JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

Using Data Science To Guide Aryl Bromide Substrate Scope Analysis in a Ni/Photoredox-Catalyzed Cross-Coupling with Acetals as Alcohol-Derived Radical Sources

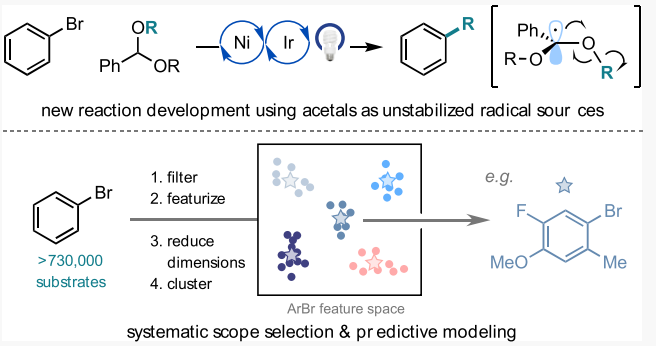
Stavros K. Kariofillis, Shutian Jiang,[§] Andrzej M. Żurański,[§] Shivaani S. Gandhi,[‖] Jesus I. Martinez Alvarado,[‖] and Abigail G. Doyle*

Cite This: J. Am. Chem. Soc. 2022, 144, 1045−1055

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Ni/photoredox catalysis has emerged as a powerful platform for C(sp^2)−C(sp^3) bond formation. While many of these methods typically employ aryl bromides as the C(sp^2) coupling partner, a variety of aliphatic radical sources have been investigated. In principle, these reactions enable access to the same product scaffolds, but it can be hard to discern which method to employ because nonstandardized sets of aryl bromides are used in scope evaluation. Herein, we report a Ni/photoredox-catalyzed (deutero)methylation and alkylation of aryl halides where benzaldehyde di(alkyl) acetals serve as alcohol-derived radical sources. Reaction development, mechanistic studies, and late-stage derivatization of a biologically relevant aryl chloride, fenofibrate, are presented. Then, we describe the integration of data science techniques, including DFT featurization, dimensionality reduction, and hierarchical clustering, to delineate a diverse and succinct collection of aryl bromides that is representative of the chemical space of the substrate class. By superimposing scope examples from published Ni/photoredox methods on this same chemical space, we identify areas of sparse coverage and high versus low average yields, enabling comparisons between prior art and this new method. Additionally, we demonstrate that the systematically selected scope of aryl bromides can be used to quantify population-wide reactivity trends and reveal sources of possible functional group incompatibility with supervised machine learning.



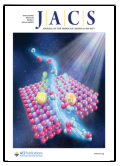
INTRODUCTION

Although a relatively new field, Ni/photoredox catalysis has seen widespread development as a general approach to C(sp^2)−C(sp^3) bond formation in chemical synthesis. In this dual catalytic strategy, a photoredox catalyst supplies substrate radicals as coupling partners to Ni and/or modulates Ni's oxidation state in catalysis.[1] Since the substrate radicals can be generated from stable, feedstock chemicals under mild conditions, a high level of functional group tolerance is often associated with these methods.[2] The advances accessible with this strategy have been particularly enabling in medicinal chemistry, where introduction of C(sp^3), rather than C(sp^2), fragments onto arenes is a well-established strategy for preparing architecturally complex molecules with improved receptor/ligand complementarity and solubility.[3]

The majority of Ni/photoredox methodologies involve the coupling of aryl bromides with aliphatic radical precursors, such as carboxylic acids,[4a] alkyl halides,[4b] trifluoroborate salts,[4c] oxalates,[4d] silicon catechols,[4e] 1,4-dihydropyridines,[4f] and C(sp^3)−H bonds[4g] (Figure 1A). In principle, these reactions afford access to the same product structures, but it is difficult to compare among the methods or to understand if there are general reactivity trends because each method features a different collection of aryl bromides in their substrate scope tables.[5] Additionally, limitations of the methods (so-called "negative" results) are often not reported. This has made it challenging for chemists to discern how well a method will translate to a new reaction partner or to select among methods to adopt.

New strategies, and a community-wide effort, are needed to address these shortcomings in substrate scope selection and reporting.[6] Toward the aim of studying functional group compatibility, Glorius and co-workers have advanced a "robustness screen," allowing for expedient examination of diverse additives for reaction inhibition or additive decomposition (Figure 1B).[7] This protocol assesses an important

Received: November 18, 2021
Published: January 5, 2022



https://doi.org/10.1021/jacs.1c12203
J. Am. Chem. Soc. 2022, 144, 1045−1055

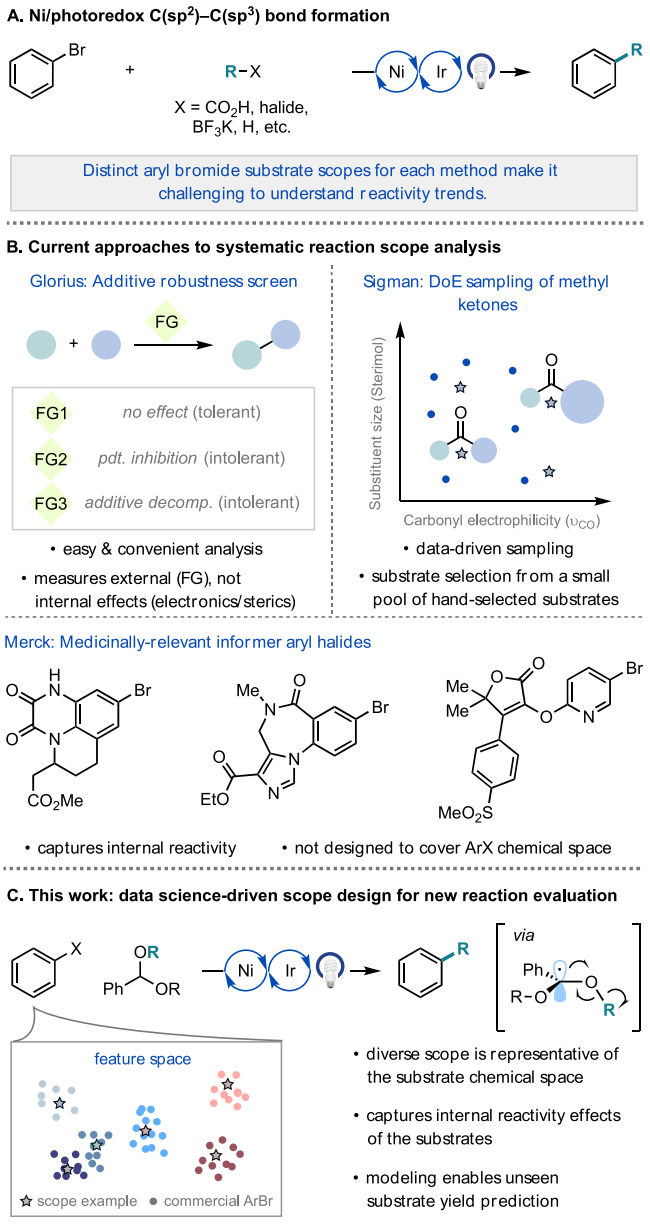


Figure 1. Challenges and new approaches to surveying reactivity from scope studies.

aspect of generality, but it does not capture the impact of internal steric or electronic effects on the reactivity of a substrate. To this end, scientists at Merck have pioneered the use of informer libraries of complex, drug-like molecules for testing the medicinal applicability of cross-coupling methodologies.[8] While highly enabling, these informer libraries were selected to represent the physicochemical space of marketed drugs rather than the chemical space of a substrate class. Moreover, due to the complexity of the informer molecules, they have yet to be broadly adopted by academic laboratories for reaction discovery.[8b] In a retrospective study, Sigman and co-workers reported a quantitative approach to substrate selection for asymmetric hydrogenation of ketones, identifying 32 ketones to span a 52-member library using Design of Experiments precepts and featurization through hand selection of two mechanistically relevant descriptors.[9] Adaptation of the workflow to accommodate the use of larger populations of substrates and higher dimensional featurization could be highly enabling and help mitigate selection bias. High-throughput experimentation (HTE) has also been used in a retrospective sense to evaluate substrate generality across synthetic methods, although the ability to gain similar information directly from reaction development manuscripts would be of immediate value to the synthetic community.[10]

To address these goals, and inspired by recent development in machine learning in chemistry,[11] we envisioned a general and quantitative scope selection workflow informed by studies in chemoinformatics and molecular library design,[12] including for informing the selection of bioactive molecules from inactive molecules,[13] virtual screening of drug-like molecules based on structure or their interaction with known protein targets,[14] and the generation of additive[15] and ligand[16] libraries for transition metal-catalyzed processes. The scope selection workflow proceeds by DFT featurization of a large virtual library of commercial substrates using our recently developed software package, auto-qchem,[17] followed by dimensionality reduction with Uniform Manifold Approximation and Projection (UMAP) and hierarchical clustering to arrive at a diverse and systematically selected substrate scope.

Herein, we describe the development, application, and assessment of this workflow for aryl bromide scope evaluation in literature examples of Ni/photoredox C(sp^2)–C(sp^3) coupling, as well as for a new Ni/photoredox method in which we show that acetals can activate methyl, primary, and secondary alcohols as aliphatic radical precursors for cross-coupling (Figure 1C). Alcohols are attractive cross-coupling partners owing to their abundance, low cost and toxicity, and functional group compatibility.[18] However, use of alcohols as aliphatic radical precursors in Ni/photoredox catalysis has seen limited development due to their high C(sp^3)–O bond strengths. Zuo and co-workers have reported C(sp^3)–CH_2OH activation of free alcohols in photocatalytic Ni/Ce cross-coupling.[19a] Alternatively, the MacMillan and Martin groups have pursued masked alcohol derivatives for Ni/photoredox cross-coupling, including oxalates[4d] and phthalimides.[19b] These methods afford access to stabilized aliphatic radicals in cross-coupling, whereas those that access methyl or primary unactivated radicals from alcohols are less developed. Very recently, Ni-catalyzed coupling reactions of free alcohols using phosphine[20a,b] and *N*-heterocyclic carbene mediators[20c] for C(sp^3)–O bond activation have been reported.

Our lab has demonstrated that another alcohol-based coupling partner, trimethyl orthoformate, can serve as a source of high-energy methyl radicals upon β-scission of its tertiary carbon-centered radical.[21] While trimethyl orthoformate is an abundant and functional group-tolerant methyl radical source, other aliphatic orthoformates are less accessible. The method also required solvent-quantity orthoformate, limiting its application as a general C(sp^2)–C(sp^3) cross-coupling method. In contrast, we anticipated that the facile synthesis of benzaldehyde di(alkyl) acetals from benzaldehyde and aliphatic alcohols would render them a promising source of aliphatic radicals for new reaction development.[22]

This article is organized as follows: first, we discuss reaction optimization and our studies aimed at elucidating the mechanism of the coupling reaction. Second, we describe a synthetic application of the method via late-stage alkylation of a biologically relevant aryl chloride. We then discuss our approach to visualizing the chemical space of aryl bromides and superimpose the reactivity of aryl bromides reported in the Ni/photoredox literature onto this space. Fifteen aryl bromides

https://doi.org/10.1021/jacs.1c12203
J. Am. Chem. Soc. 2022, 144, 1045−1055

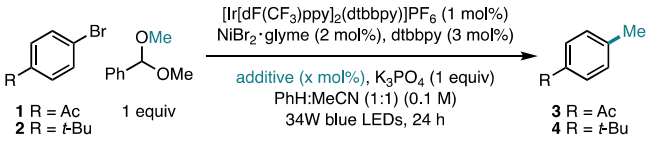
that maximally cover this substrate space are evaluated for the new synthetic method. Finally, we assess the ability of this data science-generated aryl bromide scope to capture reactivity trends via supervised machine learning.

RESULTS AND DISCUSSION

Reaction Optimization. While developing a methylation protocol with trimethyl orthoformate, we found that acetals could also serve as sources of methyl radical, but this approach required solvent quantity of the acetal to achieve a synthetically useful yield. In these reactions, activation of the acetal was mediated by chlorine radicals generated upon photoelimination from a Ni catalyst.[21] In developing a new method that employed acetals as stoichiometric coupling partners, we evaluated alternative hydrogen atom transfer (HAT) catalysts and strategies.[23] DFT calculations revealed that benzaldehyde dimethyl acetal possesses a sufficiently weak tertiary C−H bond strength (bond dissociation free energy (BDFE) = 77.0 kcal/mol)[24] for abstraction by a variety of HAT agents. These computations also revealed that the primary C−H bonds of benzaldehyde dimethyl acetal are significantly stronger (BDFE = 87.6 kcal/mol),[24] indicating high selectivity for the weaker tertiary benzylic bond was possible.

The methylation of 4′-bromoacetophenone **1** with benzaldehyde dimethyl acetal (1.5 equiv) was first investigated. Using $[Ir[dF(CF_3)ppy]_2(dtbbpy)]PF_6$ (1 mol %) as a photocatalyst, $NiBr_2$·glyme (10 mol %) and 4,4′-di-*tert*-butylbipyridine (dtbbpy) (15 mol %) as the cross-coupling catalyst, quinuclidine (15 mol %) as a HAT catalyst, and K_3PO_4 (1 equiv) in a 1:1 mixture of benzene:acetonitrile under visible light irradiation afforded **3** in 75% yield (Table 1, entry 1). We were delighted to find that reaction performance was not affected by reducing the acetal loading to 1 equiv, in addition to lowering the catalyst and ligand loadings to 2 mol % $NiBr_2$·glyme and 3 mol % dtbbpy. Under these conditions, methylation proceeded in 74% yield (Table 1, entry 2).

Table 1. Conditions Evaluation for Aryl Methylation



Entry	R	deviation	additive	% Yield[a]
1	Ac	10 mol% Ni; 15 mol% dtbbpy; 1.5 equiv acetal	quinuclidine (15 mol%)	75%
2	Ac	none	quinuclidine (15 mol%)	74%
3	**Ac**	**none**	**none**	**82%**
4	Ac	no Ni/dtbbpy	none	0%
5	Ac	no Ir	none	0%
6	Ac	no blue light	none	0%
7	*t*-Bu	none	none	9%
8	*t*-Bu	25% light intensity	TBABr (25 mol%)	78%
9	***t*-Bu**	**0.4 mmol scale; 25% light intensity**	**TBABr (25 mol%)**	**89%**

[a]Reactions performed on 0.1 mmol scale with 1,3,5-trimethoxybenzene added as an external standard (GC yield).

When performing control reactions, we found that omission of quinuclidine as a HAT catalyst resulted in a small boost in yield to 82% (Table 1, entry 3), suggesting that bromide, present from the aryl bromide or Ni precatalyst, can mediate HAT with the acetal upon oxidation to bromine radical (*vide infra*).[4b,25] This proposal is consistent with the stronger bond strength of H−Br over the tertiary C−H bond of the acetal (H−Br BDFE = 80.8 kcal/mol).[24] Independent omission of Ni/dtbbpy, photocatalyst, and visible light irradiation resulted in no cross-coupling (Table 1, entries 4−6).

Despite finding conditions for high-yielding methylation with an electron-deficient aryl bromide, methylation of the more electron-rich 1-bromo-4-*tert*-butylbenzene **2** proceeded in only 16% yield under the same conditions (Table 1, entry 7). The mass balance for this reaction could be traced to unconverted aryl bromide and protodehalogenation of **2** to *tert*-butylbenzene. After a survey of HAT catalysts and bromide additives (see Supporting Information (SI) for more details), we found that employment of exogenous bromide source tetrabutylammonium bromide (TBABr) in 25 mol % furnished the methylated product in 78% yield (Table 1, entry 8). Critical to the success of this reaction was reduction of the light intensity to 25%, which minimized protodehalogenation entirely. This modification enhances reaction yield, but also simplifies the purification process, as separation of methylated arenes from their unmethylated analogs can be challenging. A common limitation of photoredox reactions is their scalability, often due to issues of light penetration;[26] however, scale-up of this reaction to 0.4 mmol provided an additional boost in yield to 89% (Table 1, entry 9). Notably, the product derived from C−H abstraction at the primary site of benzaldehyde dimethyl acetal was never observed.

Mechanistic Investigations. Having identified conditions that enable high-yielding methylation of aryl bromides possessing varying electronic properties, we then sought to study the reaction mechanism. Our proposal for the generation of methyl radical requires stoichiometric formation of a β-scission byproduct possessing a strong C−O π bond (methyl benzoate, **7**).[22,27] In all reactions, this ester byproduct can be tracked in a nearly 1:1 ratio relative to the methylated product (Figure 2A). Cyclopropylmethyl radical ring opening from the corresponding acetal to afford **8** presented additional evidence for the intermediacy of organic radical species (Figure 2B).

We next sought to understand the nature of the HAT step. According to the mechanistic proposal, bromide undergoes oxidation to a bromine radical ($E_{1/2}[Br^-/Br\cdot]$ = +0.80 V vs SCE in DCE; $E_{1/2}[*Ir^{III}/Ir^{II}]$ = +1.21 V vs SCE in MeCN)[4b] to mediate HAT with the tertiary C−H bond of the acetal. To evaluate the feasibility of the bromine radical undergoing HAT under the coupling conditions, we performed a reaction with benzaldehyde dimethyl acetal and dimethyl maleate using 2 mol % photocatalyst and 1 equiv of TBABr. This reaction produced **9**, the product of methyl radical trapping by dimethyl maleate, in 31% yield, while methyl benzoate **7** was formed in 66% yield (Figure 2C). When TBABr was omitted from the reaction, neither methyl radical incorporation nor the ester byproduct was observed.

Halide identity studies were next explored to further understand the source and role of the bromine radical. These studies were initiated with aryl chloride **10** using $NiBr_2$·glyme as a precatalyst with and without the addition of catalytic TBABr (Table 2, entries 1−2). In both cases, methylation of the aryl chloride proceeded in >80% yield.

https://doi.org/10.1021/jacs.1c12203
J. Am. Chem. Soc. 2022, 144, 1045−1055

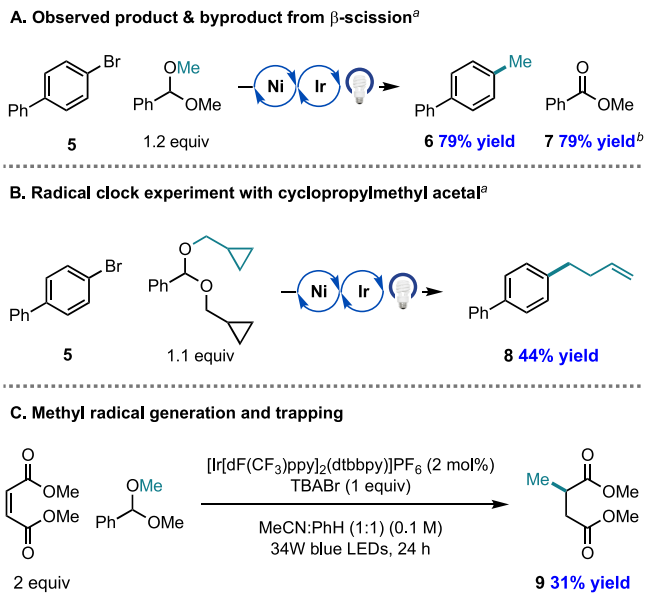
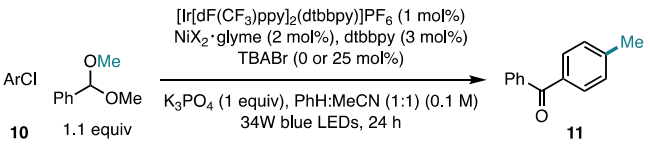


Figure 2. Selected mechanistic studies. [a]Conditions shown in Table 2 entry 9. [b]Yield relative to acetal.

Table 2. Role of Halide in the Reaction[a]



Entry	Ni source	TBABr	% Yield
1	$NiBr_2$·glyme	0 mol%	81%
2	$NiBr_2$·glyme	25 mol%	85%
3	$NiCl_2$·glyme	0 mol%	<1%
4	$NiCl_2$·glyme	25 mol%	79%

[a]ArCl = 4-chlorobenzophenone. Reactions performed on 0.1 mmol scale with 1,3,5-trimethoxybenzene as external standard (^{1}H NMR yield).

However, when $NiCl_2$·glyme was employed as the precatalyst, thereby removing all sources of bromide from the reaction, methylation proceeded in <1% yield, likely because the photocatalyst cannot oxidize chloride to a chlorine radical (Table 2, entry 3).[28,29] Reactivity was restored upon introduction of TBABr to reactions employing $NiCl_2$·glyme, providing the methylated product in 79% yield (Table 2, entry 4). These experiments demonstrate that catalytic bromide present from the Ni precatalyst is sufficient to produce high yields of methylated arenes.

An alternative pathway that is consistent with these data is direct oxidation of benzaldehyde dimethyl acetal followed by deprotonation to generate the tertiary radical of benzaldehyde dimethyl acetal. However, Stern–Volmer quenching studies revealed that the acetal does not quench the excited state of the photocatalyst; in contrast, TBABr undergoes rapid quenching. These studies also showed that TBABr undergoes more rapid quenching than (dtbbpy)Ni(*o*-Tol)(Br) (*o*-Tol = *ortho*-tolyl) in the reaction solvent system (1:1 mixture of PhH:MeCN), providing support for bromide oxidation over a bromine photoelimination pathway from Ni.[29,30]

On the basis of these studies, a catalytic cycle is shown in Figure 3. Oxidative addition of Ni(0) complex **12** into an aryl

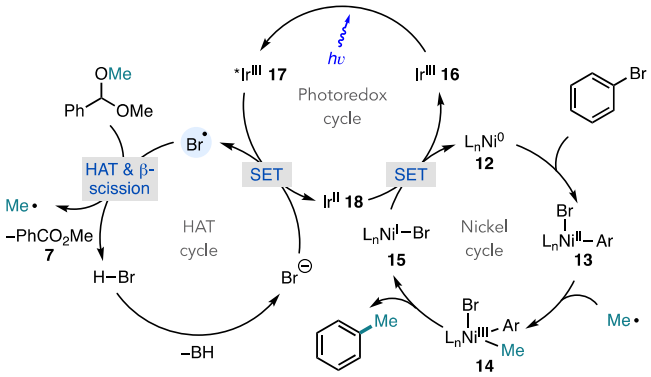


Figure 3. Catalytic cycle for aryl methylation from acetals.

halide produces a Ni(II) aryl halide intermediate (**13**). Irradiation of the iridium(III) photocatalyst **16** affords the highly oxidizing, long-lived *Ir(III) triplet excited state (**17**), which is capable of oxidizing bromide for the production of a bromine radical. Bromine then mediates selective HAT with the tertiary C–H bond of benzaldehyde dimethyl acetal, which upon *β*-scission affords a methyl radical and methyl benzoate **7**. The former species is trapped by **13**, generating Ni(III)-(Ar)(Me) (**14**) from which a facile reductive elimination forges a $C(sp^2)$–$C(sp^3)$ bond and Ni(I) complex **15**. The reduced state of the photocatalyst (**18**) can then reduce Ni(I) to regenerate both the Ir(III) and Ni(0) catalysts.

Late-Stage Application of the Methodology. After exploration of the reaction mechanism, we examined whether acetals can serve more generally as low molecular weight radical sources. This exploration was initiated with late-stage pharmaceutical product fenofibrate, which possesses an aryl chloride. Notably, aryl chlorides represent the most inexpensive, abundant, and biologically relevant aryl halide. Thus, we aimed to demonstrate that a library of alkylated derivatives of this substrate class could be prepared on the basis of the results in Table 2 which revealed that methylation of an aryl chloride is possible in the presence of a catalytic bromide additive.[31] Upon application of the reaction conditions shown in Table 2, entry 1, methylation of fenofibrate proceeded in 75% yield (Figure 4, **19**).

While aryl methylation is an established strategy for rendering compounds with improved binding affinity, bioavailability, and metabolic stability, the ability to install a variety of aliphatic groups onto a late-stage compound would be valuable for enabling rapid access to alkylated analogs.[32] Gratifyingly, employment of benzaldehyde diethyl acetal in place of the methyl analog provided ethylated product **20** in 84% yield. While benzaldehyde dimethyl and diethyl acetals are commercially available, other benzaldehyde dialkyl acetals were easily synthesized on multigram scale upon reaction of benzaldehyde, an aliphatic alcohol, and 5 Å molecular sieves at room temperature. Installation of primary aliphatic groups, including *n*-propyl (**21**), *n*-pentyl (**22**), and isoamyl (**23**) groups from the corresponding acetals proceeded in high yields. Additionally, we were excited to find that deuteromethylation of fenofibrate was possible in 78% yield (**24**). Limited methods exist to install deuteromethyl groups in cross-coupling, although iodomethane-d_3 (CD_3I),[33a] trideuterio-

https://doi.org/10.1021/jacs.1c12203
J. Am. Chem. Soc. 2022, 144, 1045–1055

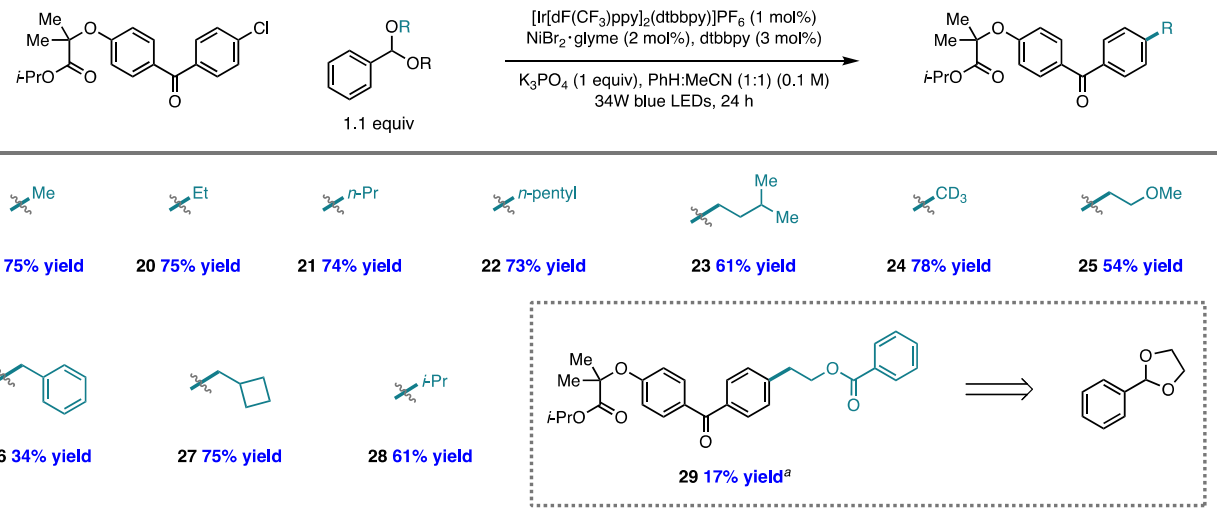


Figure 4. Late-stage functionalization of fenofibrate from benzaldehyde di(alkyl) acetals. Yields are isolated yields on 0.4 mmol scale. [a]Reaction performed with 25 mol % TBABr.

methyl *p*-toluenesulfonate (CD_3OTs),[33b] and dimethyl-d_6 carbonate ($CO(OCD_3)_2$)[33c] have previously found application as deuteromethyl sources. Acetals present a functional-group-tolerant and easily synthesized alternative deuteromethyl source derived from methanol-d_4.

Because alkyl halides containing β-heteroatoms are often not stable and may be prone to nucleophilic displacement of the halide, we also sought to install aliphatic groups containing heteroatoms, including the 2-methoxyethyl group to furnish **25**. While benzylation of fenofibrate proceeded in 35% yield (**26**), installation of the cyclobutylmethyl group provided **27** in 75% yield,[34] demonstrating that the diminished yield of the former product is likely a consequence of competitive abstraction at the secondary benzylic sites of the acetal or product rather than a steric effect. Indeed, isopropylation of fenofibrate afforded **28** in 61% yield, showing that coupling to secondary aliphatic groups is possible. Finally, we found that cyclic acetal 2-phenyl-1,3-dioxolane underwent HAT at the tertiary C−H bond, followed by ring opening and trapping of the resulting radical species to produce **29** in 17% yield.

Data Science-Driven Aryl Bromide Scope Studies and Analysis of the Chemical Space Coverage. Alkylation reactions with fenofibrate demonstrated that acetals can serve as modular sources of aliphatic radicals. Since the reaction worked with aryl bromides and chlorides, but because aryl bromides are much more commonly employed in Ni/photoredox catalysis (>100 Ni/photoredox methods published with aryl bromides versus <10 with aryl chlorides), we selected aryl bromides for scope studies to evaluate the generality of this method and to best compare to prior art.[21b] We aimed to study how well the aryl bromide chemical space has been covered in Ni/photoredox cross-coupling, as well as to use data science to generate a maximally diverse set of commercial aryl bromides.

To initiate our studies, we turned to a Reaxys structure search of aryl bromides, which provided >730,000 substrates (Figure 5). These aryl bromides were then filtered based on molecular weight (<400 amu), commercial availability, available spectral data, and functionality (see SI for full details), which provided us with an ∼2600-member data set to analyze aryl bromide chemical space. Because we filtered the aryl bromides on the basis of functionality known to be

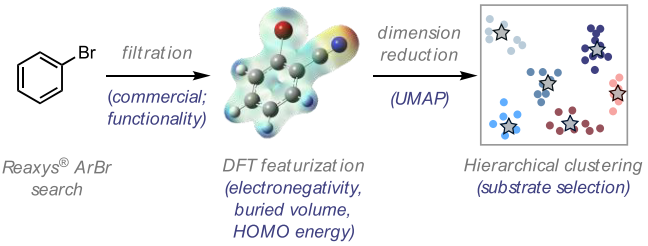


Figure 5. Workflow for construction of the chemical space and substrate selection via unsupervised learning.

incompatible with Ni/photoredox catalysis (e.g., iodides were omitted due to their preferential oxidative addition to Ni over bromides; nitro groups were omitted due to their facile reduction by the photocatalyst), this aryl bromide data set is not universally applicable, but instead designed to serve the current method. Nevertheless, the workflow and open-source code is generally applicable to other methods and substrate classes. It is also flexible, affording the ability to adapt and update as the chemical space of a substrate class changes. Indeed, we have provided the annotated code in order to make this approach accessible to the community (see SI Sections XIV and XV).

Next, we pursued molecular featurization to quantify the similarity of the aryl bromides with the goal of visualizing the chemical space of the substrate class and systematically selecting a diverse and representative scope. Several approaches were considered, including molecular fingerprints,[35] Mordred descriptors,[15,36] and DFT featurization. Neither fingerprints nor Mordred features fully capture the electronic and steric properties of aryl bromides likely relevant to their activation by oxidative addition, a shared elementary step in Ni/photoredox coupling reactions with these substrates. Accordingly, we anticipated that DFT-based featurization, which captures the underlying physical organic properties (electronegativity, buried volume, etc.), would be most appropriate for capturing reactivity trends[37] (see SI Section IX for a broader discussion and comparison between DFT and Mordred featurization).

We therefore performed DFT calculations and featurization using our previously published auto-qchem code[17] for the set

https://doi.org/10.1021/jacs.1c12203
J. Am. Chem. Soc. 2022, 144, 1045−1055

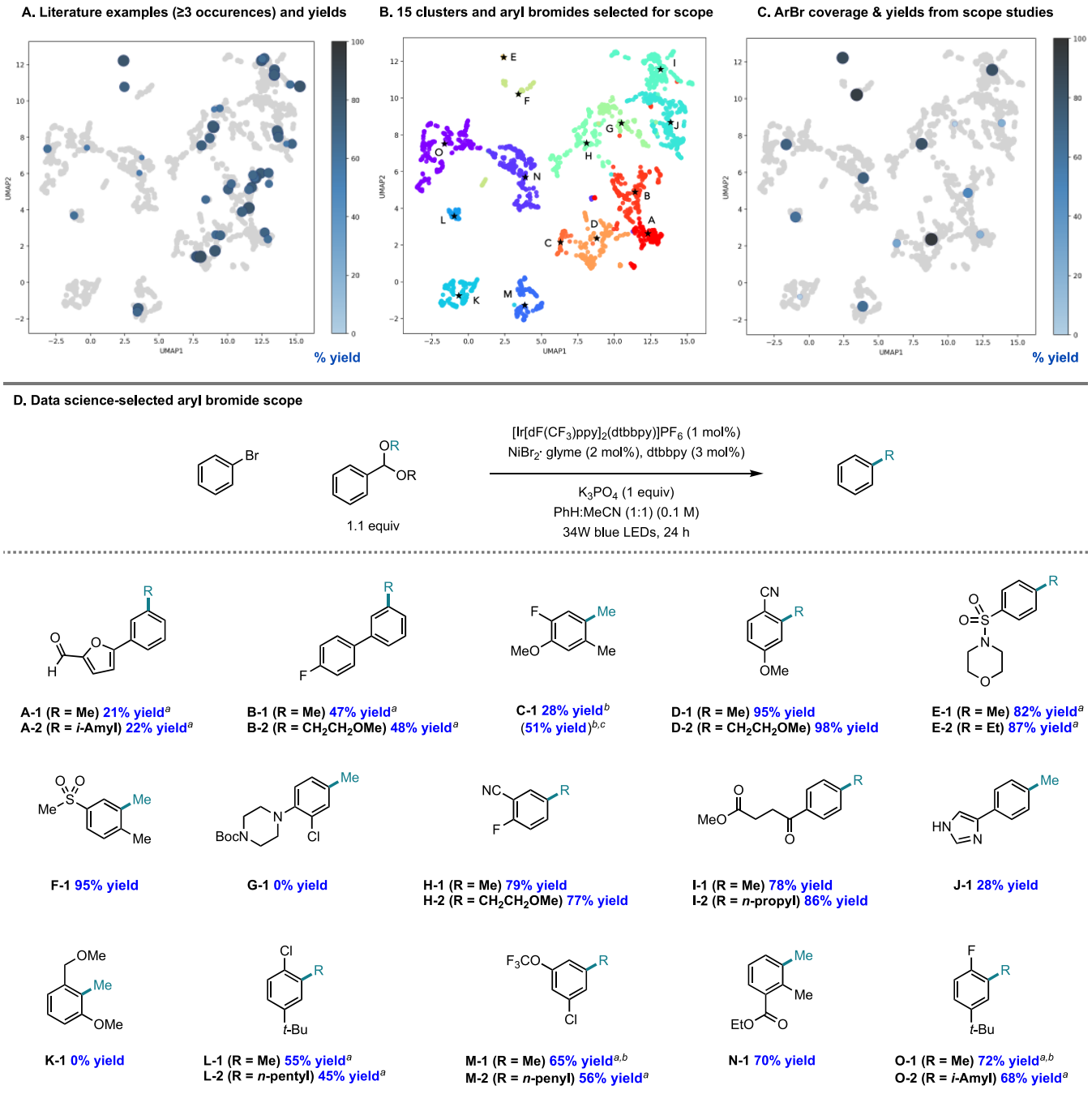


Figure 6. Evaluation of the literature coverage, clustering of the chemical space for scope selection, and scope studies (0.4 mmol scale). [a]25 mol % TBABr added. [b]^{1}H NMR yield. [c]5 mol % Ni/7.5 mol % dtbbpy.

of ~2600 aryl bromides. With auto-qchem, a system that is connected to a computing cluster and a bookkeeping database, calculations of this size complete within a few days. The calculations delivered 168 features that consist of physical properties (HOMO/LUMO energies, dipole moment, etc.), as well as partial charges and NMR shifts on the bromide and six aromatic carbon atoms (C6) (see SI Table S22 for the complete list of features). Notably, the atoms of peripheral functional groups are not directly featurized; only their impact on the common core atoms (six aromatic carbons and bromine) and the whole molecule is accounted for (see Section IX part E in the SI for an alternative featurization that includes heteroaryl bromides).[38]

Next, we conducted a literature survey of aryl bromides employed in Ni/photoredox methodologies. Our literature search produced 116 papers (see SI for details),[39] from which the yield and the frequency with which each aryl bromide appeared in substrate scope tables or Supporting Information documents were recorded. With the molecular featurization at hand, we used Uniform Manifold Approximation and Projection (UMAP)[40] to reduce the featurization to two dimensions for chemical space visualization (for alternative visualization of chemical space using principal components analysis (PCA), see SI Figure S11). We then overlaid all aryl bromides that appeared *three or more* times in the literature on the chemical space (Figure 6A). Generally, examples from the literature cover the right half of the chemical space well, while

https://doi.org/10.1021/jacs.1c12203
J. Am. Chem. Soc. 2022, 144, 1045−1055

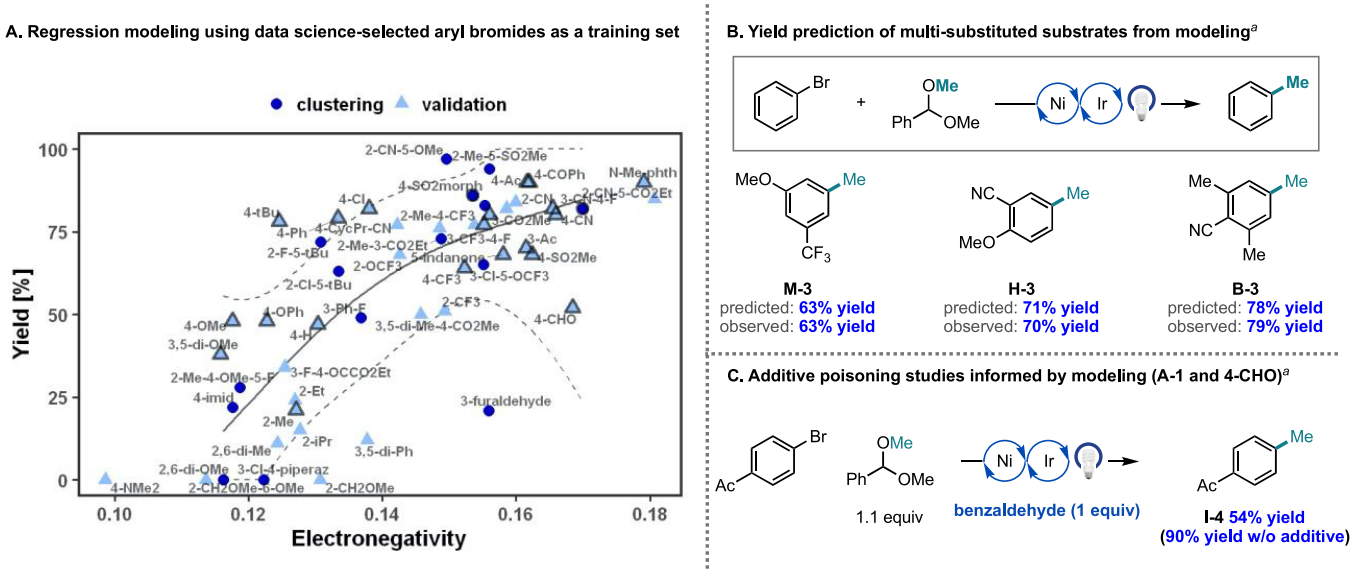


Figure 7. (A) Methylation yield as a function of aryl bromide electronegativity. The model and a 2 standard error band are trained only with the 15 data science-selected aryl bromides. Aryl bromides that appear in our Ni/photoredox literature search at least 3 times are outlined. (B) Utility of the model as a predictive tool for unseen substrate yield. (C) Additive poisoning studies. [a]Conditions shown in Figure 6D (^{1}H NMR yield using 1,3,5-trimethoxybenzene as an external standard).

the left half is sparsely covered. By virtue of using reported yields from the literature, the average yields shown are likely to be inflated, as unsuccessful substrates are seldom reported. Nevertheless, a trend is observed that the more sparsely covered chemical space corresponds to relatively lower yielding aryl bromide substrates. The five substrates covering this left region of the space all possess *ortho* substitution, revealing that steric hindrance represents a limitation in the overall methodology (see SI Figure S28). Another trend that is apparent in surveying the full space is that more electron-donating aryl bromides provide lower yields than aryl bromides with electron-withdrawing substituents at the same position.[41]

To generate a diverse substrate scope for the alkylation of aryl bromides from acetals, we sought to cluster the chemical space for substrate selection. Although no universal definition of clustering exists, it is a statistical tool used to arrange large quantities of multivariate data into natural groups.[42] We chose agglomerative hierarchical clustering with the Ward connectivity criterion.[43] This algorithm starts with each aryl bromide as a separate cluster and then iteratively connects clusters such that the total within-cluster variance of all clusters is minimized until a desired number of clusters is reached. We optimized the clustering workflow by maximizing the average Silhouette score,[44] a clustering quality index, with respect to an experimentally tractable number of clusters in the range of 3−40, and the level of dimensionality reduction with either UMAP or PCA. Silhouette score values lie between −1 and 1: large, positive values indicate that a sample matches its assigned cluster well, while at the same time does not match any other cluster. Optimization revealed that UMAP dimensionality reduction is largely superior to PCA (average silhouette score up to ∼0.5, compared to ∼0.3 for PCA), and a stable and broad maximum silhouette score was reached at about 15 clusters (average silhouette score = 0.53) with UMAP. In addition, the silhouette score varies widely as a function of the dimension for PCA, while little dependence on the dimension of UMAP is observed (see SI Figure S10). Ultimately, we selected 10-dimensional UMAP and 15 clusters as our optimal parameters. From the 15 chemically diverse clusters, we selected the center-most molecule per cluster for experimental evaluation, as indicated in Figure 6B.

To a chemist's eye, the data science-driven scope features a diverse array of functional groups, including ester, nitrile, chloride, ether, trifluoromethoxy, sulfonyl, morpholine, imidazole, and furan functionality (Figure 6D). Substituents on the aryl bromide are present on all positions of the aryl ring through mono-, di-, and trisubstitution patterns. The frequency of multisubstitution in the data science-generated scope contrasts the literature scope, which is dominated by monosubstituted aryl bromides. In fact, Clusters **C** and **K**, both of which contain highly substituted aryl bromides, were not represented in the overlay with aryl bromides from the literature (Figure 6A). The data science scope also captures a variety of steric features: among *ortho* substituents, fluoride (**O-1**), chloride (**L-1**), methyl (**C-1, F-1, N-1**), and nitrile (**D-1**) groups are represented. In addition, an *ortho,ortho*-disubstituted aryl bromide that would produce **K-1** is present in the scope.

Subjecting this collection of aryl bromides to the optimized Ni/photoredox coupling conditions with acetals demonstrated quite broad tolerance of the methodology (Figure 6D). Methylation reactions were performed for each substrate. Additional alkylation reactions were also conducted and showed that methyl and primary aliphatic acetals generally perform similarly in the cross-coupling method. The products are labeled with a letter (**A**−**O**) that corresponds to their cluster indicated in Figure 6B. With the exception of **G-1** and *ortho-,ortho*-disubstituted **K-1**, all aryl bromides underwent productive alkylation to generate at least 21% yield of the desired product. The analytical methylation yields of these products and their distribution over the chemical space are shown in Figure 6C. Figure 6A and 6C present overall similar reactivity profiles, suggesting that this method compares favorably to the collection of methods published in the literature. In particular, this method appears to be more tolerant of steric hindrance than examples reported in the

https://doi.org/10.1021/jacs.1c12203
J. Am. Chem. Soc. 2022, 144, 1045−1055

literature when examining reaction performance across clusters characteristic of substrates possessing *ortho*-substitution (Clusters **L**, **N**, and **O**).

Juxtaposition of the reactivity and substrate space of this method with other Ni/photoredox methods offers an opportunity to identify global limitations and inspire new reaction development. Preliminarily, we have found that the yield of **C-1**, derived from a trisubstituted aryl bromide, could be boosted from 28% to 51% yield by increasing the cross-coupling catalyst loading to 5 mol % $NiBr_2$·glyme and 7.5 mol % dtbbpy. However, further reaction development will likely be required to identify conditions that enable coupling of substrates in Cluster **K** that are generally both electronically deactivated and sterically encumbered.

Yield Regression from the Aryl Bromide Scope Studies. The intention of using data science techniques for substrate scope generation was to maximally cover the aryl bromide chemical space using a conserved number of aryl bromides. Given the diversity of the aryl bromides, as well as the range of yields observed when applied to the cross-coupling methodology, we wondered if this small set of aryl bromides could provide predictive generalizations on the performance of unseen aryl bromides. To do so, we evaluated univariate relationships with supervised machine learning. Using a combination of Boruta[45] and mRMR[46] feature selection methodologies on our data set of 15 aryl bromides from the clustering (**A-1** to **O-1**, using ^{1}H NMR yields from the methylation of each substrate), we found that the *electronegativity* (χ) of the aryl bromide, defined as the average of the HOMO and LUMO energies, serves as a predictive feature.[47] This finding is in line with chemical intuition, where more electron-rich substrates undergo a more challenging oxidative addition to Ni.[41,48] We then pursued training and validation of the electronegativity model using a generalized additive model (GAM).[49] The validation set is comprised of 37 additional aryl bromides present in our laboratory that were subjected to methylation using the conditions shown in Figure 6.

The regression model displayed in Figure 7A generalizes well to the validation data. The root-mean-squared-error (RMSE) is 21% and 19% in the clustering and validation data, respectively (R-squared = 0.57). By contrast, the high yield of the common literature substrates found in the validation set (outlined in gray in Figure 7A) amounts to a selection bias, making them less effective as a training set for building predictive models. Indeed, despite the literature set being larger (22 aryl bromides), a univariate regression model with these examples as a training set is worse at generalizing: training/validation RMSE of 14%/29% (see SI, Figure S19). Thus, this analysis demonstrates that, with a small and systematically diversified scope, one can provide predictive generalizations applicable to a larger population of substrates.

Despite the simplicity of the univariate model, we anticipated that it would have predictive value for multi-substituted substrates, where it is more challenging to intuit the overall electronic profile of an aryl bromide and its resulting reactivity. At the same time, multisubstitution dominates late-stage coupling examples in chemical synthesis but is severely under-represented in literature scope tables (85% of the literature aryl bromides plotted in Figure 6A are monosubstituted). To test the performance of the model on unseen substrates with complex substitution, we selected three aryl bromides possessing di- or trisubstitution with varying, nonintuitive electronic properties (Figure 7B). A 63% yield was predicted for **M-3**, which possesses *meta*-trifluoromethyl and *meta*-methoxy groups. We were pleased to find that this product was formed in 63% analytical yield under the standard reaction conditions. Two additional substrates, **H-3** possessing *meta*-nitrile and *para*-methoxy functionality, as well as **B-3** with *meta*-methyl groups and a *para*-nitrile, delivered methylated product within 1% of the predicted yield. These results suggest that the systematic substrate selection and modeling workflow is highly effective at revealing substituent impacts that affect relative reactivity (internal effects).

Nevertheless, a few substrates fall outside the 2-standard error band of the regression model in Figure 7A.[50] For example, aldehyde-containing substrates that produced **A-1** and 4-methylbenzaldehyde (**I-3**) provided noticeably lower yields than predicted, suggesting that a factor other than the internal reactivity of the substrate was overriding performance in the cross-coupling. 4-Bromobenzaldehyde belongs to Cluster **I**, one that is well covered and high-yielding in literature examples; in the methylation, two additional substrates belonging to this cluster afforded **I-1** and 4′-methylacetophenone (**I-4**) in 78 and 90% yields, respectively. Thus, we hypothesized that the aldehyde functionality was leading to reaction poisoning by an off-cycle pathway.[25a,b,51] To test this, we conducted a robustness experiment, evaluating the methylation of 4′-bromoacetophenone with 1 equiv of benzaldehyde as an additive.[7] While the reaction was not shut down by the additive, the yield of **I-4** was reduced to 54% from 90% yield (Figure 7C). This reaction was also marked by consumption of benzaldehyde (53% conversion), indicating that the reduced yields of **A-1** and **I-3** could arise from an external effect of the aldehyde functionality rather than its influence on the electronegativity of the substrate.[25a,b,51]

More broadly, this analysis shows how substituent effects can be decoupled using this workflow, with those that affect off-cycle steps (external effects) arising as outliers to the model, something that cannot be accomplished solely by using a fragment additive screen. Evaluation of only external effects on reactivity has been shown to result in false negatives and false positives.[52] As an example within this data set, robustness screening indicated that the preparation of imidazole-containing **J-1** would fail: methylation of 4-bromobenzonitrile with 1 equiv of imidazole delivered methylated product in <5% yield (compared to 80% yield without the additive). However, the data science clustering workflow showed that preparatively useful yield is possible: **J-1** was produced in 28% yield. Taken together, these data suggest that neither the robustness screening nor clustering workflow on its own is enough to sample the reactivity *and* functional group compatibility encompassed in a reaction space. However, we anticipate that these tools can be used synergistically for expedient and comprehensive modeling of internal and external effects on reactivity.

CONCLUSION

In conclusion, we have developed a $C(sp^2)-C(sp^3)$ cross-coupling reaction using acetals as sources of aliphatic coupling partners. Bromine-mediated HAT with the tertiary C−H bond of benzaldehyde di(alkyl) acetals, followed by β-scission of the resultant radical, enables the generation of carbon-centered radicals for coupling with aryl bromides and chlorides. The reaction platform works with 1:1 stoichiometry of acetal and aryl halide and installs (deutero)methyl, primary aliphatic, and

https://doi.org/10.1021/jacs.1c12203
J. Am. Chem. Soc. 2022, 144, 1045−1055

secondary aliphatic groups efficiently with high functional group tolerance. By integrating DFT featurization and data science techniques, including hierarchical clustering, we were able to design a substrate scope that is representative of the diverse chemical space of commercial aryl bromides. While 15 substrates are not enough to sample the reactivity of all regions of the space, we show that substrate selection using this workflow is better suited than traditional literature scopes for quantifying population-wide reactivity trends with supervised machine learning. Indeed, regression modeling enables population-wide yield prediction of unseen coupling partners and visually indicates possible sources of functional group incompatibility in a method. As such, we anticipate that this approach could serve as a model for standardizing scope analysis, enabling chemists to compare among methods, reduce the time and cost associated with scope evaluation, and afford literature data better suited to quantitative modeling of reactivity.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.1c12203.

Experimental procedures, spectroscopic data, aryl bromide chemical space construction and substrate selection details, information about the literature search and regression modeling (PDF)

Smiles (TXT)

■ AUTHOR INFORMATION

Corresponding Author

Abigail G. Doyle − *Department of Chemistry, Princeton University, Princeton, New Jersey 08544, United States; Department of Chemistry & Biochemistry, University of California—Los Angeles, Los Angeles, California 90095, United States;* orcid.org/0000-0002-6641-0833; Email: agdoyle@chem.ucla.edu

Authors

Stavros K. Kariofillis − *Department of Chemistry, Princeton University, Princeton, New Jersey 08544, United States; Department of Chemistry & Biochemistry, University of California—Los Angeles, Los Angeles, California 90095, United States;* orcid.org/0000-0002-5461-3190

Shutian Jiang − *Department of Chemistry, Princeton University, Princeton, New Jersey 08544, United States*

Andrzej M. Żurański − *Department of Chemistry, Princeton University, Princeton, New Jersey 08544, United States*

Shivaani S. Gandhi − *Department of Chemistry, Princeton University, Princeton, New Jersey 08544, United States; Department of Chemistry & Biochemistry, University of California—Los Angeles, Los Angeles, California 90095, United States;* orcid.org/0000-0003-1825-5450

Jesus I. Martinez Alvarado − *Department of Chemistry, Princeton University, Princeton, New Jersey 08544, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.1c12203

Author Contributions

[§]S.J. and A.M.Ż. contributed equally.

Author Contributions

[‖]S.S.G. and J.I.M.A. contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This material is based on work supported by the National Science Foundation Graduate Research Fellowship Program under Grant Number DGE-1656466 (to S.K.K.). A.M.Z. acknowledges financial support from the Schmidt DataX Fund at Princeton University made possible through a major gift from the Schmidt Futures Foundation. A.G.D. gratefully acknowledges the Princeton Innovation Fund, NIGMS (R35 GM126986) (Ni/photoredox method development), and the CCI Center for Computer Assisted Synthesis (CHE-1925607) (data science contributions) for financial support. We thank Jose Garrido Torres for helpful discussions on the data science workflow and Wendy Williams for computational support.

■ REFERENCES

(1) (a) Twilton, J.; Le, C.; Zhang, P.; Shaw, M. H.; Evans, R. W.; MacMillan, D. W. C. The Merger of Transition Metal and Photocatalysis. *Nat. Rev. Chem.* **2017**, *1*, 0052. (b) Milligan, J. A.; Phelan, J. P.; Badir, S. O.; Molander, G. A. Alkyl Carbon−Carbon Bond Formation by Nickel/Photoredox Cross-Coupling. *Angew. Chem., Int. Ed.* **2019**, *58*, 6152−6163.

(2) Cannalire, R.; Pelliccia, S.; Sancineto, L.; Novellino, E.; Tron, G. C.; Giustiniano, M. Visible light photocatalysis in the late-stage functionalization of pharmaceutically relevant compounds. *Chem. Soc. Rev.* **2021**, *50*, 766−897.

(3) Lovering, F.; Bikker, J.; Humblet, C. Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success. *J. Med. Chem.* **2009**, *52*, 6752−6756.

(4) (a) Zuo, Z.; Ahneman, D. T.; Chu, L.; Terrett, J. A.; Doyle, A. G.; MacMillan, D. W. C. Merging Photoredox Catalysis with Nickel Catalysis: Coupling of α-Carboxyl sp^3-Carbons with Aryl Halides. *Science* **2014**, *345*, 437−440. (b) Zhang, P.; Le, C.; MacMillan, D. W. C. Silyl Radical Activation of Alkyl Halides in Metallaphotoredox Catalysis: A Unique Pathway for Cross-Electrophile Coupling. *J. Am. Chem. Soc.* **2016**, *138*, 8084−8087. (c) Tellis, J. C.; Primer, D. N.; Molander, G. A. Single-Electron Transmetalation in Organoboron Cross-Coupling by Photoredox/Nickel Dual Catalysis. *Science* **2014**, *345*, 433−436. (d) Zhang, X.; MacMillan, D. W. C. Alcohols as Latent Coupling Fragments for Metallaphotoredox Catalysis: sp^3−sp^2 Cross-Coupling of Oxalates with Aryl Halides. *J. Am. Chem. Soc.* **2016**, *138*, 13862−13865. (e) Corcé, V.; Chamoreau, L.-M.; Derat, E.; Goddard, J.-P.; Ollivier, C.; Fensterbank, L. Silicates as Latent Alkyl Radical Precursors: Visible-Light Photocatalytic Oxidation of Hypervalent Bis-Catecholato Silicon Compounds. *Angew. Chem., Int. Ed.* **2015**, *54*, 11414−11418. (f) Gutiérrez-Bonet, Á.; Tellis, J. C.; Matsui, J. K.; Vara, B. A.; Molander, G. A. 1,4-Dihydropyridines as Alkyl Radical Precursors: Introducing the Aldehyde Feedstock to Nickel/Photoredox Dual Catalysis. *ACS Catal.* **2016**, *6*, 8004−8008. (g) Perry, I. B.; Brewer, T. F.; Sarver, P. J.; Schultz, D. M.; DiRocco, D. A.; MacMillan, D. W. C. Direct arylation of strong aliphatic C−H bonds. *Nature* **2018**, *560*, 70−75.

(5) This is also true of the aliphatic coupling partner, but in this case there may not be the same set of available substrates. For example, certain alkyl halides might not be available as trifluoroborate salts, and vice versa.

(6) (a) Gensch, T.; Glorius, F. The straight dope on the scope of chemical reactions. *Science* **2016**, *352*, 294−295. (b) Nguyen, T. The substrate scope conundrum. *C&EN News* **2017**, *95*, 16−18.

(7) (a) Collins, K. D.; Glorius, F. A robustness screen for the rapid assessment of chemical reactions. *Nat. Chem.* **2013**, *5*, 597−601. (b) Collins, K. D.; Glorius, F. Intermolecular Reaction Screening as a Tool for Reaction Evaluation. *Acc. Chem. Res.* **2015**, *48*, 619−627.

(c) Gensch, T.; Teders, M.; Glorius, F. Approach to Comparing the Functional Group Tolerance of Reactions. *J. Org. Chem.* **2017**, *82*, 9154−9159.
(8) (a) 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.; Krska, S. W.; Dreher, S. D. Chemistry Informer Libraries: A Chemoinformatics Enabled Approach to Evaluate and Advance Synthetic Methods. *Chem. Sci.* **2016**, *7*, 2604−2613. (b) Dreher, S. D.; Krska, S. W. Chemistry Infomer Libraries: Conception, Early Experience, and Role in the Future of Chemoinformatics. *Acc. Chem. Res.* **2021**, *54*, 1586−1596.
(9) Bess, E. N.; Bischoff, A. J.; Sigman, M. S. Designer substrate library for quantitative, predictive modeling of reaction performance. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 14698−14703.
(10) Dombrowski, A. W.; Gesmundo, N. J.; Aguirre, A. L.; Sarris, K. A.; Young, J. M.; Bogdan, A. R.; Martin, M. C.; Gedeon, S.; Wang, Y. Expanding the Medicinal Chemist Toolbox: Comparing Seven $C(sp^2)$−$C(sp^3)$ Cross-Coupling Methods by Library Synthesis. *ACS Med. Chem. Lett.* **2020**, *11*, 597−604.
(11) (a) Jorner, K.; Tomberg, A.; Bauer, C.; Sköld, C.; Norrby, P.-O. Organic reactivity from mechanism to machine learning. *Nat. Rev. Chem.* **2021**, *5*, 240−255. (b) Coley, C. W.; Eyke, N. S.; Jensen, K. F. Autonomous Discovery in the Chemical Sciences Part I: Progress. *Angew. Chem., Int. Ed.* **2020**, *59*, 22858−22893. (c) Strieth-Kalthoff, F.; Sandfort, F.; Segler, M. H. S.; Glorius, F. Machine learning the ropes: principles, applications and directions in synthetic chemistry. *Chem. Soc. Rev.* **2020**, *49*, 6154−6168. (d) de Almeida, A. F.; Moreira, R.; Rodrigues, T. Synthetic organic chemistry driven by artificial intelligence. *Nat. Rev. Chem.* **2019**, *3*, 589−604.
(12) Xu, J.; Hagler, A. Chemoinformatics and Drug Discovery. *Molecules* **2002**, *7*, 566−600.
(13) Brown, R. D.; Martin, Y. C. Use of Structure−Activity Data To Compare Structure-Based Clustering Methods and Descriptors for Use in Compound Selection. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 572−584.
(14) Mok, N. Y.; Brenk, R.; Brown, N. Increasing the Coverage of Medicinal Chemistry-Relevant Space in Commercial Fragments Screening. *J. Chem. Inf. Model.* **2014**, *54*, 79−85.
(15) Saito, M.; Kawamata, Y.; Meanwell, M.; Navratil, R.; Chiodi, D.; Carlson, E.; Hu, P.; Chen, L.; Udyavara, S.; Kingston, C.; Tanwar, M.; Tyagi, S.; McKillican, B. P.; Gichinga, M. G.; Schmidt, M. A.; Eastgate, M. D.; Lamberto, M.; He, C.; Tang, T.; Malapit, C. A.; Sigman, M. S.; Minteer, S. D.; Neurock, M.; Baran, P. S. *N*-Ammonium Ylide Mediators for Electrochemical C−H Oxidation. *J. Am. Chem. Soc.* **2021**, *143*, 7859−7867.
(16) (a) Fey, N. Lost in chemical space? Maps to support organometallic catalysis. *Chem. Cent. J.* **2015**, *9*, 38. (b) Gensch, T.; dos Passos Gomes, G.; Friederich, P.; Peters, E.; Gaudin, T.; Pollice, R.; Jorner, K.; Nigam, A.; Lindner-D'Addario, M. L.; Sigman, M. S.; Aspuru-Guzik, A. A comprehensive discovery platform for organophosphorus ligands for catalysis. 2021-04−27. *ChemRxiv*. DOI: 10.26434/chemrxiv.12996665 (accessed 2021-11-18).
(17) https://github.com/PrincetonUniversity/auto-qchem (accessed November 18, 2021).
(18) Henkel, T.; Brunne, R. M.; Müller, H.; Reichel, F. Statistical Investigation into the Structural Complementarity of Natural Products and Synthetic Compounds. *Angew. Chem., Int. Ed.* **1999**, *38*, 643−647.
(19) (a) Chen, Y.; Wang, X.; He, X.; An, Q.; Zuo, Z. Photocatalytic Dehydroxymethylative Arylation by Synergistic Cerium and Nickel Catalysis. *J. Am. Chem. Soc.* **2021**, *143*, 4896−4902. (b) Cong, F.; Lv, X.-Y.; Day, C. S.; Martin, R. Dual Catalytic Strategy for Forging sp^2−sp^3 and sp^3−sp^3 Architectures via β-Scission of Aliphatic Alcohol Derivatives. *J. Am. Chem. Soc.* **2020**, *142*, 20594−20599.
(20) (a) Li, Z.; Sun, W.; Wang, X.; Li, L.; Zhang, Y.; Li, C. Electrochemically Enabled, Nickel-Catalyzed Dehydroxylative Cross-Coupling of Alcohols with Aryl Halides. *J. Am. Chem. Soc.* **2021**, *143*, 3536−3543. (b) Chi, B. K.; Widness, J. K.; Gilbert, M. M.; Salgueiro, D. C.; Garcia, K. J.; Weix, D. J. In-Situ Bromination Enables Formal Cross-Electrophile Coupling of Alcohols with Aryl and Alkenyl Halides. *ACS Catal.* **2021**, *12*, 580−586. (c) Dong, Z.; MacMillan, D. W. C. Metallaphotoredox-enabled deoxygenative arylation of alcohols. *Nature* **2021**, *598*, 451−456.
(21) (a) Kariofillis, S. K.; Shields, B. J.; Tekle-Smith, M. A.; Zacuto, M. J.; Doyle, A. G. Nickel/Photoredox-Catalyzed Methylation of (Hetero)aryl Chlorides Using Trimethyl Orthoformate as a Methyl Radical Source. *J. Am. Chem. Soc.* **2020**, *142*, 7683−7689. (b) Kariofillis, S. K.; Doyle, A. G. Synthetic and Mechanistic Implications of Chlorine Photoelimination in Nickel/Photoredox $C(sp^3)$−H Cross-Coupling. *Acc. Chem. Res.* **2021**, *54*, 988−1000.
(22) (a) Kuhn, L. P.; Wellman, C. Reaction of *t*-Butyl Peroxide with Acetals. *J. Org. Chem.* **1957**, *22*, 774−776. (b) Hartzell, G. E.; Huyser, E. S. Generation of Methyl Radicals by Decomposition of Bibenzyl Compounds Containing α-Methoxy Substituents. *J. Org. Chem.* **1964**, *29*, 3341−3344.
(23) Capaldo, L.; Ravelli, D. Hydrogen Atom Transfer (HAT): A Versatile Strategy for Substrate Activation in Photocatalyzed Organic Synthesis. *Eur. J. Org. Chem.* **2017**, *2017*, 2056−2071.
(24) All bond strength calculations performed at the uM062x/def2tzvpp//uB3LYP/CBSB7 level of theory.
(25) Selected proposals invoking bromine radical generation in Ni/photoredox catalysis: (a) Kawasaki, T.; Ishida, N.; Murakami, M. Dehydrogenative Coupling of Benzylic and Aldehydic C−H Bonds. *J. Am. Chem. Soc.* **2020**, *142*, 3366−3370. (b) Kawasaki, T.; Ishida, N.; Murakami, M. Photoinduced Specific Acylation of Phenolic Hydroxy Groups with Aldehydes. *Angew. Chem., Int. Ed.* **2020**, *59*, 18267−18271. (c) Heitz, D. R.; Tellis, J. C.; Molander, G. A. Photochemical Nickel-Catalyzed C−H Arylation: Synthetic Scope and Mechanistic Investigations. *J. Am. Chem. Soc.* **2016**, *138*, 12715−12718. (d) Santos, M. S.; Corrêa, A. G.; Paixão, M. W.; König, B. $C(sp^3)$−$C(sp^2)$ Cross-Coupling of Alkyl Bromides and Ethers Mediated by Metal and Visible Light Photoredox Catalysis. *Adv. Synth. Catal.* **2020**, *362*, 2367−2372. (e) Huang, L.; Rueping, M. Direct Cross-Coupling of Allylic $C(sp^3)$−H Bonds with Aryl- and Vinylbromides by Combined Nickel and Visible-Light Catalysis. *Angew. Chem., Int. Ed.* **2018**, *57*, 10333−10337. (f) Shu, X.; Huan, L.; Huang, Q.; Huo, H. Direct Enantioselective $C(sp^3)$−H Acylation for the Synthesis of α-Amino Ketones. *J. Am. Chem. Soc.* **2020**, *142*, 19058−19064.
(26) Harper, K. C.; Moschetta, E. G.; Bordawekar, S. V.; Wittenberger, S. J. A Laser Driven Flow Chemistry Platform for Scaling Photochemical Reactions with Visible Light. *ACS Cent. Sci.* **2019**, *5*, 109−115.
(27) (a) Kochi, J. K. Chemistry of Alkoxy Radicals: Cleavage Reactions. *J. Am. Chem. Soc.* **1962**, *84*, 1193−1197. (b) Bacha, J. D.; Kochi, J. K. Polar and Solvent Effects in the Cleavage of *t*-Alkoxy Radicals. *J. Org. Chem.* **1965**, *30*, 3272−3278. (c) Walling, C. Some Aspects of the Chemistry of Alkoxy Radicals. *Pure Appl. Chem.* **1967**, *15*, 69−80.
(28) (a) Isse, A. A.; Lin, C. Y.; Coote, M. L.; Gennaro, A. Estimation of Standard Reduction Potentials of Halogen Atoms and Alkyl Halides. *J. Phys. Chem. B* **2011**, *115*, 678−684. (b) Shields, B. J.; Doyle, A. G. Direct $C(sp^3)$−H Cross Coupling Enabled by Catalytic Generation of Chlorine Radicals. *J. Am. Chem. Soc.* **2016**, *138*, 12719−12722. (c) Rohe, S.; Morris, A. O.; McCallum, T.; Barriault, L. Hydrogen Atom Transfer Reactions via Photoredox Catalyzed Chlorine Atom Generation. *Angew. Chem., Int. Ed.* **2018**, *57*, 15664−15669.
(29) The reaction performed under the conditions shown in Table 2 entry 3, but in a 1:1 mixture of benzaldehyde dimethyl acetal/benzene as solvent, enabled methylation in 28% yield. This result suggests that the photoelimination pathway can be turned on with an excess of the acetal coupling partner.
(30) Although more rapid quenching was observed with TBABr than Ni, MeCN can displace the bromide ligand on Ni, leading to comparable rates of quenching: Feth, M. P.; Klein, A.; Bertagnolli, H. Investigation of the Ligand Exchange Behavior of Square-Planar Nickel(II) Complexes by X-ray Absorption Spectroscopy and X-ray Diffraction. *Eur. J. Inorg. Chem.* **2003**, *2003*, 839−852.

(31) Aryl chloride methylation has previously enabled preparation of the first selective COX-2 NSAID: Sun, S.; Fu, J. Methyl-containing pharmaceuticals: Methylation in drug design. *Bioorg. Med. Chem. Lett.* **2018**, *28*, 3283−3289.
(32) (a) Barreiro, E. J.; Kümmerle, A. E.; Fraga, C. A. M. The Methylation Effect in Medicinal Chemistry. *Chem. Rev.* **2011**, *111*, 5215−5246. (b) Schönherr, H.; Cernak, T. Profound Methyl Effects in Drug Discovery and a Call for New C−H Methylation Reactions. *Angew. Chem., Int. Ed.* **2013**, *52*, 12256−12267. (c) Aynetdinova, D.; Callens, M. C.; Hicks, H. B.; Poh, C. Y. X.; Shennan, B. D. A.; Boyd, A. M.; Lim, Z. H.; Leitch, J. A.; Dixon, D. J. Installing the "magic methyl" − C−H methylation in synthesis. *Chem. Soc. Rev.* **2021**, *50*, 5517−5563.
(33) (a) Hu, L.; Liu, X.; Liao, X. Nickel-Catalyzed Methylation of Aryl Halides with Deuterated Methyl Iodide. *Angew. Chem., Int. Ed.* **2016**, *55*, 9743−9747. (b) Komeyama, K.; Yamahata, Y.; Osaka, I. Nickel and Nucleophilic Cobalt-Catalyzed Trideuteriomethylation of Aryl Halides Using Trideuteriomethyl *p*-Toluenesulfonate. *Org. Lett.* **2018**, *20*, 4375−4378. (c) Wu, Z.; Wei, F.; Wan, B.; Zhang, Y. Pd-Catalyzed *ipso*,*meta*-Dimethylation of *ortho*-Substituted Iodoarenes via a Base-Controlled C−H Activation Cascade with Dimethyl Carbonate as the Methyl Source. *J. Am. Chem. Soc.* **2021**, *143*, 4524−4530.
(34) Minor cyclobutylmethyl ring opening was observed in this reaction.
(35) (a) Senese, C. L.; Duca, J.; Pan, D.; Hopfinger, A. J.; Tseng, Y. J. 4D-Fingerprints, Universal QSAR and QSPR Descriptors. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1526−1539. (b) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742−754. (c) Pattanaik, L.; Coley, C. W. Molecular Representation: Going Long on Fingerprints. *Chem.* **2020**, *6*, 1204−1207. (d) Sandfort, F.; Strieth-Kalthoff, F.; Kühnemund, M.; Beecks, C.; Glorius, F. A Structure-Based Platform for Predicting Chemical Reactivity. *Chem.* **2020**, *6*, 1379−1390.
(36) Moriwaki, H.; Tian, Y.-S.; Kawashita, N.; Takagi, T. Mordred: a molecular descriptor calculator. *J. Cheminformatics* **2018**, *10*. DOI: 10.1186/s13321-018-0258-y.
(37) 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.
(38) Our first report here focuses on the generation of an aryl bromide scope using molecular and atomic (C6Br) features. However, this workflow can be applied to heteroaromatic bromides by changing the atomic features to the conserved C−Br bond (see SI for full details and a collection of 15 heteroaryl bromides generated from clustering, including N-, O-, and S-containing heterocycles, Figures S16−S17).
(39) Literature search performed on December 10, 2020.
(40) (a) McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. 2018-02-09. *arXiv*. https://arxiv.org/abs/1802.03426 (accessed 2021-05-15). (b) McInnes, L.; Healy, J.; Saul, N.; Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **2018**, *3*, 861.
(41) Foà, M.; Cassar, L. Oxidative Addition of Aryl Halides to Tris(triphenylphosphine)nickel(0). *J. Chem. Soc., Dalton Trans.* **1975**, 2572−2576.
(42) Izenman, A. J. Modern Multivariate Statistical Techniques, Regression, Classification, and Manifold Learning. *Springer Texts Statistics* **2008**. DOI: 10.1007/978-0-387-78189-1.
(43) (a) Hastie, T.; Tibshirani, R.; Friedman, J. Cluster Analysis. In *The Elements of Statistical Learning*, 2nd ed.; Springer: 2009; pp 501−527. DOI: 10.1007/b94608_2. (b) Kowalski, B. R.; Bender, C. F. Pattern Recognition. A Powerful Approach to Interpreting Chemical Data. *J. Am. Chem. Soc.* **1972**, *94*, 5632−5639.
(44) Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53−65.
(45) Kursa, M. B.; Jankowski, A.; Rudnicki, W. R. Boruta − A System for Feature Selection. *Fundam. Inform.* **2010**, *101*, 271−285.
(46) Peng, H.; Long, F.; Ding, C. Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226−1238.
(47) Hastie, T.; Tibshirani, R. Generalized Additive Models. *Stat Sci.* **1986**, *1*, 297−310.
(48) Zhu, C.; Yue, H.; Jia, J.; Rueping, M. Nickel-Catalyzed C-Heteroatom Cross-Coupling Reactions under Mild Conditions via Facilitated Reductive Elimination. *Angew. Chem., Int. Ed.* **2021**, *60*, 17810−17831.
(49) Using cross-validation we found that multivariate models do not offer increased predictive power, indicating that a higher fidelity regression model requires more data or dedicated feature engineering.
(50) It is interesting to note that the most positive outliers (4-tBu and 4-Ph) in the model are the substrates that we subjected to intensive reaction optimization.
(51) Poisoning by an aldehyde additive could arise from activation by Ni, HAT, or photosensitization by a photoredox catalyst. See: (a) Hoshimoto, Y.; Ohashi, M.; Ogoshi, S. Nickel-Catalyzed Selective Conversion of Two Different Aldehydes to Cross-Coupled Esters. *J. Am. Chem. Soc.* **2011**, *133*, 4668−4671. (b) Zhang, L.; Si, X.; Yang, Y.; Zimmer, M.; Witzel, S.; Sekine, K.; Rudolph, M.; Hashmi, A. S. K. The Combination of Benzaldehyde and Nickel-Catalyzed Photoredox C(sp^3)−H Alkylation/Arylation. *Angew. Chem., Int. Ed.* **2019**, *58*, 1823−1827.
(52) Lin, S.; Dikler, S.; Blincoe, W. D.; Ferguson, R. D.; Sheridan, R. P.; Peng, Z.; Conway, D. V.; Zawatzky, K.; Wang, H.; Cernak, T.; Davies, I. W.; DiRocco, D. A.; Sheng, H.; Welch, C. J.; Dreher, S. D. Mapping the dark space of chemical reactions with extended nanomole synthesis and MALDI-TOF MS. *Science* **2018**, *361*, 1−7.