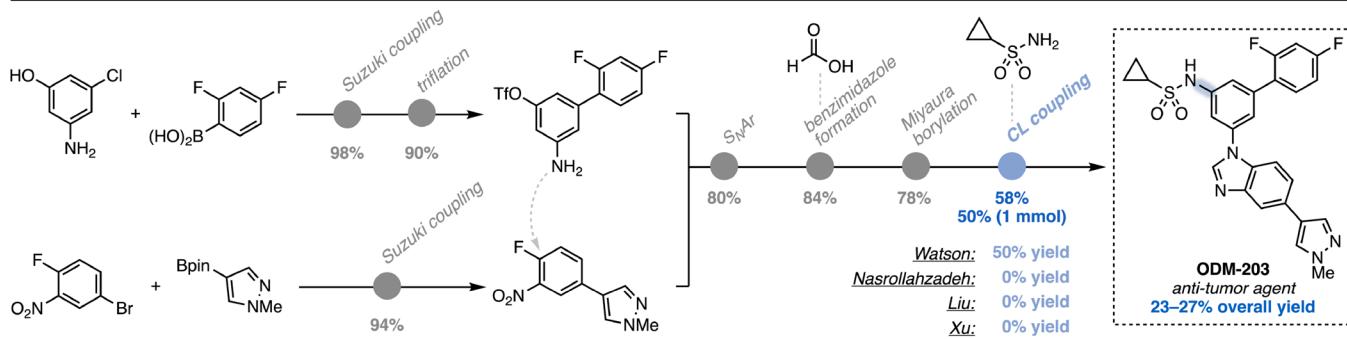


**Figure 7.** Scope of the CL coupling of **29** with boronic acids **45–53**. B) Scope of the CL coupling of **34** with Merck informer pinacol boronates **54–59**. <sup>a</sup>NMR yields.

58% yield with excellent selectivity for monoarylation as the final step in our synthesis of ODM-203 (0.05 mmol scale). Excitingly, the CL coupling was successfully scaled to 1 mmol scale in 50% yield. Ultimately, we realized a 6-step longest linear sequence synthesis in 23–27% overall yield. We also tested other literature conditions for the CL coupling of sulfonamides and found that only one other set of conditions enabled successful coupling in 50% yield, though it resulted in greater formation of the undesired overarylated product.

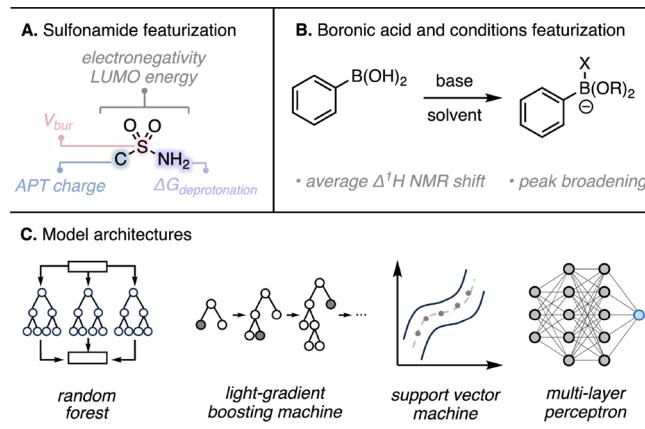
**2.4. Predictive Modeling.** Although our HTE campaign allowed for the identification of general conditions for the *N*-arylation of primary sulfonamides, the ability to make *de novo* predictions of reaction performance would be valuable given the highly substrate-dependent nature of the CL coupling. Furthermore, while our general conditions gave good yields of the desired products on average, we were unsurprised to find that the optimal conditions often differed as a function of sulfonamide identity. In fact, for 16 of the 22 sulfonamides in our training set, the highest yielding conditions diverged from the most general conditions. As such, we set out to predict high-performing conditions for unseen sulfonamide substrates in the CL coupling.

A prerequisite for any modeling effort is the selection of relevant descriptors. To this end, modeling campaigns often



**Figure 8.** Six-step synthesis of ODM-203, leveraging our optimized conditions for the final CL coupling. A comparison to other sulfonamide-specific Chan Lam conditions is shown in blue. All Chan Lam conditions tested with 4:1 sulfonamide:boronic ester stoichiometry. (Doyle (this work): Cu(OAc)<sub>2</sub> (20 mol %), KOT-Bu (2 equiv), H<sub>2</sub>O (10 equiv), EtOH (0.1 M), O<sub>2</sub>, 60 °C, 18 h). Watson: [Cu(MeCN)]PF<sub>6</sub> (0.5 equiv), K<sub>3</sub>PO<sub>4</sub> (8 equiv), MeCN (0.1 M), rt, 16 h.<sup>58</sup> Nasrollahzadeh: Cu(OAc)<sub>2</sub> (20 mol %), K<sub>2</sub>CO<sub>3</sub> (1.3 equiv), H<sub>2</sub>O (0.06 M), reflux. Liu: Cu(OAc)<sub>2</sub> (20 mol %), K<sub>2</sub>CO<sub>3</sub> (1.5 equiv), i-PrOH (0.1 M), 90 °C, 12 h.<sup>23</sup> Xu: CuCl (40 mol %), Et<sub>3</sub>N (1 equiv), dioxane (0.2 M), rt.<sup>26</sup> See SI for full details about synthetic route.

rely on hand-selection of descriptors based on a mechanistic hypothesis or chemical intuition.<sup>70</sup> Given the mechanistic ambiguity that plagues the CL coupling, we undertook a broad featurization campaign, relying on both algorithmic feature selection and chemical intuition to identify relevant descriptors. Despite the numerous reaction components present, we focused our efforts on generating molecular descriptors for the sulfonamide coupling partner, as we observed a strong influence of sulfonamide identity on reaction outcome in our HTE studies. Using our group's AutoQChem software,<sup>55</sup> we computed global and common atomic features, including electronegativity, atomic polar tensor (APT) charge, natural bond orbital (NBO) charge, and buried volume (Figure 9A).



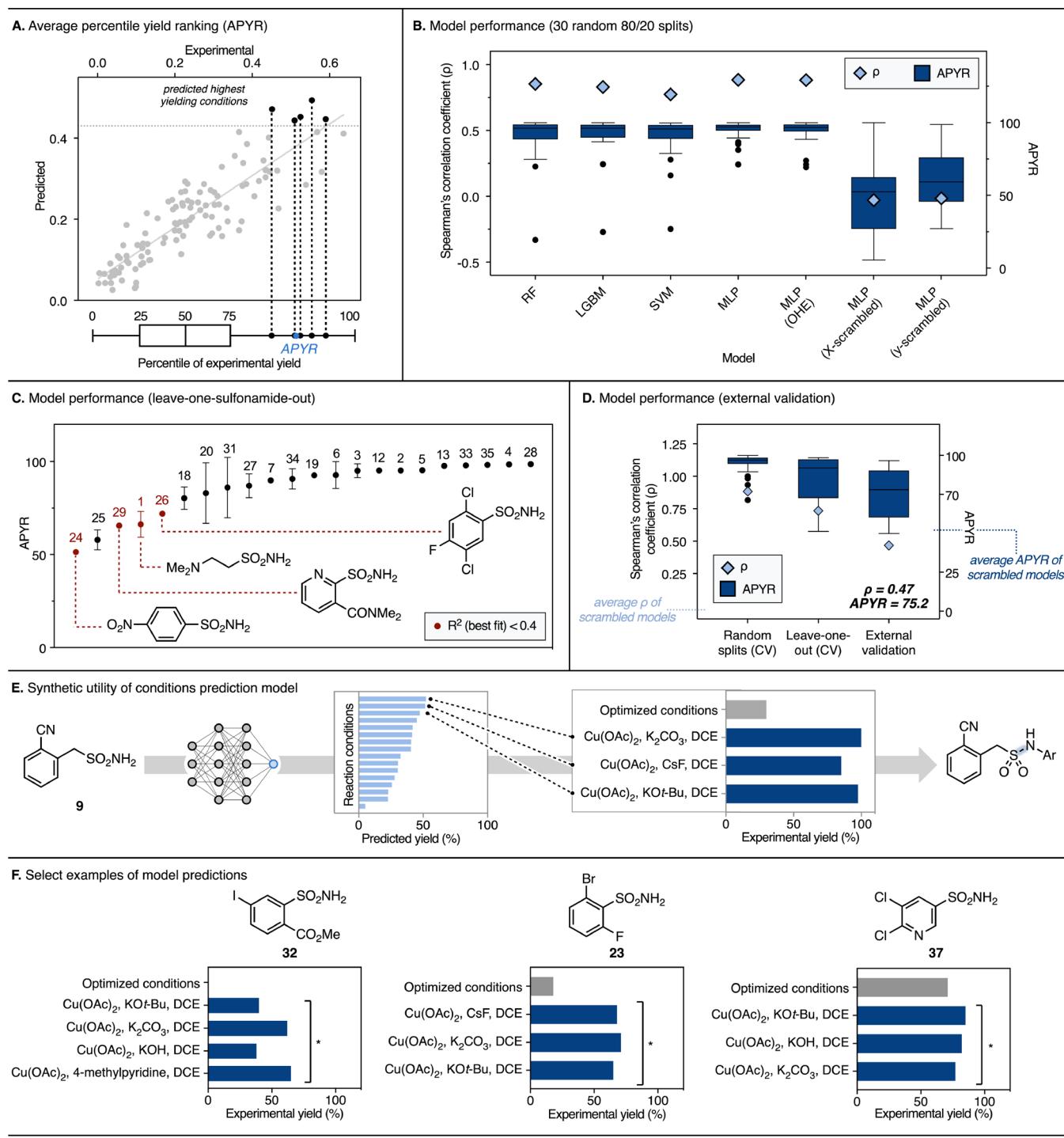
**Figure 9.** A) Computed and B) experimental features used for modeling. C) ML model architectures evaluated.

Given the observation from our HTE campaign (*vide supra*) that electron-deficient sulfonamides perform better on average than other sulfonamides, we also computed the gas phase free energy of sulfonamide deprotonation. Our optimization campaign also revealed the significant impact that base and solvent identity can have on the reaction outcome, particularly boronic acid speciation. To encode for this joint influence, we measured the average change in NMR shift for each boronic acid in each solvent as a function of base identity (Figure 9B); from these experiments, we also encoded categorical descriptors for whether remaining boronic acid was present,

and whether broadening of the boronic acid peaks was observed. As certain sulfonamides were observed to have reduced solubility, we also included  $\Delta G_{\text{solvation}}$  for each sulfonamide in each reaction solvent, predicted by a previously published ML model.<sup>71</sup> Experimentally determined categorical features were also generated, describing whether each sulfonamide forms a homogeneous solution in each solvent, and whether it forms a uniform mixture (including suspensions).

In total, 41 descriptors were generated across these methods to evaluate for model generation. We planned to evaluate an array of supervised ML methods—namely, multilayer perceptron (MLP), light gradient-boosting machine (LGBM), random forest (RF), and support vector machine (SVM) (Figure 9C) in our modeling campaign. However, we used LGBM models for feature selection due to the speed with which they can be trained. Ultimately, in the interest of generating an interpretable model, Shapley additive explanations<sup>72</sup> (SHAP) was used to select 9 descriptors that were best correlated to reaction outcome for the training set (see SI for details).

Several electronic properties were identified by feature selection, prompting us to also include Hirschfeld charge at the sulfonamide C and N, which has been shown to better correlate to the experimental physical organic Hammett parameter than other computed electronic features.<sup>73</sup> Although the computed  $\Delta G_{\text{deprotonation}}$  of the sulfonamides was not identified through feature selection, we included it in the model given our observation that more electron-poor sulfonamides performed better on average (*vide supra*). We also opted to include one-hot-encoded features for the base, catalyst, boronic acid, and solvent. While this limits the model's extrapolative ability in these dimensions, our screening indicated that the sulfonamide identity has the greatest influence on reaction outcome. We included sulfonamide one-hot-encoded features, as we saw improved performance upon their inclusion. These features complement the existing chemically informed descriptors by offering a clearer distinction when the model encounters substrates not present in the training set. Specifically, for an out-of-sample substrate, all one-hot-encoded columns related to training set substrates would equal 0.



**Figure 10.** A) Average percentile yield ranking (APYR) is calculated per substrate, by averaging the percentile of each experimental yield corresponding to the top yield predictions. B) Model performance comparison for random forest (RF), light-gradient boosting machine (LGBM), support-vector machine (SVM), and multilayer perceptron (MLP). C) Leave-one-sulfonamide-out cross validation performance per substrate. D) Comparison of MLP performance for random cross validation, leave-one-out cross validation, and external validation. E) The prediction workflow involves feeding substrate features into the MLP model, requesting predictions for all reaction conditions, and experimentally validating top predicted conditions. F) Select examples of experimental performance of top-predicted conditions.

Given the notorious fickleness of the CL coupling, we sought to focus on predicting reactivity trends across different conditions, rather than precise yield values. For this reason, we evaluated model performance with percentiles and the ranking-based Spearman's correlation coefficient ( $\rho$ ),<sup>74</sup> rather than traditional metrics like  $R^2$ , MAE, or RMSE. For each substrate, we took all predictions within 5% yield of the highest predicted

yield and determined the percentile rank of the corresponding experimental yields. The average of these percentile ranks was used to assess overall predictive performance (Figure 10A). Henceforth, we refer to this metric as average percentile yield ranking (APYR). A high APYR value indicates that the model can successfully identify high-yielding conditions within the yields attainable for a given substrate. This trend-based

approach is more practical for a chemist seeking to identify top-yielding reaction conditions, allowing them to focus on a few promising options instead of relying on trial and error.

We evaluated the predictions from an array of supervised ML methods using 30 random 80:20 splits, wherein 80% of the data is used to train a model that is then used to predict high-yielding reaction conditions for the remaining 20% (validation set) (see SI for model hyperparameters) (Figure 10B). The highest-performing model (MLP) has  $\rho = 0.884$  and APYR of  $94.8 \pm 6.3$ . This suggests that on average, the model is able to predict reaction conditions for a given substrate that fall above the  $94.8 \pm 6.3$  percentile of experimental yield. To benchmark model performance, we generated three baseline MLP models using 1) only OHE, 2) X-randomized, and 3) y-randomized features. As expected, the chemically informed MLP model significantly ( $P < 0.001$  for permutation test) outperformed the X- and y-randomized baseline models, which had negative  $\rho$  values and APYR of  $49.0 \pm 24.7$  and  $60.9 \pm 20.3$ , respectively. Although the informed MLP model did not show significant improvement over the OHE baseline model ( $\rho = 0.882$  and APYR =  $94.2 \pm 7.9$ ), we believe this may be a demonstration of the ability of the complex neural network framework to learn underlying reactivity patterns in the OHE data set.

While random splits are often used to evaluate model performance, they provide an optimistic estimate of a model's ability to extrapolate to unseen substrates. To simulate an out-of-sample test case more closely, we estimated generalization error with leave-one-group-out (LOGO) cross-validation, wherein one sulfonamide substrate is withheld from model training entirely, and all the data for that substrate is used instead as the validation set. This process is repeated for each sulfonamide, and the final metrics are averaged across all substrates. This evaluation tends to result in larger error but is more representative of a real-life scenario. When faced with this more challenging assessment, the MLP model performed well, with  $\rho = 0.735$  and APYR of  $85.1 \pm 14.5$  (vs  $\rho = 0.717$  and APYR of  $83.4 \pm 16.6$  for the OHE baseline model). The lower value for Spearman's coefficient shows that for out-of-sample predictions, it is more difficult for the model to accurately rank reaction conditions from best to worst. Closer analysis reveals that only four substrates (1, 24, 25, and 29) had APYR values below the 75th percentile; the remaining 18 substrates from the training set scored  $90.7 \pm 8.7$  on average (Figure 10C). The APYR values suggest that the model would still be highly effective for predicting promising reaction conditions for unseen sulfonamides.

The LOGO assessment provided the opportunity to gain further insight into the predictive capabilities and limitations of the model. We analyzed the experimental vs predicted yields for each individual substrate, using the  $R^2$  for the line of best fit to evaluate how effectively our model captures underlying reactivity trends and identifies high-yielding conditions. All the substrates had lines of best fit with  $R^2 > 0.4$  with the exceptions of 1, 24, 26, and 29 (Figure 10C, red points). Computational studies revealed that 1 and 29 can access bidentate coordinate modes with a copper catalyst (computed with  $\text{Cu(OAc)}_2$ ); although other substrates such as 6 were also favored to access bidentate complexes, it is possible that the solution-phase behavior of 1 and 29 differ sufficiently to alter observed reactivity; inclusion of additional substrates for which similar binding modes are accessible may improve model performance in these cases. 24 and 26 are highly electron-poor examples that may not be adequately represented in the data set.

Finally, we tested our model's ability to predict high-yielding reaction conditions for CL coupling of previously selected unseen test substrates (8–11, 14–17, 21–23, 30, 32, 36–44) (*vide supra*, Figure 5). The model was able to capture reactivity trends and predict high-yielding conditions for the external validation data set, with  $\rho = 0.468$  and APYR =  $75.2 \pm 15.8$  (vs  $\rho = 0.474$  and APYR =  $77.3 \pm 20.0$  for the OHE baseline model) (Figure 10D). Importantly, these statistics are significantly more predictive ( $P < 0.005$ ) than the X- and y-scrambled baseline models, suggesting that the chemically informed model is better than random. The model predicts high-yielding conditions with APYR > 70th percentile for 12 of the 22 external validation substrates, and  $\geq 50$ th percentile for all 22. Interestingly, the substrates with APYR values below 60 tended to be coordinating and/or highly electron-deficient, in line with our LOGO analysis.

While the Spearman's correlation coefficient is lower for the external validation data set, this is expected given that external data often presents new challenges that may not be fully represented in the training set. These results suggest that some substrate-dependent effects, especially pertaining to sulfonamide electronics and interactions with catalyst, may be inadequately represented in the training set, and/or that the features used do not fully capture the relevant information.

The challenges of modeling notwithstanding, our results indicate that our model successfully captures the underlying trends in reactivity for this notoriously difficult-to-predict reaction. Though the chemically informed model offers marginal improvements to overall performance over the OHE baseline model, it is better able to capture substrate-level reactivity trends. The excellent APYR values indicate that a chemist could use our model to identify high-yielding reaction conditions with confidence, particularly for an unseen substrate (Figure 10E). For example, for validation substrates 32 and 23, our optimized conditions gave 0 and 18% yield, respectively. We queried the model for high-yielding conditions for these substrates, and it delivered successful conditions that afforded above 60% yield (Figure 10F). The model was also able to predict high-yielding conditions for substrates that worked well under our optimized conditions, such as 37.

Although the model may not perfectly predict precise yields, its ability to rank reaction conditions by relative performance provides significant value in narrowing down the conditions most likely to succeed. This is particularly advantageous in medicinal chemistry, where rapid diversification is required, as well as in scenarios where material is precious or experiments are resource-intensive. Our model may serve as a decision-making tool for selecting promising reaction conditions without exhaustive and expensive trial and error.

### 3. CONCLUSIONS

In conclusion, we applied an unsupervised learning workflow to select a diverse training set, collected an HTE data set of 3,904 unique experiments, and constructed a predictive ML model. The diversity in our training set led to the identification of general conditions for the selective monoarylation of primary sulfonamides; we established the broadest scope of primary sulfonamides to date, filling a previously outstanding gap in the CL methodology. Careful feature engineering enabled the generation of an ML model that successfully predicts high-yielding conditions for the CL coupling of unseen sulfonamide substrates, ultimately saving valuable time

and resources for such a substrate-dependent reaction. Analysis of our data set and the resulting models revealed several opportunities for future improvement, including the study of substrates that can ligate copper, as well as substrates of varying nucleophile acidity. Importantly, this work outlines a workflow for employing various data science tools to study challenging reactions. It is our hope that this study not only provides practicing organic chemists with general and selective reaction conditions and a useful ML model, but also inspires similar investigations of other reactions for which a panacea has yet to be discovered.

## ■ ASSOCIATED CONTENT

### Data Availability Statement

All scripts are available on GitHub at [github.com/doyle-lab-ucla/CL-sulfonamides](https://github.com/doyle-lab-ucla/CL-sulfonamides). The HTE data set is available at [open-reaction-database.org](https://open-reaction-database.org) (ORD data set ID: ord\_dataset5c9a10329a8a48968d18879a48bb8ab2).

### ■ Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acscatal.4c07972>.

Experimental procedures, experimental data, characterization and spectral data, and modeling details ([PDF](#))

## ■ AUTHOR INFORMATION

### Corresponding Authors

**Iulia I. Strambeanu** — *Chemistry Capabilities, Analytical and Purification, Global Discovery Chemistry, Johnson & Johnson, Spring House, Pennsylvania 19477, United States;*  [orcid.org/0000-0002-1502-5484](https://orcid.org/0000-0002-1502-5484); Email: [istrambe@its.jnj.com](mailto:istrambe@its.jnj.com)

**Kristi A. Leonard** — *Global Discovery Chemistry, Johnson & Johnson, Spring House, Pennsylvania 19477, United States;*  [orcid.org/0000-0002-9398-3354](https://orcid.org/0000-0002-9398-3354); Email: [kleonar1@its.jnj.com](mailto:kleonar1@its.jnj.com)

**Abigail G. Doyle** — *Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095, United States;*  [orcid.org/0000-0002-6641-0833](https://orcid.org/0000-0002-6641-0833); Email: [agdoyle@chem.ucla.edu](mailto:agdoyle@chem.ucla.edu)

### Authors

**Shivaani S. Gandhi** — *Department of Chemistry, Princeton University, Princeton, New Jersey 08544, United States;* *Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095, United States;*  [orcid.org/0000-0003-1825-5450](https://orcid.org/0000-0003-1825-5450)

**Giselle Z. Brown** — *Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095, United States*

**Santeri Aikonen** — *Drug Discovery Data Science, In Silico Discovery, Johnson & Johnson, Spring House, Pennsylvania 19477, United States;*  [orcid.org/0000-0003-2675-7290](https://orcid.org/0000-0003-2675-7290)

**Jordan S. Compton** — *Chemistry Capabilities, Analytical and Purification, Global Discovery Chemistry, Johnson & Johnson, Spring House, Pennsylvania 19477, United States;*  [orcid.org/0000-0001-7099-3456](https://orcid.org/0000-0001-7099-3456)

**Paulo Neves** — *Drug Discovery Data Science, In Silico Discovery, Johnson & Johnson, 2740-244 Porto Salvo, Portugal;*  [orcid.org/0000-0001-9556-6418](https://orcid.org/0000-0001-9556-6418)

**Jesus I. Martinez Alvarado** — *Department of Chemistry, Princeton University, Princeton, New Jersey 08544, United States;* Present Address: Discovery High-Throughput

Chemistry, GlaxoSmithKline, Collegeville, Pennsylvania 19426

Complete contact information is available at:  
<https://pubs.acs.org/10.1021/acscatal.4c07972>

### Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

### Author Contributions

<sup>a</sup>S.A., J.S.C., and P.N. contributed equally.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This project has been funded through the generous contributions of the Princeton Catalysis Initiative and Janssen R&D (systematic dataset design, HTE data collection and analysis, predictive modeling) and the NSF CCI Center for Computer Assisted Synthesis (CHE-1925607) (synthetic application). These studies were supported by shared instrumentation grants from the NSF (CHE-1048804) and NIH Office of Research Infrastructure Programs (S10OD028644). DFT calculations were performed using resources managed and supported by Princeton Research Computing. G.Z.B. gratefully acknowledges the Clare Boothe Luce Scholars Program for financial support. We thank Dr. Kiran Kumar for providing the Janssen sulfonamide library and Paris Dee for conducting bench-scale solvent screens.

## ■ REFERENCES

- (1) Scott, K. A.; Njardarson, J. T. Analysis of US FDA-Approved Drugs Containing Sulfur Atoms. *Top. Curr. Chem.* **2018**, *376*, 5.
- (2) Ovung, A.; Bhattacharyya, J. Sulfonamide Drugs: Structure, Antibacterial Property, Toxicity, and Biophysical Interactions. *Biophys. Rev.* **2021**, *13*, 259–272.
- (3) Pierson, D. A.; Olsen, B. A.; Robbins, D. K.; DeVries, K. M.; Varie, D. L. Approaches to Assessment, Testing Decisions, and Analytical Determination of Genotoxic Impurities in Drug Substances. *Org. Process Res. Dev.* **2009**, *13*, 285–291.
- (4) Snodin, D. J. Genotoxic Impurities: From Structural Alerts to Qualification. *Org. Process Res. Dev.* **2010**, *14*, 960–976.
- (5) He, H.; Wu, Y.-J. Copper-Catalyzed N-Arylation of Sulfonamides with Aryl Bromides and Iodides Using Microwave Heating. *Tetrahedron Lett.* **2003**, *44*, 3385–3386.
- (6) Deng, W.; Liu, L.; Zhang, C.; Liu, M.; Guo, Q.-X. Copper-Catalyzed Cross-Coupling of Sulfonamides with Aryl Iodides and Bromides Facilitated by Amino Acid Ligands. *Tetrahedron Lett.* **2005**, *46*, 7295–7298.
- (7) Hosseinzadeh, R.; Tajbakhsh, M.; Mohadjerani, M.; Alikarami, M. Copper-Catalysed N-Arylation of Arylsulfonamides with Aryl Bromides and Aryl Iodides Using KF/Al<sub>2</sub>O<sub>3</sub>. *J. Chem. Sci.* **2010**, *122*, 143–148.
- (8) Shekhar, S.; Dunn, T. B.; Kotecki, B. J.; Montavon, D. K.; Cullen, S. C. A General Method for Palladium-Catalyzed Reactions of Primary Sulfonamides with Aryl Nonaflates. *J. Org. Chem.* **2011**, *76*, 4552–4563.
- (9) Rosen, B. R.; Ruble, J. C.; Beauchamp, T. J.; Navarro, A. Mild Pd-Catalyzed N-Arylation of Methanesulfonamide and Related Nucleophiles: Avoiding Potentially Genotoxic Reagents and By-products. *Org. Lett.* **2011**, *13*, 2564–2567.
- (10) Wang, X.; Guram, A.; Ronk, M.; Milne, J. E.; Tedrow, J. S.; Paul, M. M. Copper-Catalyzed N-Arylation of Sulfonamides with Aryl Bromides under Mild Conditions. *Tetrahedron Lett.* **2012**, *53*, 7–10.

- (11) Teo, Y.; Yong, F.; Ithnin, I. K.; Yio, S. T.; Lin, Z. Efficient Manganese/Copper Bimetallic Catalyst for *N*-Arylation of Amides and Sulfonamides Under Mild Conditions in Water. *Eur. J. Org. Chem.* **2013**, 515–524.
- (12) Tan, B. Y.; Teo, Y.; Seow, A. Low Catalyst Loadings for Ligand-Free Copper(I)-Oxide-Catalyzed *N*-Arylation of Methanesulfonamide in Water. *Eur. J. Org. Chem.* **2014**, 2014, 1541–1546.
- (13) Zhang, W.; Yang, D.; Wang, W.; Wang, S.; Zhao, H. Iridium(III)-Catalyzed Directed Ortho-C(sp<sup>2</sup>)–H Amidation of Arenes with Sulfonamides. *Eur. J. Org. Chem.* **2018**, 2071–2077.
- (14) Zhao, Z.; Lian, Y.; Zhao, C.; Wang, B. Palladium-Catalyzed Desulfitative Arylation of Sulfonamides with Sodium Arylsulfinate. *Synth. Commun.* **2018**, 48, 1436–1442.
- (15) Laffoon, J. D.; Chan, V. S.; Fickes, M. G.; Kotecki, B.; Ickes, A. R.; Henle, J.; Napolitano, J. G.; Franczyk, T. S.; Dunn, T. B.; Barnes, D. M.; Haight, A. R.; Henry, R. F.; Shekhar, S. Pd-Catalyzed Cross-Coupling Reactions Promoted by Biaryl Phosphorinane Ligands. *ACS Catal.* **2019**, 9, 11691–11708.
- (16) McGuire, R. T.; Simon, C. M.; Yadav, A. A.; Ferguson, M. J.; Stradiotto, M. Nickel-Catalyzed Cross-Coupling of Sulfonamides With (Hetero)Aryl Chlorides. *Angew. Chem., Int. Ed.* **2020**, 59, 8952–8956.
- (17) Vu, H.-M.; Yong, J.-Y.; Chen, F.-W.; Li, X.-Q.; Shi, G.-Q. Rhodium-Catalyzed C(sp<sup>2</sup>)–H Amidation of Azine with Sulfonamides. *J. Org. Chem.* **2020**, 85, 4963–4972.
- (18) Li, Q.; Xu, L.; Ma, D. Cu-Catalyzed Coupling Reactions of Sulfonamides with (Hetero)Aryl Chlorides/Bromides. *Angew. Chem., Int. Ed.* **2022**, 61, No. e202210483.
- (19) Teo, Y. C.; Tan, Y. R.; Kyra Saanvi, M.; Loh, C. K.; Tan, S. N. Copper Catalyzed *N*-Arylation of Sulfonamides with Aryl Bromides under Ligand-Free Conditions. *Synth. Commun.* **2023**, 53, 1143–1152.
- (20) Kim, T.; McCarver, S. J.; Lee, C.; MacMillan, D. W. C. Sulfonamidation of Aryl and Heteroaryl Halides through Photo-sensitized Nickel Catalysis. *Angew. Chem., Int. Ed.* **2018**, 57, 3488–3492.
- (21) Zhu, C.; Kale, A. P.; Yue, H.; Rueping, M. Redox-Neutral Cross-Coupling Amination with Weak *N*-Nucleophiles: Arylation of Anilines, Sulfonamides, Sulfoximines, Carbamates, and Imines via Nickelaelectrocatalysis. *JACS Au* **2021**, 1, 1057–1065.
- (22) Zhao, T.-T.; Qin, H.-N.; Xu, P.-F. Light-Promoted Nickel-Catalyzed C–O/C–N Coupling of Aryl Halides with Carboxylic Acids and Sulfonamides. *Org. Lett.* **2023**, 25, 636–641.
- (23) Pan, C.; Cheng, J.; Wu, H.; Ding, J.; Liu, M. Cu(OAc)<sub>2</sub>-Catalyzed *N*-Arylation of Sulfonamides with Arylboronic Acids or Trimethoxy(Phenyl)Silane. *Synth. Commun.* **2009**, 39, 2082–2092.
- (24) Meshram, G. A.; Patil, V. D. A Simple and Efficient *N*-Arylation of Amines and Sulfonamides with Cu(BF<sub>4</sub>)<sub>2</sub>•SiO<sub>2</sub>. *Org. Chem. Ind. J.* **2009**, 5, 434–440.
- (25) Nasrollahzadeh, M.; Ehsani, A.; Maham, M. Copper-Catalyzed *N*-Arylation of Sulfonamides with Boronic Acids in Water under Ligand-Free and Aerobic Conditions. *Synlett* **2014**, 25, 505–508.
- (26) Zu, W.; Liu, S.; Jia, X.; Xu, L. Chemoselective *N*-Arylation of Aminobenzene Sulfonamides via Copper Catalysed Chan–Evans–Lam Reactions. *Org. Chem. Front.* **2019**, 6, 1356–1360.
- (27) Raju, S.; Teimouri, M.; Adhikari, B.; Donnadieu, B.; Stokes, S. L.; Emerson, J. P. Copper Complexes for the Chemoselective *N*-Arylation of Arylamines and Sulfonamides via Chan–Evans–Lam Cross-Coupling. *Dalton Trans.* **2023**, 52, 15986–15994.
- (28) Nasrollahzadeh, M.; Rostami-Vartooni, A.; Ehsani, A.; Moghadam, M. Fabrication, Characterization and Application of Nanopolymer Supported Copper (II) Complex as an Effective and Reusable Catalyst for the CN Bond Cross-Coupling Reaction of Sulfonamides with Arylboronic Acids in Water under Aerobic Conditions. *J. Mol. Catal. A: Chem.* **2014**, 387, 123–129.
- (29) Azarifar, D.; Soleimanei, F. Natural Indian Natrolite Zeolite-Supported Cu Nanoparticles: A New and Reusable Heterogeneous Catalyst for *N*-Arylation of Sulfonamides with Boronic Acids in Water under Ligand-Free Conditions. *RSC Adv.* **2014**, 4, 12119–12126.
- (30) Nasrollahzadeh, M.; Nezafat, Z.; Pakzad, K.; Ahmadpoor, F. Synthesis of Magnetic Chitosan Supported Metformin-Cu(II) Complex as a Recyclable Catalyst for *N*-Arylation of Primary Sulfonamides. *J. Organomet. Chem.* **2021**, 948, No. 121915.
- (31) Kaboudin, B.; Abedi, Y.; Yokomatsu, T. Cu<sup>II</sup>– $\beta$ -Cyclodextrin Complex as a Nanocatalyst for the Homo- and Cross-Coupling of Arylboronic Acids under Ligand- and Base-Free Conditions in Air: Chemoselective Cross-Coupling of Arylboronic Acids in Water. *Eur. J. Org. Chem.* **2011**, 6656–6662.
- (32) Chandrasekharappa, A. P.; Badiger, S. E.; Dubey, P. K.; Panigrahi, S. K.; Manukonda, S. R. V. V. V. Design and Synthesis of 2-Substituted Benzoxazoles as Novel PTP1B Inhibitors. *Bioorg. Med. Chem. Lett.* **2013**, 23, 2579–2584.
- (33) Saikia, R.; Das, S.; Almin, A.; Mahanta, A.; Sarma, B.; Thakur, A. J.; Bora, U. N. *N'*-Dimethylurea as an Efficient Ligand for the Synthesis of Pharma-Relevant Motifs through Chan–Lam Cross-Coupling Strategy. *Org. Biomol. Chem.* **2023**, 21, 3143–3155.
- (34) Lan, J.-B.; Zhang, G.-L.; Yu, X.-Q.; You, J.-S.; Chen, L.; Yan, M.; Xie, R.-G. A Simple Copper Salt Catalyzed *N*-Arylation of Amines, Amides, Imides, and Sulfonamides with Arylboronic Acids. *Synlett* **2004**, 6, 1095–1097.
- (35) Gunning, P. T.; Park, J. S.; Ahmar, S.; Cabral, A. D.; Tin, G. K. C.; Rasheed, S.; Abdeldayem, A.; Armstrong, D.; Frere, G. A.; Quilates, E. J.; Rosa, D. A.; Gozhina, O.; Omeara, J. A.; Simpson, G. Benzenesulfonamide Derivatives and Uses Thereof. U.S. Patent WO20219568, January 21, 2021.
- (36) Kraskouskaya, D.; Ahmar, S.; Gunning, P. T.; Omeara, J.; Rosa, D. A.; Park, J. S.; Simpson, G. L.; Abdeldayem, A. Pentafluorobenzenesulfonamide Derivatives and Uses Thereof. U.S. Patent WO202199842, May 27, 2021.
- (37) Andrez, J.-C.; Burford, K. N.; Dehnhardt, C. M.; Focken, T.; Grimwood, M. E.; Jia, Q.; Lofstrand, V. A.; Wesolowski, S. S.; Wilson, M. S. Heteroaryl-Substituted Sulfonamide Compounds and Their Use As Sodium Channel Inhibitors. U.S. Patent 202071313, March 5, 2020.
- (38) Qian, P.; Zhang, P.; Shi, L.; Wang, F.; Hou, Y. Method for Preparing Sulfentrazone through Catalytic Coupling of Copper Reagent. Chinese Patent CN109796419, May 24, 2019.
- (39) Schiemann, K. Substituted Indazoles and Related Heterocycles. U.S. Patent WO201641618, March 24, 2016.
- (40) Langkopf, E.; Himmelsbach, F.; Mack, J.; Pautsch, A.; Schoelch, C.; Schuler-Metz, A.; Streicher, R.; Wagner, H. New Arylsulphonylglycine Derivatives, the Preparation Thereof and Their Use As Medicaments. U.S. Patent WO2009127723, October 22, 2009.
- (41) Vantourout, J. C.; Miras, H. N.; Isidro-Llobet, A.; Sproules, S.; Watson, A. J. B. Spectroscopic Studies of the Chan–Lam Amination: A Mechanism-Inspired Solution to Boronic Ester Reactivity. *J. Am. Chem. Soc.* **2017**, 139, 4769–4779.
- (42) Bose, S.; Dutta, S.; Koley, D. Entering Chemical Space with Theoretical Underpinning of the Mechanistic Pathways in the Chan–Lam Amination. *ACS Catal.* **2022**, 12, 1461–1474.
- (43) Dong, Z.-B.; Chen, J.-Q.; Li, J.-H. A Review on the Latest Progress of Chan–Lam Coupling Reaction. *Adv. Synth. Catal.* **2020**, 3311.
- (44) Meuwly, M. Machine Learning for Chemical Reactions. *Chem. Rev.* **2021**, 121, 10218–10239.
- (45) Shi, Y.-F.; Yang, Z.-X.; Ma, S.; Kang, P.-L.; Shang, C.; Hu, P.; Liu, Z.-P. Machine Learning for Chemistry: Basics and Applications. *Engineering* **2023**, 70.
- (46) Tu, Z.; Stuyver, T.; Coley, C. W. Predictive Chemistry: Machine Learning for Reaction Deployment, Reaction Development, and Reaction Discovery. *Chem. Sci.* **2023**, 14, 226–244.
- (47) Su, A.; Wang, X.; Wang, L.; Zhang, C.; Wu, Y.; Wu, X.; Zhao, Q.; Duan, H. Reproducing the Invention of a Named Reaction: Zero-Shot Prediction of Unseen Chemical Reactions. *Phys. Chem. Chem. Phys.* **2022**, 24, 10280–10291.
- (48) Kariofillis, S. K.; Jiang, S.; Źurański, A. M.; Gandhi, S. S.; Martinez Alvarado, J. I.; Doyle, A. G. Using Data Science To Guide Aryl Bromide Substrate Scope Analysis in a Ni/Photoredox-Catalyzed

- Cross-Coupling with Acetals as Alcohol-Derived Radical Sources. *J. Am. Chem. Soc.* **2022**, *144*, 1045–1055.
- (49) Gensch, T.; Smith, S. R.; Colacot, T. J.; Timsina, Y. N.; Xu, G.; Glasspoole, B. W.; Sigman, M. S. Design and Application of a Screening Set for Monophosphine Ligands in Cross-Coupling. *ACS Catal.* **2022**, *12*, 7773–7780.
- (50) Zahrt, A. F.; Henle, J. J.; Rose, B. T.; Wang, Y.; Darrow, W. T.; Denmark, S. E. Prediction of Higher-Selectivity Catalysts by Computer-Driven Workflow and Machine Learning. *Science* **2019**, *363*, No. eaau5631.
- (51) Rana, D.; Pflüger, P. M.; Höltner, N. P.; Tan, G.; Glorius, F. Standardizing Substrate Selection: A Strategy toward Unbiased Evaluation of Reaction Generality. *ACS Cent. Sci.* **2024**, *10*, 899–906.
- (52) Mennen, S. M.; Alhambra, C.; Allen, C. L.; Barberis, M.; Berritt, S.; Brandt, T. A.; Campbell, A. D.; Castañón, J.; Cherney, A. H.; Christensen, M.; Damon, D. B.; Eugenio de Diego, J.; García-Cerrada, S.; García-Losada, P.; Haro, R.; Janey, J.; Leitch, D. C.; Li, L.; Liu, F.; Lobben, P. C.; MacMillan, D. W. C.; Magano, J.; McInturff, E.; Monfette, S.; Post, R. J.; Schultz, D.; Sitter, B. J.; Stevens, J. M.; Strambaneau, I. I.; Twilton, J.; Wang, K.; Zajac, M. A. The Evolution of High-Throughput Experimentation in Pharmaceutical Development and Perspectives on the Future. *Org. Process Res. Dev.* **2019**, *23*, 1213–1242.
- (53) Moriwaki, H.; Tian, Y. S.; Kawashita, N.; Takagi, T. Mordred: A Molecular Descriptor Calculator. *J. Cheminform.* **2018**, *10*, 4.
- (54) 2-, 3-, and 4-Aminobenzenesulfonamide are excluded from this chemical space plot, as competitive nucleophilic functionality were filtered from the chemical space search. See Supporting Information for detailed list of filters applied.
- (55) Żurański, A. M.; Wang, J. Y.; Shields, B. J.; Doyle, A. G. Auto-QChem: An Automated Workflow for the Generation and Storage of DFT Calculations for Organic Molecules. *React. Chem. Eng.* **2022**, *7*, 1276–1284.
- (56) McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* **2018**. DOI: [10.48550/arXiv.1802.03426](https://doi.org/10.48550/arXiv.1802.03426)
- (57) West, M. J.; Fyfe, J. W. B.; Vantourout, J. C.; Watson, A. J. B. Mechanistic Development and Recent Applications of the Chan–Lam Amination. *Chem. Rev.* **2019**, *119*, 12491–12523.
- (58) Vantourout, J. C.; Li, L.; Bendito-Moll, E.; Chabarra, S.; Arrington, K.; Bode, B. E.; Isidro-Llobet, A.; Kowalski, J. A.; Nilson, M. G.; Wheelhouse, K. M. P.; Woodard, J. L.; Xie, S.; Leitch, D. C.; Watson, A. J. B. Mechanistic Insight Enables Practical, Scalable, Room Temperature Chan–Lam N-Arylation of N-Aryl Sulfonamides. *ACS Catal.* **2018**, *8*, 9560–9566.
- (59) Bolz, R. E.; Tuve, G. L. *CRC Handbook of Tables for Applied Engineering Science*; CRC Press: Boca Raton, FL, 1973.
- (60) Bard, A. J.; Parsons, R.; Jordan, J. *Standard Potentials in Aqueous Solution*; CRC Press: Boca Raton, FL, 1985.
- (61) In our investigations, comparable reactivity is possible with an air atmosphere, provided there is a sufficiently large headspace. However, for reactions on 0.1 mmol scale, we have found it most practical to utilize a smaller vessel, such as a reaction tube, under an O<sub>2</sub> atmosphere. Detailed information regarding vessel size and atmosphere screening is available in the Supporting Information.
- (62) Evans, D. A.; Katz, J. L.; West, T. R. Synthesis of Diaryl Ethers through the Copper-Promoted Arylation of Phenols with Arylboronic Acids. An Expedient Synthesis of Thyroxine. *Tetrahedron Lett.* **1998**, *39*, 2937–2940.
- (63) Grimaud, L.; Jutand, A. Role of Fluoride Ions in Palladium-Catalyzed Cross-Coupling Reactions. *Synthesis* **2017**, *49*, 1182–1189.
- (64) Carrow, B. P.; Hartwig, J. F. Distinguishing Between Pathways for Transmetalation in Suzuki–Miyaura Reactions. *J. Am. Chem. Soc.* **2011**, *133*, 2116–2119.
- (65) Battersby, D. J. High-Throughput Methods For Reaction Development Using The Mosquito® Liquid Handling Robot, Ph.D. Dissertation, University of Cambridge, Cambridge, UK, 2018.
- (66) Kutchukian, P. S.; Dropinski, J. F.; Dykstra, K. D.; Li, B.; DiRocco, D. A.; Streckfuss, E. C.; Campeau, L.-C.; Cernak, T.; Vachal, P.; Davies, I. W.; Kraska, S. W.; Dreher, S. D. Chemistry Informer Libraries: A Chemoinformatics Enabled Approach to Evaluate and Advance Synthetic Methods. *Chem. Sci.* **2016**, *7*, 2604–2613.
- (67) Holmström, T. H.; Moilanen, A.-M.; Ikonen, T.; Björkman, M. L.; Linnanen, T.; Wohlfahrt, G.; Karlsson, S.; Oksala, R.; Korjamo, T.; Samajdar, S.; Rajagopalan, S.; Chelur, S.; Narayan, K.; Ramachandra, R. K.; Mani, J.; Nair, R.; Gowda, N.; Anthony, T.; Dhodheri, S.; Mukherjee, S.; Ujjinamatada, R. K.; Srinivas, N.; Ramachandra, M.; Kallio, P. J. ODM-203 a Selective Inhibitor of FGFR and VEGFR, Shows Strong Anti-Tumor Activity, and Induces Anti-Tumor Immunity. *Mol. Cancer Ther.* **2018**, *18*, 28.
- (68) Bono, P.; Massard, C.; Peltola, K. J.; Azaro, A.; Italiano, A.; Kristeleit, R. S.; Curigliano, G.; Lassen, U.; Arkenau, H.-T.; Hakulinen, P.; Garratt, C.; Ikonen, T.; Mustonen, M. V. J.; Rodon, J. A. Phase I/IIa, Open-Label, Multicentre Study to Evaluate the Optimal Dosing and Safety of ODM-203 in Patients with Advanced or Metastatic Solid Tumours. *ESMO Open* **2020**, *5*, No. e001081.
- (69) Elgammal, W. E.; Halawa, A. H.; Eissa, I. H.; Elkady, H.; Metwaly, A. M.; Hassan, S. M.; El-Agrody, A. M. Design, Synthesis, and Anticancer Evaluation of N-Sulfonylpiperidines as Potential VEGFR-2 Inhibitors. *Apoptotic Inducers. Bioorg. Chem.* **2024**, *145*, No. 107157.
- (70) Żurański, A. M.; Martinez Alvarado, J. I.; Shields, B. J.; Doyle, A. G. Predicting Reaction Yields via Supervised Learning. *Acc. Chem. Res.* **2021**, *54*, 1856–1865.
- (71) Kim, Y.; Jung, H.; Kumar, S.; Paton, R. S.; Kim, S. Designing Solvent Systems Using Self-Evolving Solubility Databases and Graph Neural Networks. *Chem. Sci.* **2024**, *15*, 923–939.
- (72) Lundberg, S.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *arXiv* **2017**. DOI: [10.48550/arXiv.1705.07874](https://doi.org/10.48550/arXiv.1705.07874)
- (73) Luchini, G.; Paton, R. S. Bottom-Up Atomistic Descriptions of Top-Down Macroscopic Measurements: Computational Benchmarks for Hammett Electronic Parameters. *ACS Phys. Chem. Au* **2024**, *4*, 259–267.
- (74) Spearman, C. The Proof and Measurement of Association between Two Things. *Int. J. Epidemiology* **2010**, *39*, 1137–1150.

**CAS BIOFINDER DISCOVERY PLATFORM™**

# PRECISION DATA FOR FASTER DRUG DISCOVERY

CAS BioFinder helps you identify targets, biomarkers, and pathways

**Unlock insights**

**CAS** A division of the American Chemical Society