Weighted Reaction Fingerprints for Visualizing Reactivity Cliffs and Generality

Babak Mahjour[1], Jillian Hoffstadt[1], Daniel Schorin[2], Tim Cernak[*1,3]

[1]Department of Medicinal Chemistry, University of Michigan
[2]School of Information, University of Michigan
[3]Department of Chemistry, University of Michigan

ABSTRACT: Visualization of reaction space is a critical step in improving human understanding of bulk chemical reaction data. We present weighted reaction fingerprinting, a simple method to rapidly analyze and evaluate the results of large reaction corpuses which we have found to be helpful in the analysis of high-throughput experimentation (HTE) campaigns. Weighted reaction fingerprints can be utilized to rapidly identify successful and failing conditions and systems for chemical transformations. Reactions are encoded into a standard "reaction anatomy" comprising popular reagent components such as nucleophile, electrophile, catalyst, ligand, and solvent. Each reaction is converted into a typical fingerprint matrix and multiplied by a weight vector to generate the weighted reaction fingerprint. These fingerprints are fed into dimensionality reduction algorithms such as principal component analysis (PCA) or t-stochastic neighbor embedding (t-SNE) to create visualizable 2-D manifolds that visually reveal reactivity trends such as catalysts with high substrate generality or reactivity cliffs.

Introduction

Statistical modeling of chemical reactions requires large datasets of chemical reactions from high-throughput experimentation (HTE) campaigns or curations of published data. Our group has been using HTE[1-14] to explore chemical reactivity. The primary goals of HTE analysis include statistical data profiling, allowing chemists to rapidly identify best, worst, and average performing reaction conditions, to understand which reaction conditions work best for certain substrate pairs, and to inspire the generation of ideas for new experimental space to explore. HTE reaction array visualizations as simple rectangular heatmaps are common ways to show the physical location on a wellplate where each reaction took place, but these rectangular heatmaps cannot easily communicate trends such as catalyst generality against substrate pairs or reactivity cliffs. Similarly, common machine learning or artificial intelligence algorithms for reaction informatics are opaque and challenge human-interpretable analyses. As reaction data[8] is deposited into centralized databases,[15] visual tools that communicate reaction performance in HTE campaigns, or related reaction datasets, are needed to identify patterns and trends.

Machine-readable molecular representations[16] are critical in understanding trends of chemical reactivity from large datasets. Graph representations[17] and molecular fingerprints have been used for the prediction of chemical properties, similarity searching, and structure optimization. In a reaction context, embeddings of the molecules from chemoinformatic, chemometric and quantum descriptors in conjunction with reaction outcomes are used to train models to predict reactivity and elucidate mechanisms.[18, 19] In particular, fingerprinting methods, such as the Morgan instantiation[20] of extended connectivity fingerprints (ECFP),[21] provide a fast and computationally non-intensive method to analyzing chemical data and the influx of reported reaction data in a standardized format. In recent literature, fingerprinting methods for reactions have included the concatenation of reagent fingerprints,[22, 23] feature binning fingerprints,[24] and reaction difference fingerprints.[25]

Herein, we demonstrate the utility of weighted reaction fingerprinting (Figure 1) – a simple abstractable method applicable to any large reaction corpus stored in a standardized format. Reaction discovery campaigns are analyzed to evaluate the results of multiplexed reaction arrays and large reaction datasets. We show that this visual tool can be used to analyze and draw conclusions from large corpuses of reaction data.

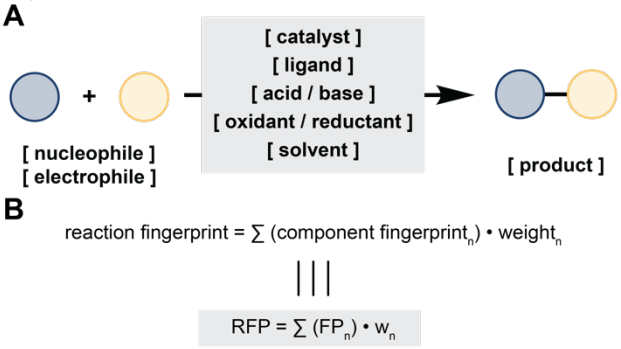

Figure 1. A) Anatomy of a reaction. Electrophiles and nucleophiles are arbitrarily defined as the reacting partners that form the product. B) The reaction fingerprint can be calculated by taking the product between a matrix of component fingerprints and a vector of weight arrays. Any fingerprint or feature vector can be utilized as long as the vector lengths for each reaction components are equal. Reaction component weights (w_n) are initialized at 1 and can be set to any value by the user. A component * weight product is added to the sum for each reaction component *n*.

Methods

We introduce reaction array fingerprinting with a dataset consisting of 2,786 Suzuki reactions.[3] Each reaction is composed of an electrophile, nucleophile, catalyst, base, and solvent. In Figure 2, the Morgan Fingerprint with radius 4 and 2,048 bits was calculated for each substance, then summed such that the five 2,048-bit binary fingerprints for each reaction become a single 2,048-bit count reaction array fingerprint – a one-dimensional array representation

of the coupling. This array of 2,786 reaction fingerprints was fed into a t-SNE algorithm with no hyperparameter tuning to generate Figure 2A, where reaction points are colored by the product yield for the reaction – high, medium, low, and 0% – based on product/internal standard (PROD/IS) integrations. The observed clusters are chemically distinguishable, with each cluster composed of similar or identical topological fragments in the reaction mixture. This reaction fingerprinting method (Figure S1) contrasts with currently published methods that include reagent concatenation,[22, 23] reaction feature binning,[24] and reaction difference fingerprints,[25] although weighing can at times be used in other reaction fingerprints to a similar effect. Results of different dimensionality reduction algorithms such as PCA, UMAP, and SOMs as well as different fingerprint representations, including one-hot encoding, are shown in the Supporting Information (Figure S2). Effects of modifying the perplexity of the t-SNE for this analysis is shown in Figure S3 and the same analysis is shown with molecular access system (MACCS) fingerprints[26] in Figure S4. Results of unsupervised versus supervised UMAP for the dataset is shown in Figure S5, and comparisons of weighted fingerprints against concatenated and difference fingerprints are shown in Figure S6. The six plots of Figure 2C display the same embedding shown in Figure 2B with six alternative color scales representing different features. The first plot's points are colored by the exact product/internal standard value for each reaction as calculated in the dataset. The remaining five plots are each colored by reagent per specific reagent class as defined in the template (this reaction dataset consists entirely of electrophile, nucleophile, catalyst, base, and solvent components.)

An example of a reaction cluster identified from the manifold is shown in Figure 2D. All reactions in the dataset using bromide electrophile **1** and boronate nucleophile **2** exist within this cluster. It is rapidly identified that this substrate pair was tested with 16 different catalysts and six different bases. Figure 2E showcases a simple extension of the fingerprinting algorithm using the data from this cluster. When summing the fingerprints of the individual components, a weight factor can be multiplied into a reagent's fingerprint to influence the clustering within the manifold. The 96 reactions between **1** and **2** were encoded as reaction fingerprints in two different formats. Once where the catalyst fingerprint was weighed by a factor of three, and again where the base fingerprint was weighed by a factor of three. These two datasets were then fed into the t-SNE reduction algorithm, and three plots colored by product/internal standard integrations, catalyst, and base for each of the two datasets are displayed in Figure 2E. As revealed by the color encoded reagents, elevating the catalyst weight produces manifolds with catalyst clusters, and similar behavior is seen with base clusters when elevating the base weight. This fingerprinting method can be used to identify reagents with high generality. For the same dataset, weighing is used to cluster electrophile and nucleophile pairs (Figure S7). When weighing the catalyst component, RuPhos Pd G3 was found as a high generality palladium catalyst and resulted in a 0.49 PROD/IS integration across 20 electrophile/nucleophile pairs. Similarly, 2-tert-butyl-1,1,3,3-tetramethylguanidine was found as a high generality base and a mixture of water and DMSO was found as a high generality solvent that worked consistently across the substrate scope used in the campaign.

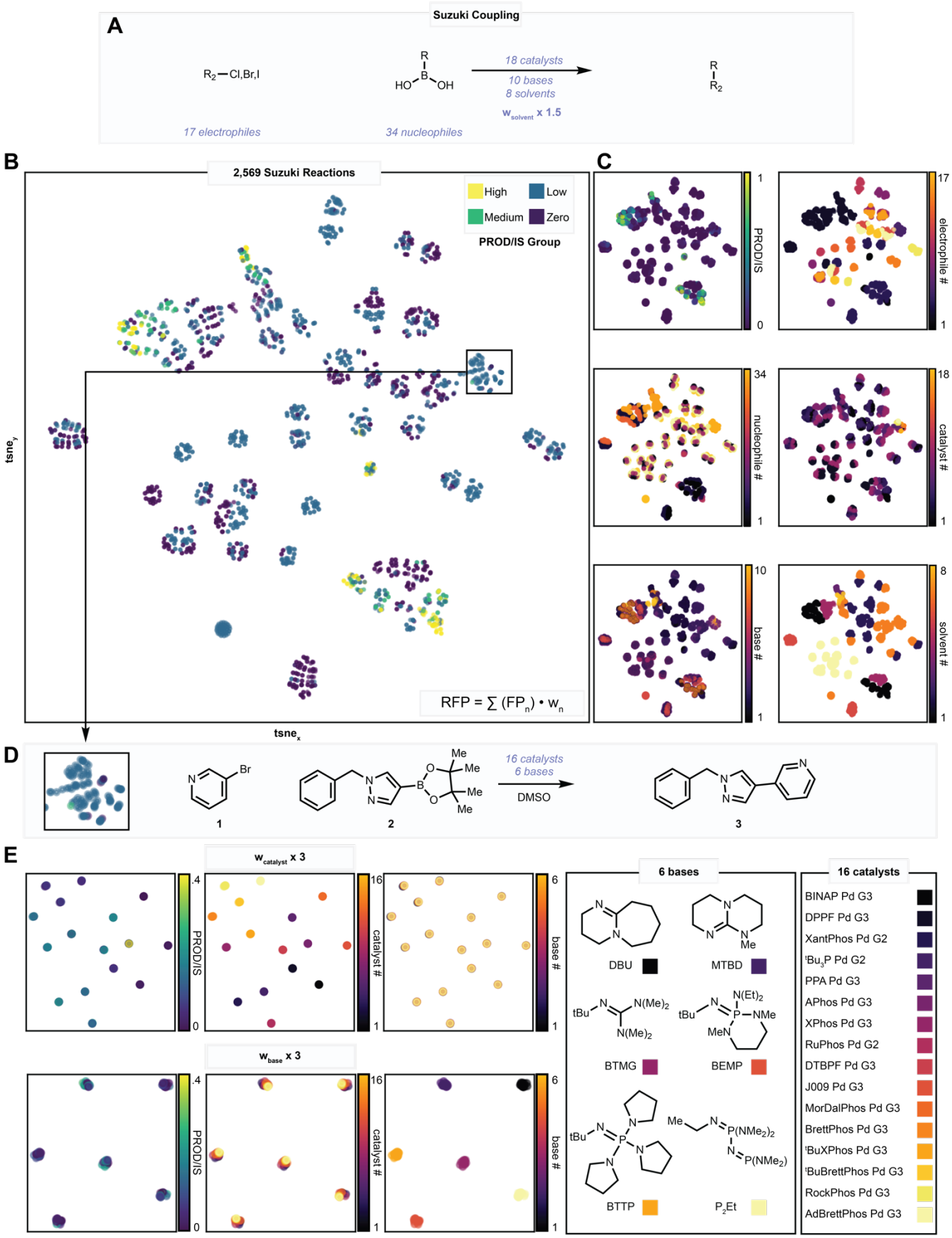

Figure 2. 2,786 Suzuki reactions performed in a high-throughput format plotted using a t-SNE trained on *reaction fingerprints*, each of which formed through the sum or concatenation of a reactant fingerprint for all components involved a particular reaction. (A) The template of the reaction performed across all experiments. (B) The 2,048-bit radius four Morgan Fingerprint of each reaction's electrophile, nucleophile, base, catalyst, and solvent were summed, plotted, and colored by output group. (C) The same manifold colored by exact product/internal standard values and reagent components (electrophile, nucleophile, catalyst,

base, solvent.) (D) A specific reaction cluster containing all reactions between substrates **1** and **2**. (E) t-SNE of the reactions in the selected cluster when the catalyst fingerprint is multiplied by three (top) and when the base fingerprint is multiplied by three (bottom). The three manifolds for each result are colored by product/internal standard values, catalyst, and base.

Results and Discussion

Figure 3 shows a reaction landscape from the discovery of a sp^3–sp^3 deaminative–decarboxylative carbon–carbon cross coupling reaction over multiple HTE experiments. In this study, the fingerprints of all acid electrophiles were multiplied by 3 ($w_e = 3$) before summing with the fingerprints of the other reagents. This trivial modification resulted in clusters of reactions based on electrophile.

As shown in Figure 3, with the acid electrophiles having an elevated weight, the clusters within the manifold are split into three distinct regimes reflecting the stages in the discovery and development of the chemistry. The discovery campaign began with using the free acid **4** as the model substrate (Figure 3 – Initial Reactions) but after a limit to the reactivity was realized, efforts moved to the NHPI activated acid **5** (Figure 3 – Methodology Development). In this case, the activated acid electrophile NHPI ester N-Boc-proline was used as the model substrate for the reaction, resulting in a large cluster with few substrates and many conditions. Thus, most of the reactions in this dataset fall within the NHPI activated (**5**) and free acid (**4**) N-Boc proline clusters colored in black (Figure 3B – Initial Reactions) and yellow (Figure 3B – Methodology Development) respectively. In the PROD/IS manifold shown in Figure 3C, a direct path can be followed as experiments drive the product output from 0% to close to 100% within the NHPI activated N-Boc proline cluster (Methodology Development). Once ideal conditions were developed for N-Boc proline, this reaction system was tested with a variety of other acid electrophile substrates, creating a tight cluster of many electrophile acids and few conditions (**6**-**26**, Scope Exploration). The efficacy of this system on these substrates is shown in the t-SNE with clusters of various acid electrophile substrates forming in different locations, each with their own PROD/IS distributions. Average PROD/IS results of each acid electrophile tested are shown below the plots in Figure 3C as well as the number of reactions they were tested in.

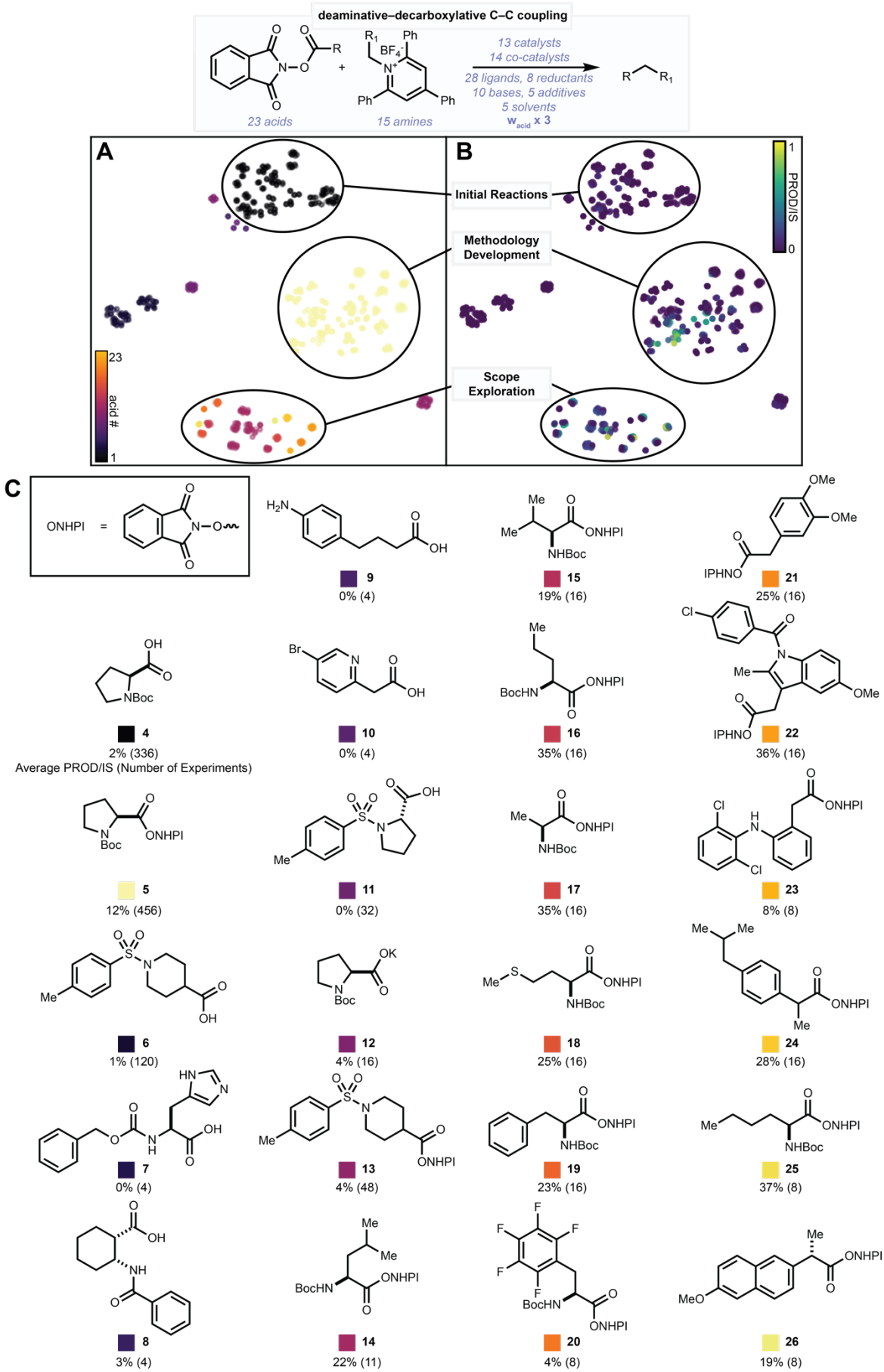

Figure 3. 1,296 reactions performed in an HTE format towards the development of a novel sp^3–sp^3 deaminative-decarboxylative carbon–carbon cross coupling (ref. 13). The manifold splits into three regimes, each representing a different stage of the discovery campaign. (A) The manifold points are colored by the acid used in the experiment. (B) The manifold points are colored

by the recorded PROD/IS value of the reaction. (C) All acids used in the experiment, with average PROD/IS values and the number of experiments they were used in.

In Figure 4, a case study analyzing the Suzuki dataset reveals reactivity cliffs – clusters that contain some reactions that fail and some that are successful, indicating minor changes to the system that cause the reaction to "flip" on or off.[27] An electrophile nucleophile substrate pair is identified to have a reactivity cliff in Figure 4A. The plots in Figure 4A are reaction array fingerprints only containing the fingerprints for the electrophile and nucleophile to emphasize focus on the substrate flexibility of the Suzuki reaction as opposed to the specifics of the reagents. Using this technique, a cluster containing a single nucleophile/electrophile pair was identified. On further analysis of this cluster with reactions that use electrophile **27** and nucleophile **28** to form **29**, it is revealed that this substrate pair was tested with four different solvents, six bases, and eight catalysts. In Figure 4B, we plot the reaction array fingerprints for these reactions, producing clean clusters that separate all components and producing a humanly interpretable explanation of the behavior behind the reactivity. Since experimental chemistry is rife with reactivity cliffs, oftentimes as subtle as a switch in solvent or order of reagent addition, it is critical to be able to visualize and interpret this behavior. From the color-coded solvent plot, it is clear that the reactivity of the substrate pair **28** and **29** is controlled by the solvent system used. The manifold directly identifies failing and working solvents when traversing the space from 1:9 ^{t}AmOH:NMP to 1:3 water:NMP. The addition of water in the solvent system is found to be critical in achieving desired reactivity as shown by another reactivity cliff between 1:3 water:NMP and pure NMP. A chi-squared analysis is reported in the Supporting Information to validate this finding (Figure S8). Thus, it can be rapidly identified which solvent systems poison the reaction even though a variety of reagents are being changed. This analysis is compared to boxplots of the solvent system used (Figure S9).

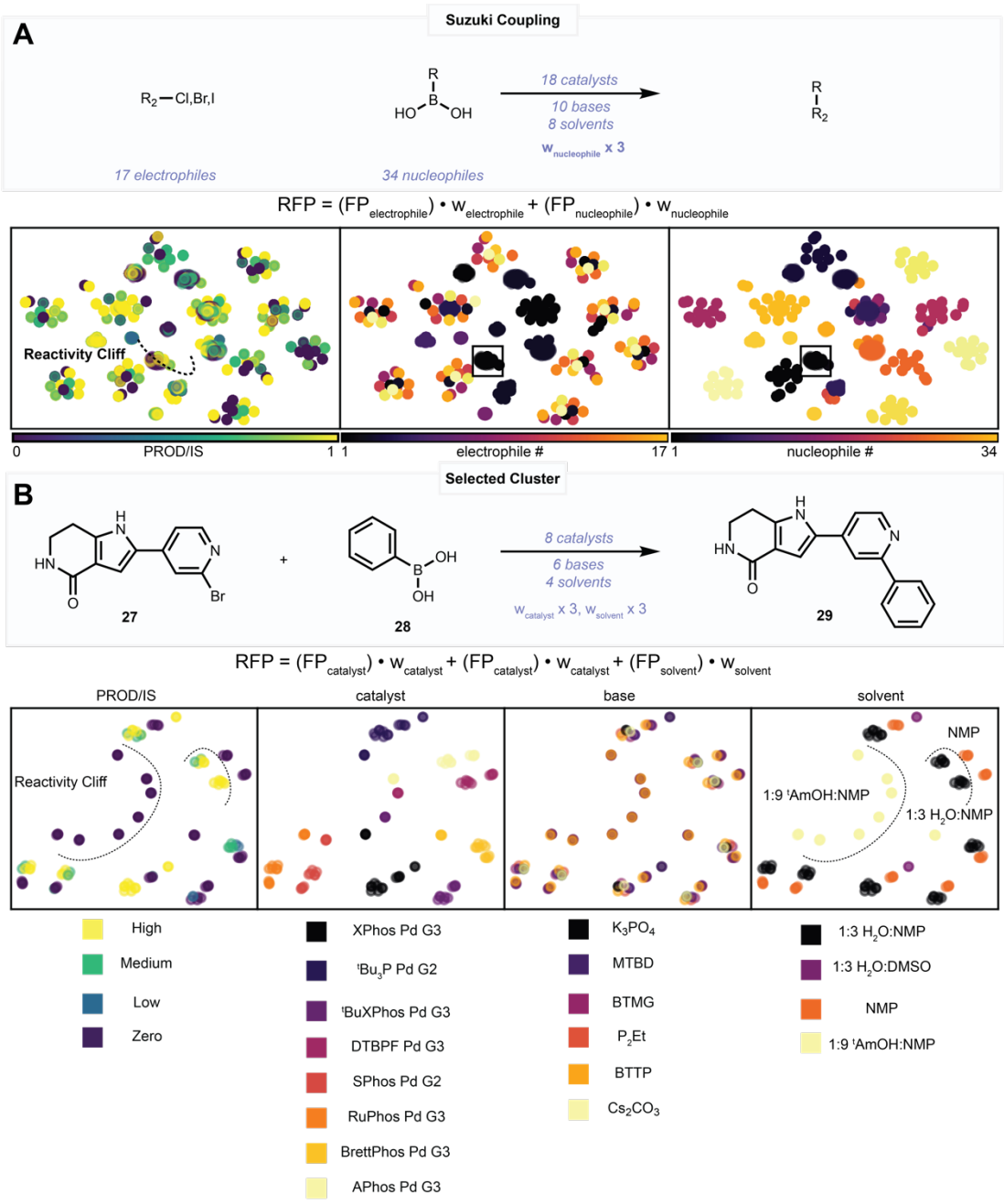

Figure 4. Reactivity cliffs are identified when focusing on specific substrate pairs. (A) Compounds **27** and **28** cluster together with high and poor performing reactions. (B) Repeating the analysis on the cluster that form **29** from **27** and **28** reveals reagents that cause this reactivity to flip on. Reactivity cliffs are readily identified between solvent regimes. The 1:3 water:NMP regime sits between two reactivity cliffs, separating it from the two failing solvent systems 1:9 *t*AmOH:NMP and pure NMP.

The creation of reaction array fingerprint manifolds is automatable. When using a HTE management system such as phactor or obtaining machine readable reaction data in a standardized format, reaction arrays can be rapidly analyzed using this method. In Figure 5, we show the automated creation of six reaction fingerprint t-SNEs utilizing output files procedurally generated from phactor. Hyperparameters including t-SNE perplexity and reagent weights (w_x) were optimized to best illuminate reactivity trends (see Supporting Information). In all cases, specific reagent classes were clustered. For instance, there are four clusters in Figure 5A representing the four ligands used in the reaction array which couples **30** and **14** to form **31**. These clusters are split into two subclusters, each representing one of the two reductants used. These clusters are then further split into three column-shaped clusters that represent the three catalysts and each of these columns consist of four points each corresponding to one of the four additives in the reaction array. Four clusters were formed in Figure 5B which shows 12 electrophile acids and 8 nucleophile amines, producing a successful reaction that uses **32** and **5** to form **33**. Each of the 12 acids formed a cluster in the manifold. In Figure 5C eight ligands and 12 catalysts were used to make **36** from **34** and **35**. The t-SNE clustered each reaction by ligand, resulting in eight clusters. In substrate scope ultraHTE experiments shown in Figure 5D, the t-SNE is clustered by the eight diazoniums used in the screen. In Figure 5D the reaction where **39** was formed from **37** and **38** is highlighted in the manifold. Finally, in the ultraHTE direct-to-biology assay shown in Figure 5E, a cluster is formed for each of the 80 amines

used in the reaction array. The reaction which amide inhibitor **42** is generated from the coupling of **40** and **41** is identified in the t-SNE. This visual format is compared against pivot table heatmaps in Figure S10.

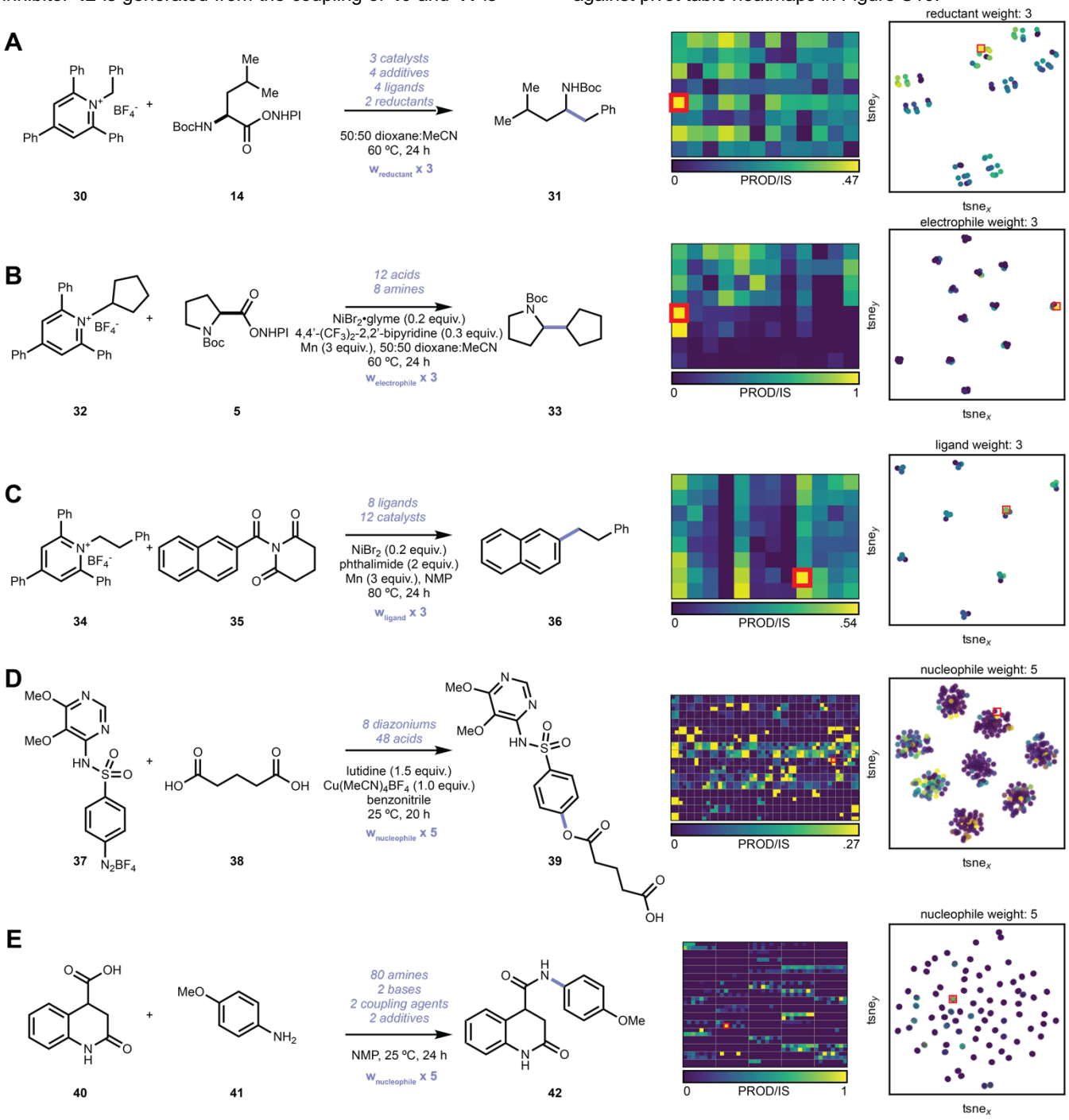

Figure 5. Reaction array manifolds can be procedurally generated from the output files produced by phactor. (A)-(F) reaction array results as reported in ref. 8 and their corresponding manifolds colored by output value. Perplexity and weights were modified as described in the Supporting Information to optimize the latent space for visualization.

Reaction fingerprints can be used to rapidly identify the best conditions for a given substrate pair and transformations. As part of our studies, we have been exploring mild benzoic acid decarboxylation methods. We discovered simple reaction conditions for a decarboxylative–deaminative amine–acid sp^2–sp^2 C–C coupling reaction between quinidinyl acids and amine diazoniums. This transformation proceeds when both substrates are added to solvent and irradiated by light, based on a recent report on irradiating diazoniums in methanol to achieve various chemistries.[28]

Ultimately 424 experiments were performed over a series of 24 and 96-well reaction arrays (Figure 6A).

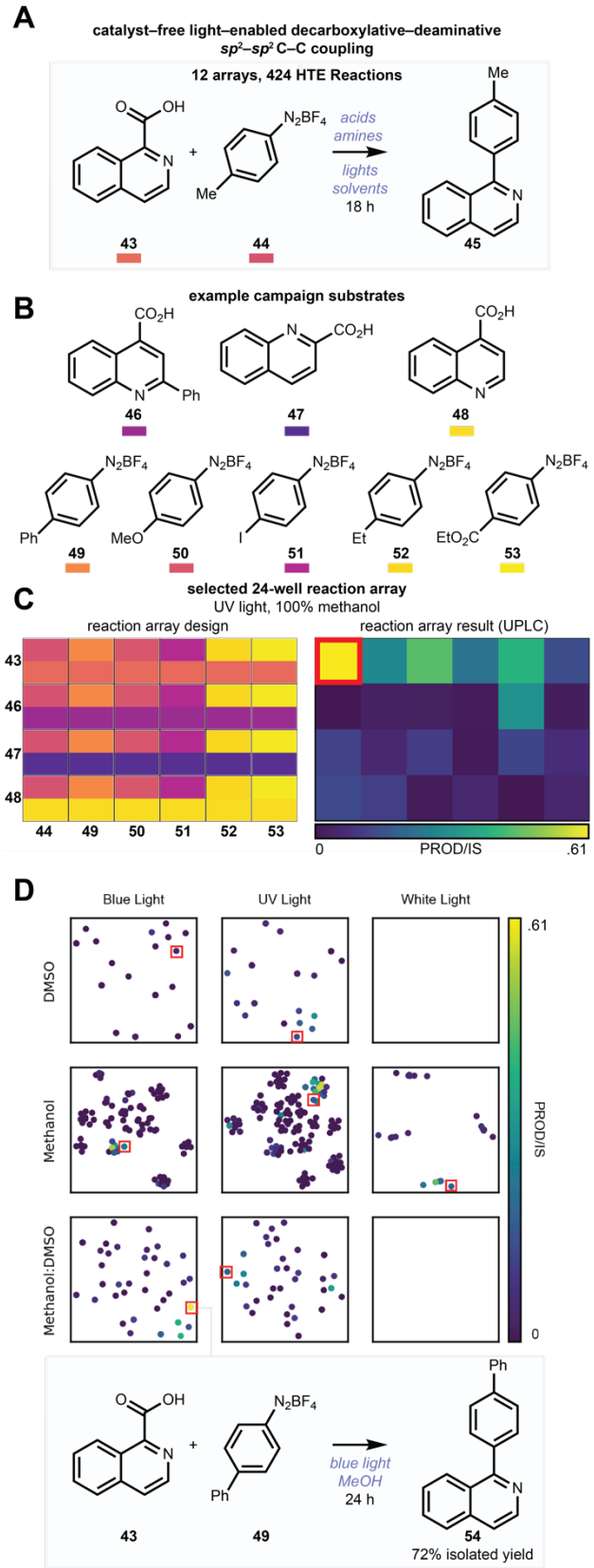

Figure 6. (a) Standard conditions for catalyst-free light-enabled decarboxylative–deaminative sp^2–sp^2 C–C coupling. (b) Sampling of substrates explored. (c) Screen design and execution. (d) Trellised reaction array fingerprint of experimental campaign. Each box contains all reactions performed for the given light and solvent regime. Reactions that perform well in one regime but not others are quickly identified.

Isoquinolines **43**, **46**, **47**, and **48**, and diazonium salts **44**, **49**, **50**, **51**, **52**, and **53** produced appreciable yields (measured by product/internal standard integrations) in methanol, DMSO, and a 1:1 mixture of methanol and DMSO (Figure 6B) to form biaryl products such as **45**. An array performed in the campaign is shown in Figure 6C, where the four acids and six diazonium salts in Figure 6B were irradiated in methanol under UV light for 18 hours. For the reaction between **43** and **49** to form **54**, we are able to rapidly identify from the HTE results and reaction fingerprint arrays (Figure 6D) that the reaction works best in blue light and with 50:50 DMSO:methanol, as we can directly compare the PROD/IS result against the reaction performed in other regimes.

Conclusion

Weighted reaction fingerprinting is an easy-to-perform method for the analysis of massive reaction datasets. The algorithm allows chemists to rapidly navigate through large collections of reaction data. Reaction embeddings can be easily optimized and modified by changing the reagent weights.

Corresponding Author

* Tim Cernak – Department of Medicinal Chemistry, College of Pharmacy, University of Michigan, Ann Arbor, MI, USA 48104; orcid.org/0000-0001-5407-0643.

Email: tcernak@med.umich.edu

Author Contributions

B.M performed computational experiments and wet chemistry experiments described in Fig. 6. J.H. performed computational experiments and D.S. performed data harmonization. T.C and B.M. wrote the manuscript.

Conflicts of Interest

T.C. holds equity in Scorpion Therapeutics, and is a co-Founder and equity holder of Entos, Inc.

Acknowledgements

We gratefully acknowledge funding from the National Science Foundation (CHE-2236215) (T.C.) and an ACS MEDI Predoctoral Fellowship (B.M.). Initial studies were funded by startup funds from the University of Michigan College of Pharmacy.

Notes and references

(1) Cernak, T.; Gesmundo, N. J.; Dykstra, K.; Yu, Y.; Wu, Z.; Shi, Z.-C.; Vachal, P.; Sperbeck, D.; He, S.; Murphy, B. A. Microscale high-throughput experimentation as an enabling technology in drug discovery: Application in the discovery of (Piperidinyl) pyridinyl-1 H-benzimidazole diacylglycerol acyltransferase 1 inhibitors. *Journal of Medicinal Chemistry* **2017**,

60 (9), 3594-3605. DOI: 10.1021/acs.jmedchem.6b01543.
(2) Douthwaite, J. L.; Zhao, R.; Shim, E.; Mahjour, B.; Zimmerman, P.; Cernak, T. Formal Cross-Coupling of Amines and Carboxylic Acids to Form sp3–sp2 Carbon–Carbon Bonds. *Journal of the American Chemical Society* **2023**, *145* (20), 10930-10937. DOI: 10.1021/jacs.2c11563.
(3) Gesmundo, N.; Dykstra, K.; Douthwaite, J.; Mahjour, B.; Ferguson, R.; Dreher, S.; Sauvagnat, B.; Sauri, J.; Cernak, T. Miniaturization of Popular Reactions from the Medicinal Chemists' Toolbox for Ultrahigh-Throughput Experimentation. **2022**.
(4) 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. Chemistry informer libraries: a chemoinformatics enabled approach to evaluate and advance synthetic methods. *Chemical Science* **2016**, *7* (4), 2604-2613. DOI: 10.1039/C5SC04751J.
(5) Lin, S.; Dikler, S.; Blincoe, W. D.; Ferguson, R. D.; Sheridan, R. P.; Peng, Z.; Conway, D. V.; Zawatzky, K.; Wang, H.; Cernak, T. Mapping the dark space of chemical reactions with extended nanomole synthesis and MALDI-TOF MS. *Science* **2018**, *361* (6402), eaar6236. DOI: 10.1126/science.aar6236.
(6) Mahjour, B.; Shen, Y.; Liu, W.; Cernak, T. A map of the amine–carboxylic acid coupling system. *Nature* **2020**, *580* (7801), 71-75. DOI: 10.1038/s41586-020-2142-y.
(7) Mahjour, B.; Shen, Y.; Cernak, T. Ultrahigh-throughput experimentation for Information-Rich Chemical Synthesis. *Accounts of Chemical Research* **2021**, *54* (10), 2337-2346. DOI: 10.1021/acs.accounts.1c00119.
(8) Mahjour, B.; Zhang, R.; Shen, Y.; McGrath, A.; Zhao, R.; Mohamed, O. G.; Lin, Y.; Zhang, Z.; Douthwaite, J. L.; Tripathi, A.; Cernak, T. Rapid Planning and Analysis of High-Throughput Experiment Arrays for Reaction Discovery. **2022**.
(9) McGrath, A.; Zhang, R.; Shafiq, K.; Cernak, T. Repurposing amine and carboxylic acid building blocks with an automatable esterification reaction. *Chemical Communications* **2023**, *59*, 1026-1029. DOI: 10.1039/D2CC05670D.
(10) Buitrago Santanilla, A.; 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. Nanomole-scale high-throughput chemistry for the synthesis of complex molecules. *Science* **2015**, *347* (6217), 49. DOI: 10.1126/science.1259203.
(11) Shen, Y.; Mahjour, B.; Cernak, T. Development of copper-catalyzed deaminative esterification using high-throughput experimentation. *Communications Chemistry* **2022**, *5* (1), 83. DOI: 10.1038/s42004-022-00698-0.
(12) Uehling, M. R.; King, R. P.; Krska, S. W.; Cernak, T.; Buchwald, S. L. Pharmaceutical diversification via palladium oxidative addition complexes. *Science* **2019**, *363* (6425), 405. DOI: 10.1126/science.aac6153.
(13) Zhang, Z.; Cernak, T. The Formal Cross-Coupling of Amines and Carboxylic Acids to Form sp3–sp3 Carbon–Carbon Bonds. *Angewandte Chemie International Edition* **2021**, *60* (52), 27293-27298. DOI: 10.1002/anie.202112454.
(14) Gesmundo, N. J.; Sauvagnat, B.; Curran, P. J.; Richards, M. P.; Andrews, C. L.; Dandliker, P. J.; Cernak, T. Nanoscale synthesis and affinity ranking. *Nature* **2018**, *557* (7704), 228-232. DOI: 10.1038/s41586-018-0056-8.
(15) Kearnes, S. M.; Maser, M. R.; Wleklinski, M.; Kast, A.; Doyle, A. G.; Dreher, S. D.; Hawkins, J. M.; Jensen, K. F.; Coley, C. W. The Open Reaction Database. *Journal of the American Chemical Society* **2021**, *143* (45), 18820-18826. DOI: 10.1021/jacs.1c09820.
(16) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Modeling* **1988**, *28* (1), 31-36. DOI: 10.1021/ci00057a005.
(17) Zhang, R.; Mahjour, B.; Cernak, T. Exploring the Combinatorial Explosion of Amine–Acid Reaction Space via Graph Editing. **2022**.
(18) Shim, E.; Kammeraad, J. A.; Xu, Z.; Tewari, A.; Cernak, T.; Zimmerman, P. M. Predicting reaction conditions from limited data through active transfer learning. *Chemical Science* **2022**, *13* (22), 6655-6668. DOI: 10.1039/D1SC06932B.
(19) Shen, Y.; Borowski, J. E.; Hardy, M. A.; Sarpong, R.; Doyle, A. G.; Cernak, T. Automation and computer-assisted planning for chemical synthesis. *Nature Reviews Methods Primers* **2021**, *1* (1), 1-23. DOI: 10.1038/s43586-021-00022-5.

(20) Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *Journal of Chemical Documentation* **1965**, *5* (2), 107-113. DOI: 10.1021/c160017a018.
(21) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling* **2010**, *50* (5), 742-754. DOI: 10.1021/ci100050t.
(22) Shields, B. J.; Stevens, J.; Li, J.; Parasram, M.; Damani, F.; Alvarado, J. I. M.; Janey, J. M.; Adams, R. P.; Doyle, A. G. Bayesian reaction optimization as a tool for chemical synthesis. *Nature* **2021**, *590* (7844), 89-96. DOI: 10.1038/s41586-021-03213-y.
(23) Sandfort, F.; Strieth-Kalthoff, F.; Kühnemund, M.; Beecks, C.; Glorius, F. A structure-based platform for predicting chemical reactivity. *Chem* **2020**, *6* (6), 1379-1390. DOI: 10.1016/j.chempr.2020.02.017.
(24) ChemAxon. Reaction fingerprints. *https://docs.chemaxon.com/display/docs/reaction-fingerprint-rf.md* **2022**.
(25) Schneider, N.; Lowe, D. M.; Sayle, R. A.; Landrum, G. A. Development of a novel fingerprint for chemical reactions and its application to large-scale reaction classification and similarity. *J Chem Inf Model* **2015**, *55* (1), 39-53. DOI: 10.1021/ci5006614. From NLM Medline.
(26) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL keys for use in drug discovery. *J Chem Inf Comput Sci* **2002**, *42* (6), 1273-1280. DOI: 10.1021/ci010132r. From NLM Medline.
(27) Stumpfe, D.; Bajorath, J. r. Exploring Activity Cliffs in Medicinal Chemistry. *Journal of Medicinal Chemistry* **2012**, *55* (7), 2932-2942. DOI: 10.1021/jm201706b.
(28) Witzel, S.; Hoffmann, M.; Rudolph, M.; Kerscher, M.; Comba, P.; Dreuw, A.; Hashmi, A. S. K. Excitation of aryl cations as the key to catalyst-free radical arylations. *Cell Reports Physical Science* **2021**, *2* (2), 100325. DOI: 10.1016/j.xcrp.2021.100325.