

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^{*}

Department of Chemistry, Princeton University, Princeton, New Jersey 08544, United States

Supporting Information Placeholder

ABSTRACT: Ni/photoredox catalysis has emerged as a powerful platform for C(sp²)-C(sp³) bond formation. While many of these methods typically employ aryl bromides as the C(sp²) 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 non-standardized 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 with supervised machine learning.

INTRODUCTION

Although a relatively new field, Ni/photoredox catalysis has seen widespread development as a general approach to C(sp²)-C(sp³) 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(sp³), rather than C(sp²), fragments onto arenes is a well-established

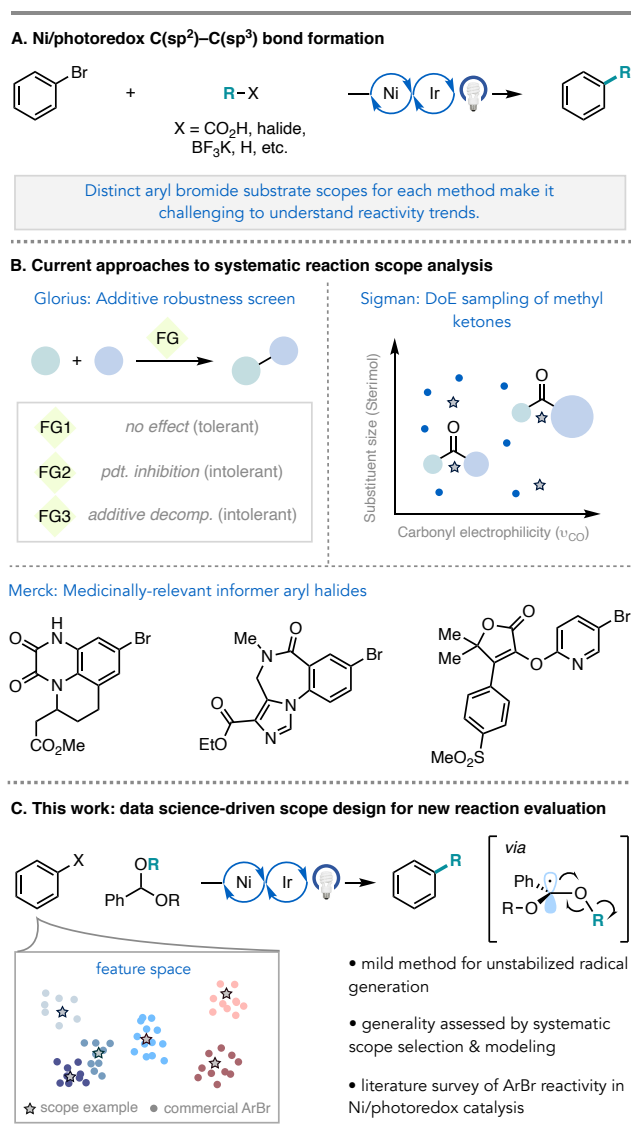


Figure 1. Challenges & new approaches to surveying reactivity from scope studies.

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(sp³)-H bonds^{4g} (Figure 1A). In principle, these reactions afford access to the same product structures, but it is difficult to compare amongst 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.

Thus, 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 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.

Alternatively, high-throughput experimentation (HTE) has been used in a retrospective sense to evaluate substrate generality across synthetic methods. For example, scientists at Abbvie recently studied seven different Pd-, Ni-, and Ni/photoredox-catalyzed methodologies with four common aryl bromides, providing insight into which reactions may provide the highest success for installing methyl, primary aliphatic, secondary aliphatic, benzylic, α -heteroatom, or tertiary aliphatic groups.¹⁰ Nevertheless, the ability to gain similar information directly from reaction development manuscripts without performing additional experimentation would be of immediate value to the synthetic community.

To address these goals, we pursued a quantitative method informed by related studies in chemoinformatics and molecular library design,¹¹ including for informing the selection of bioactive molecules from inactive molecules,¹² virtual screening of drug-like molecules based on structure or their interaction with known protein targets,¹³ and the

generation of additive¹⁴ and ligand¹⁵ libraries for transition metal-catalyzed processes. Specifically, we used our recently developed software package, auto-qchem, to run DFT calculations for a commercial set of >2600 aryl bromides and extract their shared molecular and atomic features. After dimensionality reduction with Uniform Manifold Approximation and Projection (UMAP), we employed hierarchical clustering to identify 15 maximally distinct groups of aryl bromides and selected one substrate from each group for a subsequent scope study.

We assess this data science-selected scope 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 with aryl bromides. Alcohols are attractive cross-coupling partners owing to their abundance, low cost and toxicity, and functional group compatibility.¹⁶ However, use of alcohols as aliphatic radical precursors in Ni/photoredox catalysis has seen limited development due to their high C(sp³)-O bond strengths. Zuo and co-workers have reported C(sp³)-CH₂OH activation of free alcohols in photocatalytic Ni/Ce cross-coupling.^{17a} Alternatively, the MacMillan and Martin groups have pursued masked alcohol derivatives for Ni/photoredox cross-coupling, including oxalates^{4d} and phthalimides.^{17b} 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, the Li group reported an electrochemical Ni-catalyzed coupling of free alcohols using a phosphine mediator for C(sp³)-O bond activation.¹⁸

Recently, our lab demonstrated that trimethyl orthoformate, another alcohol-based coupling partner, can serve as a source of high-energy methyl radicals upon β -scission of its tertiary carbon-centered radical.¹⁹ 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(sp²)-C(sp³) 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.²⁰

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 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.¹⁹ In developing a new method that could use acetals as stoichiometric coupling partners, we sought to evaluate alternative hydrogen atom transfer (HAT) catalysts and strategies.²¹ 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)²² 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),²² 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(CF₃)ppy)₂(dtbbpy)]PF₆ (1 mol%) as a photocatalyst, NiBr₂·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 K₃PO₄ (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% NiBr₂·glyme and 3 mol% dtbbpy. Under these conditions, methylation proceeded in 74% yield (Table 1, Entry 2). 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,23} 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).²² 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 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;²⁴ 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.

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	<i>t</i> -Bu	none	none	16%
8	<i>t</i> -Bu	25% light intensity	TBABr (25 mol%)	78%
9	<i>t</i> -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).

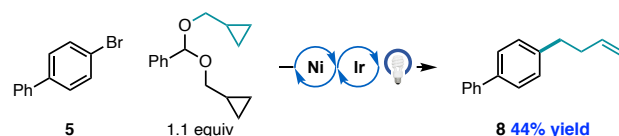
Mechanistic Investigations

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**).^{20,25} 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).

A. Observed product & byproduct from β -scission^a



B. Radical clock experiment with cyclopropylmethyl acetal^a



C. Methyl radical generation and trapping

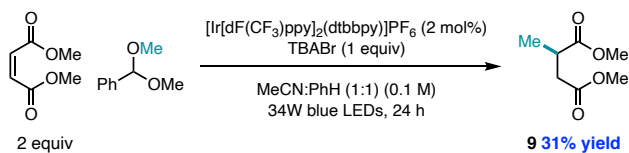


Figure 2. Selected mechanistic studies. ^a Conditions shown in Table 2 Entry 9. ^b Yield relative to acetal.

We next sought to understand the nature of the HAT step. According to the mechanistic proposal, bromide undergoes oxidation to bromine radical ($E_{1/2}[\text{Br}^-/\text{Br}^\bullet] = +0.80$ V vs SCE in DCE; $E_{1/2}[*\text{Ir}^{\text{III}}/\text{Ir}^{\text{II}}] = +1.21$ V vs SCE in MeCN)^{4b} to mediate HAT with the tertiary C–H bond of the acetal. To evaluate the feasibility of 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 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 bromine radical. These studies were initiated with aryl chloride **10** using NiBr₂·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. However, when NiCl₂·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 chlorine radical (Table 2, Entry 3).^{26,27} Reactivity was restored upon introduction of TBABr to reactions employing NiCl₂·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.

Table 2. Role of halide in the reaction

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%

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

An alternative pathway that is consistent with this 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.^{27,28}

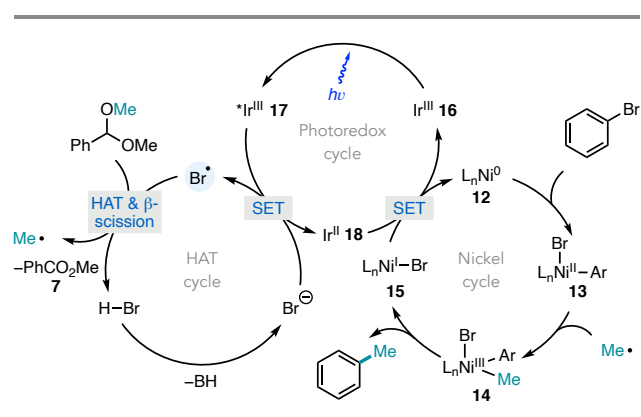


Figure 3. Catalytic cycle for aryl methylation from acetals.

On the basis of these studies, a catalytic cycle is shown in Figure 3. Oxidative addition of Ni(0) complex **12** into an aryl 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 bromine radical. Bromine then mediates selective HAT with the tertiary C–H bond of benzaldehyde dimethyl acetal, which upon β -scission affords 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(sp²)–C(sp³) 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 next sought to examine a synthetic application of acetals as low molecular weight radical sources. Given the feasibility of methylating an aryl chloride in the presence of a catalytic bromide additive (Table 2), this exploration was initiated with late-stage pharmaceutical product fenofibrate, which possesses an aryl chloride. 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.²⁹ 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 multi-gram 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 *iso*-amyl (**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 (CD₃I),^{30a} trideuteriomethyl *p*-toluenesulfonate (CD₃OTs),^{30b} and dimethyl-*d*₆ carbonate (CO(OCD₃)₂)^{30c} have previously found application as deuteromethyl

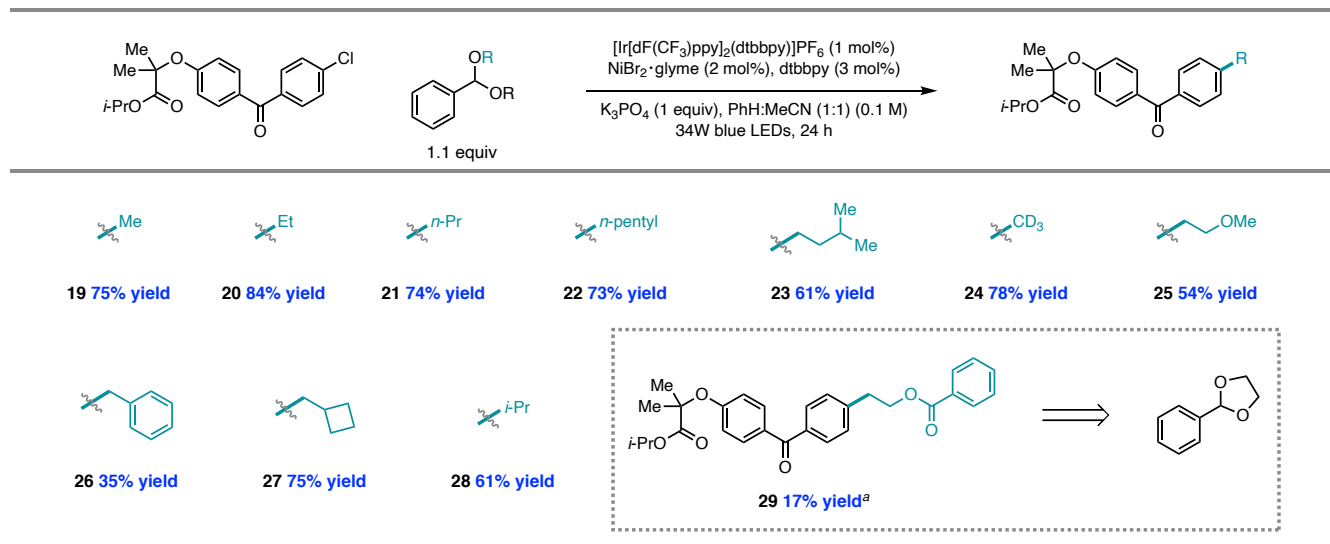


Figure 4. Late-stage functionalization of fenofibrate from benzaldehyde di(alkyl) acetals. Yields are isolated yields on 0.4 mmol scale. ^a Reaction performed with 25 mol% TBABr

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,³¹ 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, *iso*-propylation 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. We next initiated aryl bromide scope studies to evaluate the generality of this method. We aimed to study how well the aryl bromide chemical space has been covered in Ni/photoredox cross-coupling, as well as 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 a ~2600-member dataset to analyze aryl bromide chemical space. Because we filtered the aryl bromides on the basis of functionality known to be 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 dataset is not universally applicable, but instead designed to serve the current method. An analogous pre-selection approach, perhaps informed by a Glorius robustness screen⁷ or prior knowledge, can be applied to other methods.

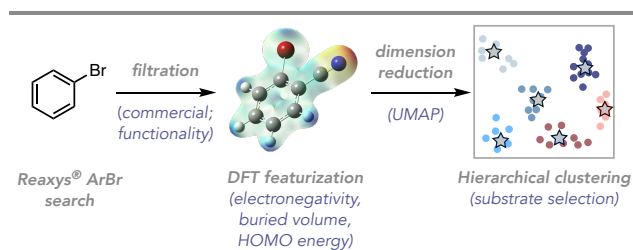
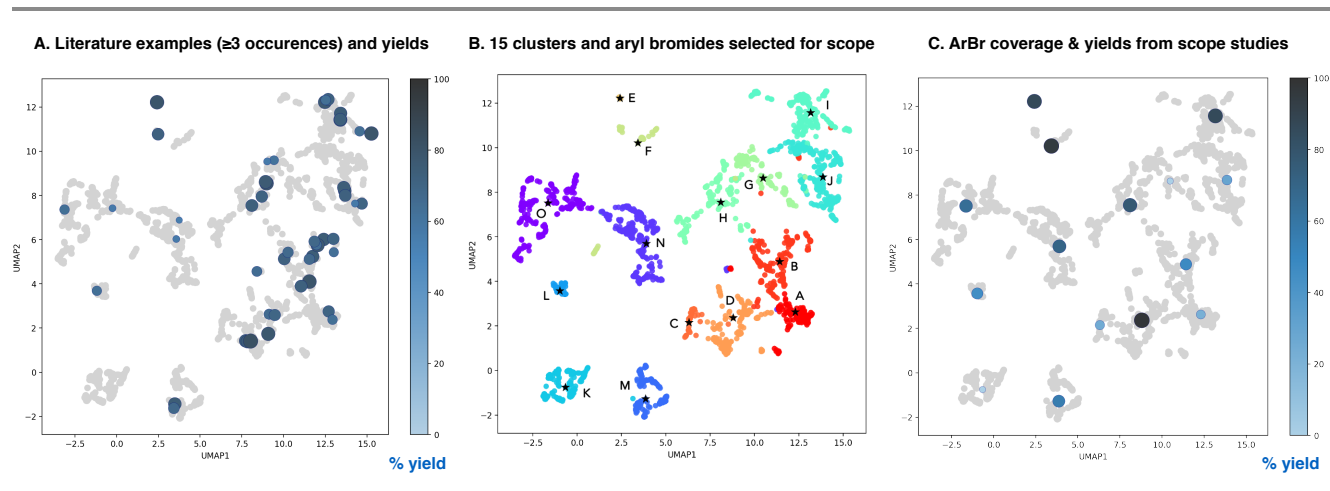


Figure 5. Workflow for construction of the chemical space and substrate selection via unsupervised learning.

To quantify similarity, we next pursued molecular featurization of the aryl bromides. Several approaches were considered, including molecular fingerprints,³² Mordred descriptors,^{14,33} 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 trends³⁴ (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 code³⁵ for the set 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



D. Data science-selected aryl bromide scope

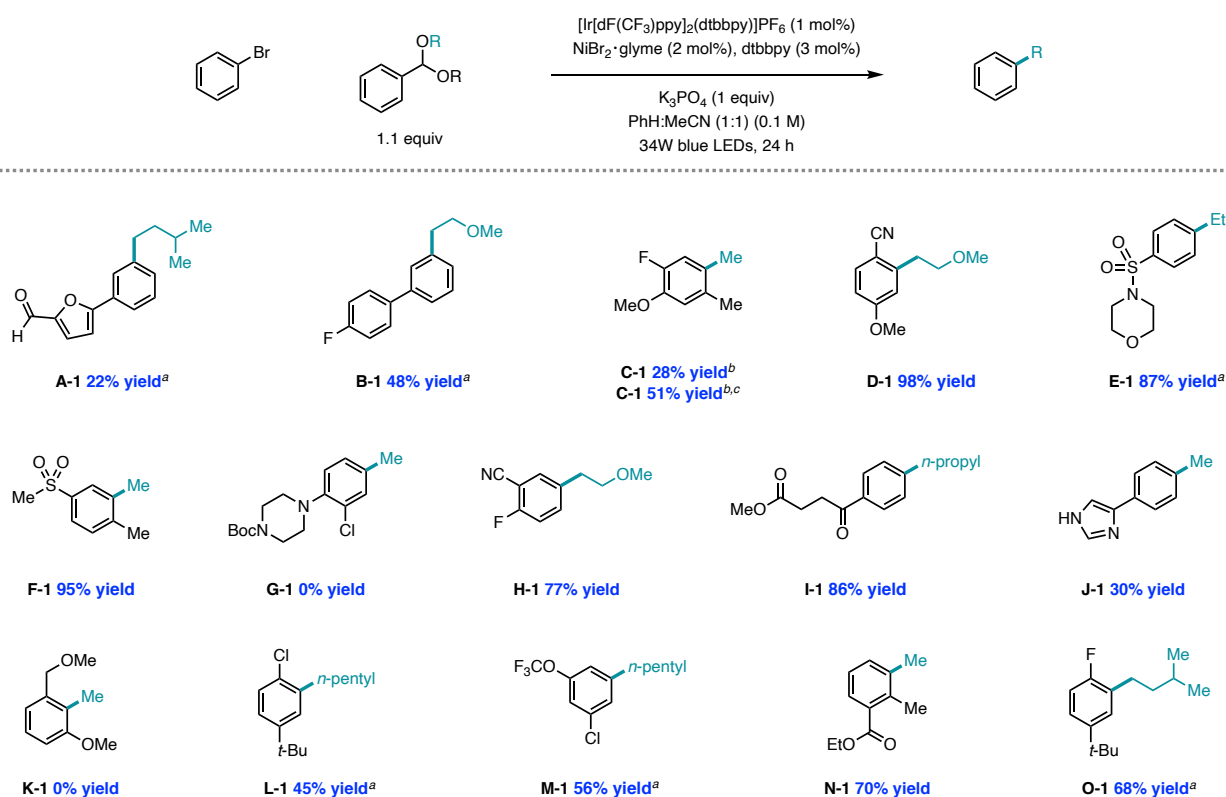


Figure 6. Evaluation of the literature coverage, clustering of the chemical space for scope selection, and scope studies (0.4 mmol scale). ^a 25 mol% TBABr added; ^b ¹H NMR yield; ^c 5 mol % Ni/7.5 mol % dtbbpy; ^d GC yield.

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 featured; only their impact on the common core atoms (six aromatic carbons and bromine) and the whole molecule is accounted for.³⁶

Next, we conducted a literature survey of aryl bromides employed in Ni/photoredox methodologies. Our literature search produced 116 papers (see SI for details),³⁷ 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)³⁸ 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 S10). 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 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 S16). 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.³⁹

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.⁴⁰ We chose agglomerative hierarchical clustering with the Ward connectivity criterion.⁴¹ 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,⁴² 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 in the range of -1 and 1: large 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 fifteen 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 S9). Ultimately, we selected 10-dimensional UMAP and 15 clusters as our optimal parameters. From the fifteen chemically diverse clusters, we selected the center-most molecule per cluster for experimental evaluation, as indicated in Figure 6B. Because methyl and primary aliphatic acetals perform similarly with aryl bromides in the cross-coupling method, we limited the scope evaluation only to these partners (Figure 6D). The products are labeled with a letter (A-O) that corresponds to their cluster 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 tri-substitution patterns. The frequency of multi-substitution in the data science-generated scope contrasts the literature scope, which is dominated by mono-substituted 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 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. With the exception of G-1 and *ortho,ortho*-disubstituted K-1, all aryl bromides underwent productive alkylation to generate at least 22% yield of the desired product. The yields of these products and their distribution over the chemical space are shown in Figure 6C. Figures 6A and 6C present overall similar reactivity profiles, suggesting that this method compares favorably to the collection of methods published in the literature.

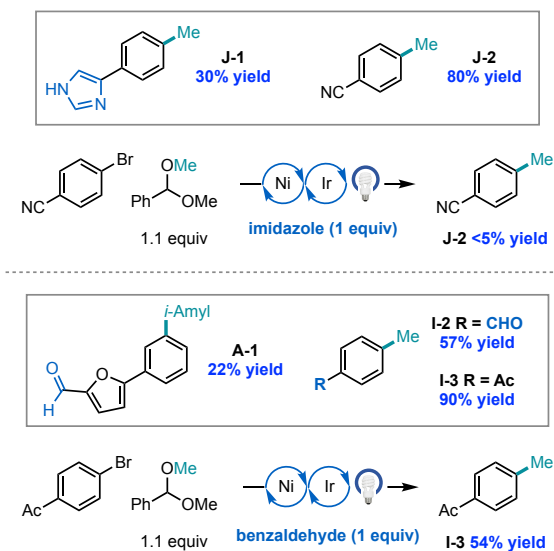
Nevertheless, we noticed that Cluster J generally comprises high-performing literature examples, yet in the methylation reaction, J-1 was obtained in only 30% yield. In optimization studies, we found that 4-bromo-*tert*-butylbenzene (3), another aryl bromide that belongs to Cluster J, underwent methylation to produce 4 in 89% yield. We therefore hypothesized that the lower yield of J-1 was a consequence of functional group tolerance of the imidazole rather than one of relative reactivity across the cluster. To test this hypothesis, methylation of another common literature substrate from Cluster J, 4-bromobenzonitrile, was examined. Under our standard conditions, this substrate underwent methylation to produce J-2 in 80% yield. However, when methylation of this substrate was re-evaluated with 1 equiv imidazole, J-2 was formed in <5% yield (Figure 7A, top). This example highlights that 15 substrates is simply not enough to sample the reactivity *and* functional group compatibility encompassed in the chemical space. Nevertheless, use of substrate clustering in this manner can serve as a complement to the Glorius fragment approach to reveal reactivity and functional group tolerance trends.⁷ In fact, robustness screening would have indicated that the preparation of J-1 would fail, whereas the data science clustering approach showed that preparatively useful yield is possible.

Not all cases of low reactivity are a consequence of functional group intolerance. For example, C-1 underwent methylation in 28% yield and K-1 was completely unreactive, likely as a result of steric hindrance. We hope that identification of these less reactive clusters can inspire new reaction development toward improving reactivity in a cluster. Preliminarily, we have found that the yield of tetrasubstituted C-1 could be boosted to 51% by increasing the cross-coupling catalyst loading to 5 mol% NiBr₂-glyme and 7.5 mol% dtbbpy.

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 some predictive generalizations of the performance of unseen aryl bromides. To do so, we evaluated univariate relationships with supervised machine learning. Using a combination of Boruta⁴³ and mRMR⁴⁴ feature selection methodologies on our dataset of 15 aryl bromides from the clustering (A-1 to O-1), we found that that *electronegativity* (χ) of the aryl bromide, defined as the average of the HOMO and LUMO energies, serves as a predictive feature.⁴⁵ This finding is in line with chemical intuition,

A. Effects of additive doping on reaction performance



B. Regression modeling using data science-selected aryl bromides as a training set

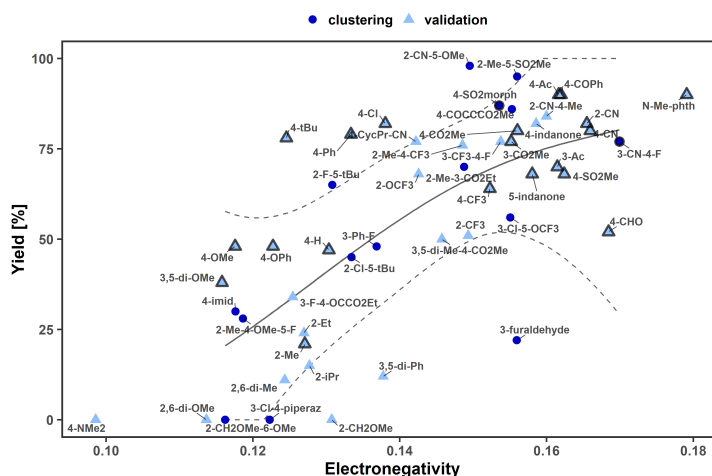


Figure 7. (A) Additive poisoning studies using the conditions shown in Figure 6 (^1H NMR yields with 1,3,5-trimethoxybenzene as an external standard). (B) Alkylation yield as a function of aryl bromide electronegativity. The model and a 2 standard error band are trained only with the 15 data science-selected aryl bromides. Aryl bromides that appear in our Ni/photoredox literature search at least 3 times are outlined.

where more electron-rich substrates undergo a more challenging oxidative addition to Ni.^{39,46} We then pursued training and validation of the electronegativity model using a generalized additive model (GAM).⁴⁷ 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 7B generalizes well to the validation data. The root-mean-squared-error (RMSE) is 21% and 19% in the clustering and validation data, respectively. Considering that the training examples are comprised of reactions with different alkyl coupling partners, the observation of a well-generalizable univariate model (R -squared = 56%) is significant. By contrast, the high yield of the common literature substrates found in the validation set (outlined in gray in Figure 7B) amounts to a selection bias and would likely make 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 S13). Thus, this analysis demonstrates that with a small and systematically diversified scope, one can provide predictive generalizations applicable to a larger population of substrates.

Examination of the regression model led us to make additional hypotheses about substrate reactivity.⁴⁸ For example, aldehyde-containing substrates that produced **A-1** and 4-methylbenzaldehyde (**I-2**) provided noticeably lower yields than predicted, seemingly model outliers. 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-3**) in 86 and 90% yields, respectively. Thus, we hypothesized that the aldehyde functionality was leading to reaction poisoning. To test this, we

evaluated the methylation of 4'-bromoacetophenone with 1 equiv benzaldehyde. While the reaction was not shut down by the additive, the yield of **I-3** was reduced to 54% from 90% yield (Figure 7A, bottom). This reaction was also marked by consumption of benzaldehyde (53% conversion), further suggesting that the reduced yields of **A-1** and **I-2** could be a result of the aldehyde functionality rather than influence of electronegativity on the reactivity of the substrate.^{23a-b,49} More broadly, an analysis like this suggests that the systematic substrate selection and modeling can be used to reveal substituent impacts that are both internal (affecting relative reactivity) and external (affecting off-cycle steps).

Limitations of the Data Science-Driven Scope

Clustering techniques, which are based on drawing a diverse subset of molecules from a large population, have several drawbacks. First, small changes to the starting population, e.g., new aryl bromides that become commercially available (or are no longer commercially available), may result in shifting the cluster boundaries. With a slightly modified starting population of aryl bromides, the central molecules for some clusters may change. Second, the molecular similarity may include features that are not relevant to the given reactivity, hence diversification along these features is not necessary. Accordingly, features that are relevant may not be emphasized enough. Finally, the number of clusters selected may be too small, resulting in diverse reactivity within the same cluster. As shown in our analysis above, this issue is likely compounded by the fact that functional groups can have both internal and external influences on the reactivity of a given substrate, but the DFT features likely only capture internal effects. Combining our approach with robustness screens, enhancing featurization and using an iterative approach to clustering (i.e., eliminating irrelevant features and re-clustering) could allow chemists to

identify a scope tailored to a given reaction and provide rich information about reactivity trends.

CONCLUSION

In conclusion, we have developed a C(sp²)-C(sp³) 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 (deuterio)methyl, primary aliphatic, and 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 fifteen substrates is not enough to sample the reactivity of all regions of the space, we show that substrate selection using this workflow and comparison to literature reactivity does allow us to formulate and test hypotheses about internal versus external impacts of functional groups on reactivity. Moreover, since a diverse set of substrates and outputs is essential for building accurate models that can be used for the exploration of the chemical space through interpolation, the data-science selected scope is better suited than traditional literature scopes for quantifying population-wide reactivity trends with supervised machine learning. As such, we anticipate this approach could serve as a model for standardizing scope analysis, enabling chemists to compare amongst methods, reduce the time and cost associated with scope evaluation, and afford literature data better suited to quantitative modeling of reactivity.

ASSOCIATED CONTENT

The Supporting Information is available free of charge on the ACS Publications website.

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

Smiles (Smiles_list_2683.txt).

Annotated Code (Bromide_clustering.ipynb, Bromide_regression.ipynb).

AUTHOR INFORMATION

Corresponding Author

Abigail G. Doyle – *Department of Chemistry, Princeton University, Princeton, New Jersey 08544, United States*; orcid.org/0000-0002-6641-0833; Email: agdoyle@princeton.edu

Authors

Stavros K. Kariofillis – *Department of Chemistry, Princeton University, Princeton, New Jersey 08544, 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*; orcid.org/0000-0002-8640-7551

Shivaani S. Gandhi – *Department of Chemistry, Princeton University, Princeton, New Jersey 08544, United States*; orcid.org/0000-0003-1825-5450

Jesus I. Martinez Alvarado – *Department of Chemistry, Princeton University, Princeton, New Jersey 08544, United States*; orcid.org/0000-0002-7061-8792

Author Contributions

*S.J. and A.M.Ż. contributed equally.

^S.S.G. and J.I.M.A. contributed equally

Funding Sources

The authors declare no competing financial interest.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

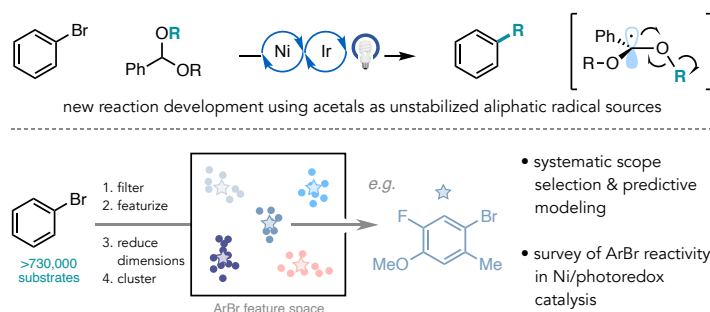
This material is based on work supported by the National Science Foundation Graduate Research Fellowship Program under Grant Number DGE-1656466 (to S.K.K.). A.M.Z. acknowledges financial support from the Schmidt DataX Fund at Princeton University made possible through a major gift from the Schmidt Futures 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 sp³-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: sp³-sp² 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 Informer 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(sp²)-C(sp³) Cross-Coupling Methods by Library Synthesis. *ACS Med. Chem. Lett.* **2020**, *11*, 597-604.
- [11] Xu, J.; Hagler, A. Chemoinformatics and Drug Discovery. *Molecules* **2002**, *7*, 566-600.
- [12] 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.
- [13] 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.
- [14] 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.; Minter, S. D.; Neurock, M.; Baran, P. S. *J. Am. Chem. Soc.* **2021**, *143*, 7859-7867.
- [15] (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. *ChemRxiv* **2021**, Preprint. <https://doi.org/10.26434/chemrxiv.12996665>.
- [16] Henkel, T.; Brunne, R. M.; Müller, H.; Reichel, F. *Angew. Chem., Int. Ed.* **1999**, *38*, 643-647.
- [17] (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 sp²-sp³ and sp³-sp³ Architectures via β -Scission of Aliphatic Alcohol Derivatives. *J. Am. Chem. Soc.* **2020**, *142*, 20594-20599.
- [18] 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.
- [19] (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(sp³)-H Cross-Coupling. *Acc. Chem. Res.* **2021**, *54*, 988-1000.
- [20] (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.
- [21] Capaldo, L.; Ravelli, D. Hydrogen Atom Transfer (HAT): A Versatile Strategy for Substrate Activation in Photocatalyzed Organic Synthesis. *Eur. J. Org. Chem.* **2017**, 2056-2071.
- [22] All bond strength calculations performed at the uM062x/def2tzvpp//uB3LYP/CBSB7 level of theory.
- [23] 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(sp³)-C(sp²) 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(sp³)-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(sp³)-H Acylation for the Synthesis of α -Amino Ketones. *J. Am. Chem. Soc.* **2020**, *142*, 19058-19064.
- [24] 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.
- [25] (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.
- [26] (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(sp³)-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.
- [27] 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.
- [28] 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**, 839-852.

- [29] (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.
- [30] (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.
- [31] Minor cyclobutylmethyl ring opening was observed in this reaction.
- [32] (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*, 1203-1207. (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.
- [33] Moriwaki, H.; Tian, Y.-S.; Kawashita, N.; Takagi, T. Mordred: a molecular descriptor calculator. *J. Cheminformatics* **2018**, *10*.
- [34] Żurański, A. M.; Martínez Alvarado, J. I.; Shields, B. J.; Doyle, A. G. Predicting Reaction Yields via Supervised Learning. *Acc. Chem. Res.* **2021**, *54*, 1856-1865.
- [35] <https://github.com/PrincetonUniversity/auto-qchem>
- [36] The process does not directly translate to heteroaryl bromides since the diversity of their structures would lead to different results for the (algorithmic) extraction of the common core atoms. Preliminary, we have found the reaction does extend to heteroaromatic compounds (see SI).
- [37] Literature search performed on December 10, 2020.
- [38] (a) McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv Preprint* Feb 9, **2018**, arXiv:1802.03426 [stat.ML]. <https://arxiv.org/abs/1802.03426> (accessed 2021-5-15). (b) McInnes, L.; Healy, J.; Saul, N.; Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **2018**, *3*, 861.
- [39] Foà, M.; Cassar, L. Oxidative Addition of Aryl Halides to Tris(triphenylphosphine)nickel(0). *J. Chem. Soc., Dalton Trans.* **1975**, 2572-2576.
- [40] Izenman, A. J. Modern Multivariate Statistical Techniques, Regression, Classification, and Manifold Learning. *Springer Texts Statistics* **2008**.
- [41] (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.
- [42] Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53-65.
- [43] Kurşa, M. B.; Jankowski, A.; Rudnicki, W. R. Boruta – A System for Feature Selection. *Fundam. Inform.* **2010**, *101*, 271-285.
- [44] 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.
- [45] Hastie, T.; Tibshirani, R. Generalized Additive Models. *Stat Sci* **1986**, *1*, 297-310.
- [46] 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.* **2020**. DOI: 10.1002/anie.202013852.
- [47] 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.
- [48] It is interesting to note that the most positive outliers (4-*t*Bu and 4-Ph) in the model are the substrates that we subjected to intensive reaction optimization.
- [49] 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(sp³)-H Alkylation/Arylation. *Angew. Chem., Int. Ed.* **2019**, *58*, 1823-1827.



Click or tap here to enter text.