JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Cite This: J. Am. Chem. Soc. 2018, 140, 5004−5008
Communication
pubs.acs.org/JACS

Deoxyfluorination with Sulfonyl Fluorides: Navigating Reaction Space with Machine Learning

Matthew K. Nielsen, Derek T. Ahneman, Orestes Riera, and Abigail G. Doyle*

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

Supporting Information

ABSTRACT: Through fine-tuning of reagent and base structure, sulfonyl fluorides can efficiently fluorinate diverse classes of alcohols. We show that machine learning can map the intricate reaction landscape and enable accurate prediction of high-yielding conditions for untested substrates.

Due to the multidimensionality of chemical reactivity, vast resources are currently expended on the translation of reported synthetic methods to new chemical structures.[1] An ongoing conundrum in methods development is the trade-off between reporting a single protocol for many substrates vs developing highly optimized conditions tailored to individual substrates. Chemists tend to seek a single set of "optimal" conditions with the objective of rendering the method easily adoptable. However, if these conditions are low-yielding or fail to deliver product for a new substrate, a user has little recourse for advancement. Alternatively, chemists can report distinct sets of conditions for different substrate classes. In this case, new users may be dissuaded from adopting the method due to the complexity of interpreting multidimensional chemical data. Several strategies have been advanced to address this multitude of options, including the commercialization of broad spectrum catalyst mixtures,[2] the development of high-throughput informer libraries,[3] and the publication of reviews featuring "how-to" manuals.[4] As an alternative, we questioned whether it would be possible to employ machine learning (ML) to analyze existing reaction data in its full complexity and then predict optimal conditions on a per substrate basis.[5] This approach would enable researchers to evaluate the feasibility of a specific transformation and identify high-yielding conditions for a new substrate *a priori*.

Recently, our laboratory, in collaboration with scientists at Merck Research Laboratories, demonstrated that ML algorithms can successfully model and predict the outcomes of Buchwald−Hartwig aminations based on a training set comprising thousands of reactions performed via ultra-high-throughput experimentation (ultra-HTE).[6] One challenge was the necessity of isolating standards to calibrate ultra-HTE yields, a time-consuming process when hundreds of products are in play. To streamline our workflow, we employed a Glorius additive screen,[7] testing numerous isoxazole additives against a small subset of amination reactions.

Following this initial study, we hoped to expand our evaluation of ML utility to data sets containing more structural diversity and mechanistic non-uniformity, representative of the challenges faced during adoption of most synthetic methods. In this publication, we apply ML to the analysis of deoxyfluorination with sulfonyl fluorides. Since fluorination yields may be accurately determined from reaction mixtures by ^{19}F NMR, we were able to rapidly screen structurally and stereochemically diverse alcohols. This study identifies sulfonyl fluorides as a reagent class with tunable and broad-spectrum reactivity. However, distinct combinations of sulfonyl fluorides and bases are necessary to obtain high yields across different alcohol classes. Despite the complexity of the reaction space, we show that a random forest ML algorithm trained on HTE deoxyfluorination data can be used to predict yield and high-yielding conditions for entirely new substrates.

Deoxyfluorination transforms alcohols to aliphatic fluorides, which are high-value targets in pharmaceutical development.[8] Traditionally, most deoxyfluorinations have been carried out with DAST or other sulfur(IV) reagents that suffer from thermal instability.[9] Recently, the Ritter group has developed Phenofluor and other fluoroimidazolium compounds capable of fluorinating complex substrates with exquisite selectivity.[10] Our group and others have identified sulfonyl fluorides such as PyFluor and perfluorobutanesulfonyl fluoride (PBSF) as noteworthy alternatives.[11] In addition to being inexpensive and thermally stable, the sulfonyl fluoride motif is highly modular. Diverse sulfonyl fluorides may be synthesized by both classical and modern methods,[12] enabling one to fine-tune the reagent to specific alcohols. Notwithstanding, there has been no systematic investigation into the substrate or reagent scope of deoxyfluorination with sulfonyl fluorides.

In our previous report, we found that the combination of PyFluor and DBU was optimal for minimizing elimination in the preparation of acyclic secondary fluoride **2a** (see Figure 1).[11c] Although this protocol translates well to most unactivated alcohols, PyFluor delivers poor yields for activated substrates that are susceptible to nucleophilic attack by DBU, such as benzylic alcohol **1b**. Conversely, strained cyclic alcohols (**1c** and **1d**) are largely unreactive.

To better understand sulfonyl fluoride reactivity, we screened alcohols **1a**−**1d** against a selection of sulfonyl fluorides and bases (Figure 1A). We evaluated the commercial reagents PyFluor and PBSF as well as the arylsulfonyl fluorides **3-Cl**, **3-CF$_3$**, and **3-NO$_2$**, each of which is readily obtained from the commercial sulfonyl chloride. Based on experimentally determined relative rate constants, these sulfonyl fluorides span a broad range of reactivity that correlates with the leaving

Received: February 7, 2018
Published: March 27, 2018

DOI: 10.1021/jacs.8b01523
J. Am. Chem. Soc. 2018, 140, 5004−5008

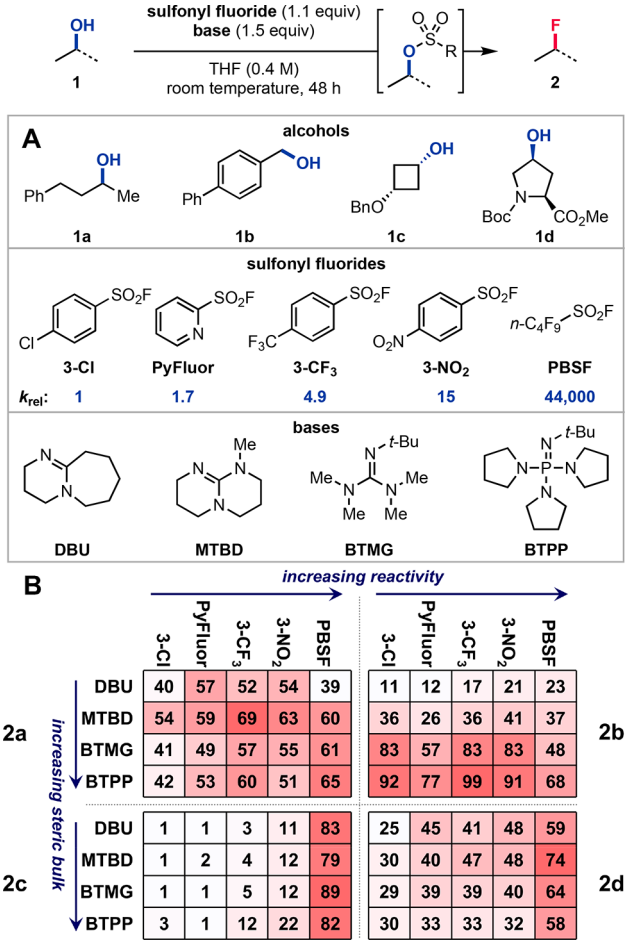

Figure 1. Multi-dimensional screening of alcohols **1a**–**1d**. The ^{19}F NMR yield for each combination of sulfonyl fluoride and base is displayed in the accompanying heat map. See SI for k_{rel} data.

group ability of the corresponding sulfonate ester (Supporting Information (SI) Figures S1–S3). Reactions with PBSF are typically complete within seconds, whereas arylsulfonyl fluorides may require hours or even days to reach full conversion. Likewise, the bases feature steric diversity, ranging from the compact amidine DBU to the bulky phosphazene BTPP.

As shown in Figure 1, alcohols **1a**–**1d** displayed very distinct reactivity profiles. Secondary fluoride **2a** and benzylic fluoride **2b** were obtained in highest yield with **3-CF$_3$**, whereas cyclic fluorides **2c** and **2d** were favored by the more reactive PBSF. The reaction surfaces of **2a** and **2d** show a complex interplay between base and sulfonyl fluoride, with relatively unencumbered MTBD performing best. In contrast, bulky BTPP proved optimal for benzylic alcohol **1b**. Finally, cyclobutanol **1c** displayed almost no reactivity except with PBSF, for which yields were largely independent of base identity.

These results hint at a complex reaction space that defies simple axiomatic interpretation. As such, we expected that this method would be an ideal candidate for reaction modeling via ML. To develop a suitable training set, we subjected a diverse range of alcohols **1a**–**1af** to the same 20-reaction screen as shown in Figure 1, resulting in a total of 640 distinct reactions. The highest yielding conditions for each substrate are summarized in Figure 2.[13]

Unactivated primary alcohols (Figure 2A) were fluorinated in as high as 96% yield with arylsulfonyl fluorides and non-nucleophilic bases such as BTMG and BTPP (**2e**, **2g**, **2i**, **2j**). For acyclic secondary alcohols, we were unable to identify superior conditions to those previously disclosed with PyFluor (**2a**, **2h**).[11c] Time point studies indicated that second-order kinetics are operative in the fluorination of both primary and secondary alcohols, consistent with a rate-limiting S_N2 substitution between fluoride and the intermediate sulfonate ester. In contrast, tertiary alcohols required the exceptionally reactive perfluorobutanesulfonate leaving group and displayed distinct first-order kinetics indicative of a unimolecular S_N1 pathway (**2f**). In the case of α-amino alcohol **1j**, we suspected that the aniline might engage in neighboring group participation. In support of our hypothesis, piperidine **1k** reacted to form both **2k** and the rearranged azepane **2k′**, presumably via the intermediacy of an aziridinium ion.[14]

Cyclic substrates (Figure 2B) typically have higher energy transition states arising from deformation of the ideal S_N2 trigonal bipyramidal geometry. Indeed, we found that cyclic alcohols **1c**, **1d**, and **1l**–**1q** were fluorinated in up to 94% yield with highly activated PBSF. Full stereochemical inversion was observed including for hydroxyproline diastereomers **2c** and **2l** as well as cyclohexanol diastereomers **2o** and **2p**. In contrast to 4- and 5-membered cyclic alcohols, cyclohexanols performed poorly due to the absence of ring strain in the ground state, which raises the $\Delta G^{\ddagger}$ for substitution.[15] Cyclohexanol **1o**, for which the alcohol is locked in an equatorial conformation, delivered **2o** in only modest yield due to the highly congested backside trajectory. However, axial diastereomer **1p** afforded almost exclusively E2 elimination owing to the constrained *anti*-periplanar geometry.

Benzylic fluorides (Figure 2C) were obtained in as high as 90% yield by employing bulky non-nucleophilic bases (**2b**, **2r**–**2v**). Fluoride **2v** exhibited retention of configuration, consistent with selective ring-opening of an aziridinium intermediate.[16] Primary allylic fluoride **2w** was formed with high selectivity (12:1) against the branched S_N2' product; however, the secondary fluoride **2x** exhibited only 4:1 rr, and the tertiary fluoride **2y** was dominated in 1.5:1 rr by the linear fluoride. Homobenzylic alcohols **1z** and **1aa** (Figure 2D) afforded high yields, but may form via an arenium ion. The analogous competing homoallylic rearrangement resulted in low yield of **2ab**.[17] The α-hydroxycarbonyl compounds **1ac** and **1ad** performed reasonably well, but β-hydroxycarbonyl substrates afforded exclusively α,β-unsaturated products. A rare exception, **2ae** was formed in 24% yield with BTPP due to shielding of the α carbon by the bulky trityl group. Fluorination of hemiacetal **1af** likely involves an oxocarbenium intermediate as evidenced by the dependency of diastereoselectivity on the sterics of the base, which serves as the fluoride counterion.[13]

In general, activated substrates perform optimally with comparatively electron-rich sulfonyl fluorides that form stable sulfonate esters, thus minimizing elimination and undesired side reactions, whereas deactivated cyclic and tertiary alcohols react best with the highly reactive PBSF. Unencumbered primary alcohols require the bulky bases BTMG or BTPP, while congested secondary alcohols are frequently higher-yielding with the more compact bases. Notwithstanding, these observations fail to encompass the full complexity of the enclosed reaction space. In this regard, ML is the ideal tool for capturing such intricacies.

DOI: 10.1021/jacs.8b01523
J. Am. Chem. Soc. 2018, 140, 5004–5008

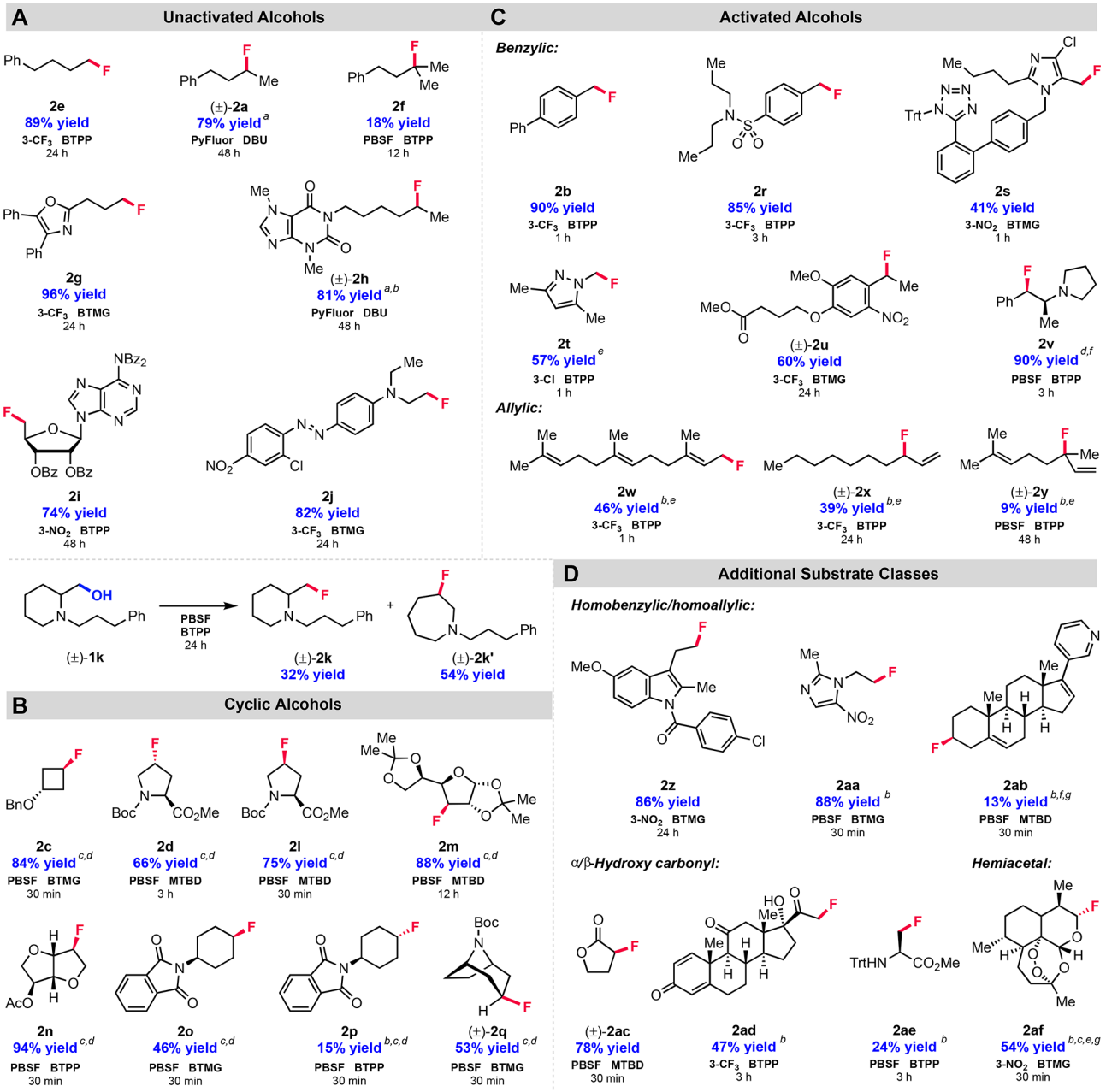

Figure 2. Scope of deoxyfluorination. The reagents used for each alcohol are listed below the yield. Reactions were performed on 0.5−1.0 mmol scale with 1.1 equiv sulfonyl fluoride and 1.1−2 equiv base in THF (0.1−0.5 M) at room temperature. [a]Solvent: toluene (1 M). [b]Co-isolated with side product (see SI). [c]Inversion observed. [d]>20:1 diastereoselectivity. [e]^{19}F NMR yield. [f]Retention observed. [g]4:1 dr.

Previously, we demonstrated that the random forest ML algorithm, which is specifically designed to avoid overfitting by averaging the output of many randomly generated decision trees,[18] affords superior predictive ability over a range of ML algorithms. In order to train a random forest model for sulfonyl fluoride deoxyfluorination, we assembled a table of descriptors for the alcohols, sulfonyl fluorides, and bases employed in each of the 640 screening reactions. These descriptors included computed atomic and molecular properties (i.e., electrostatic charge, NMR shift) as well as binary categorical identifiers (i.e., primary, secondary, tertiary, cyclic). Using R statistical software,[13] we applied a random forest algorithm to a training set containing 70% of the screening entries, and evaluated the model using a test set comprising the remaining 192 reactions. Remarkably, the trained model predicted the yields of the test set reactions with a root-mean-square error (RMSE) of only 7.4% yield (Figure 3A). In comparison to our previous work, this training set was smaller (448 vs 2898 training reactions), encompassed much broader substrate diversity, and incorporated multiple mechanisms (S_N1, S_N2, anchimeric assistance), highlighting the robust predictive ability of the random forest algorithm.

As further validation, we evaluated substrates **1ag**−**1ak** which do not appear in the training set (Figure 3B). Again, our model successfully predicted the yields of **2ag**−**2ai** with reasonable accuracy (8−13% RMSE). Substantially higher variation (~20% RMSE) was observed in the case of **2aj** and **2ak** which contain distinct functionality not found in the training set; however, the error was largely systematic, indicating that the algorithm successfully identified underlying trends. Notably, for each

DOI: 10.1021/jacs.8b01523
J. Am. Chem. Soc. 2018, 140, 5004−5008

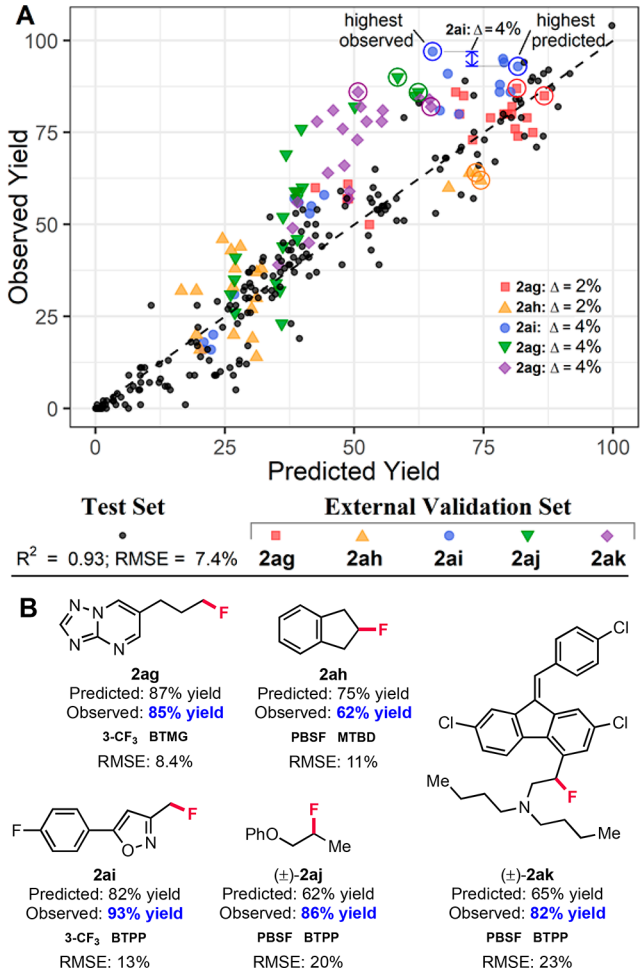

Figure 3. Performance of random forest model. (A) Plot of observed vs predicted yields for the test set and validation substrates **1ag**–**1ak** (see legend). The highest predicted and highest observed yields are circled. (B) Comparison of predicted and observed yields for the predicted optimal conditions of **1ag**–**1ak**.

substrate, the experimental yield corresponding to the predicted optimal conditions lay within 4% of the highest-yielding conditions. This level of precision is more than sufficient for enabling new users to evaluate reaction feasibility and select initial reaction conditions. For example, although experimental yields for **2ai** ranged from 16% to 97%, the predicted optimal conditions afforded 93% yield, only 4% below the best experimental conditions. On the other hand, the best conditions for **2ai**, which delivered 97% yield, were predicted to provide only 65% yield, suggesting that expanded training sets along with advances in the description and modeling of the chemical space are necessary to use ML for predicting the best conditions.

In conclusion, we have demonstrated that the flexibility of the sulfonyl fluoride scaffold allows deoxyfluorination of a broad range of alcohols, significantly expanding upon the scope initially reported with PyFluor. Moreover, we show that a random forest model can accurately predict reaction outcomes and aid in identifying optimal conditions for new alcohols. We envision that this tool will enable users to navigate the multitude of distinct deoxyfluorination conditions, and that similar models could prove beneficial in the development of other synthetic methods. Continual expansion of our training set with new substrates and additional variables (i.e., stoichiometry, concentration, solvent, and temperature) will lead to more accurate and comprehensive coverage of the sulfonyl fluoride deoxyfluorination reaction space.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.8b01523.

Experimental procedures, additional reaction optimization, and spectroscopic data for all new compounds (PDF)

"rxnpredict" files (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

*agdoyle@princeton.edu

ORCID

Abigail G. Doyle: 0000-0002-6641-0833

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Financial support was generously provided by NSF (CHE-1565983), Janssen, and Celgene Corporation.

■ REFERENCES

(1) Santanilla, A. B.; Regalado, E. L.; Pereira, T.; Shevlin, M.; Bateman, K.; Campeau, L.-C.; Schneeweis, J.; Berritt, S.; Shi, Z.-C.; Nantermet, P.; Liu, Y.; Helmy, R.; Welch, C. J.; Vachal, P.; Davies, I. W.; Cernak, T.; Dreher, S. D. *Science* **2015**, *347*, 49.

(2) Fors, B. P.; Buchwald, S. L. *J. Am. Chem. Soc.* **2010**, *132*, 15914.

(3) 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. *Chem. Sci.* **2016**, *7*, 2604.

(4) Chatterjee, A. K.; Choi, T.-L.; Sanders, D. P.; Grubbs, R. H. *J. Am. Chem. Soc.* **2003**, *125*, 11360.

(5) The Sigman lab has introduced a data-intensive strategy to predict selectivity and gain mechanistic insight in catalytic reactions using linear regression analysis. For a leading reference, see: Milo, A.; Neel, A. J.; Toste, F. D.; Sigman, M. S. *Science* **2015**, *347*, 737.

(6) Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. *Science* **2018**, DOI: 10.1126/science.aar5169.

(7) Collins, K. D.; Glorius, F. *Acc. Chem. Res.* **2015**, *48*, 619.

(8) Yerien, D. E.; Bonesi, S.; Postigo, A. *Org. Biomol. Chem.* **2016**, *14*, 8398.

(9) (a) Middleton, W. J. *J. Org. Chem.* **1975**, *40*, 574. (b) Lal, G. S.; Pez, G. P.; Pesaresi, R. J.; Prozonic, F. M.; Cheng, H. *J. Org. Chem.* **1999**, *64*, 7048. (c) Beaulieu, F.; Beauregard, L.-P.; Courchesne, G.; Couturier, M.; LaFlamme, F.; L'Heureux, A. *Org. Lett.* **2009**, *11*, 5050. (d) Umemoto, T.; Singh, R. P.; Xu, Y.; Saito, N. *J. Am. Chem. Soc.* **2010**, *132*, 18199. (e) Messina, P. A.; Mange, K. C.; Middleton, W. J. *J. Fluorine Chem.* **1989**, *42*, 137.

(10) (a) Sladojevich, F.; Arlow, S. I.; Tang, P.; Ritter, T. *J. Am. Chem. Soc.* **2013**, *135*, 2470. (b) Goldberg, N. W.; Shen, X.; Li, J.; Ritter, T. *Org. Lett.* **2016**, *18*, 6102.

(11) (a) Shimizu, M.; Nakahara, Y.; Yoshioka, H. *Tetrahedron Lett.* **1985**, *26*, 4207. (b) Bennua-Skalmowski, B.; Vorbrüggen, H. *Tetrahedron Lett.* **1995**, *36*, 2611. (c) Nielsen, M. K.; Ugaz, C. R.; Li, W.; Doyle, A. G. *J. Am. Chem. Soc.* **2015**, *137*, 9571.

(12) (a) Dong, J.; Krasnova, L.; Finn, M. G.; Sharpless, B. K. *Angew. Chem., Int. Ed.* **2014**, *53*, 9430. (b) Davies, A. T.; Curto, J. M.; Bagley, S. W.; Willis, M. C. *Chem. Sci.* **2017**, *8*, 1233. (c) Tribby, A. L.; Rodríguez, I.; Shariffudin, S.; Ball, N. D. *J. Org. Chem.* **2017**, *82*, 2294.

(13) See Supporting Information for full data set and documentation.

DOI: 10.1021/jacs.8b01523
J. Am. Chem. Soc. 2018, 140, 5004–5008

(14) Déchamps, I.; Pardo, D. G.; Cossy, J. *Eur. J. Org. Chem.* **2007**, *2007*, 4224.
(15) Mange, K. C.; Middleton, W. J. *J. Fluorine Chem.* **1989**, *43*, 405.
(16) Hamman, S.; Beguin, C. G.; Charlon, C.; Luu-Duc, C. *J. Fluorine Chem.* **1987**, *37*, 343.
(17) Néder, Á.; Uskert, A.; Nagy, É.; Méhesfalvi, Z.; Kuszmann, J. *Acta Chim. Acad. Sci. Hung.* **1980**, *103*, 231.
(18) (a) Mitchell, J. B. *Comput. Mol. Sci.* **2014**, *4*, 468. (b) Kleinberg, E. M. *Ann. Statist.* **1996**, *24*, 2319. (c) Breiman, L. *Mach. Learn.* **2001**, *45*, 5.

DOI: 10.1021/jacs.8b01523
J. Am. Chem. Soc. 2018, 140, 5004−5008