EurJOC
European Journal of Organic Chemistry

Research Article
doi.org/10.1002/ejoc.202400428

Chemistry Europe
European Chemical Societies Publishing

www.eurjoc.org

Multi-Threshold Analysis for Chemical Space Mapping of Ni-Catalyzed Suzuki-Miyaura Couplings

Austin LeSueur,[a] Nari Tao,[b] Abigail Doyle,*[b] and Matthew Sigman*[a]

A key challenge in synthetic chemistry is the selection of high-performing ligands for cross-coupling reactions. To address this challenge, this work presents a classification workflow to identify physicochemical descriptors that bin monophosphine ligands as active or inactive in Ni-catalyzed Suzuki-Miyaura coupling reactions. Using five previously published high-throughput experimentation datasets for training, we found that a binary classifier using a phosphine's minimum buried volume and Boltzmann-averaged minimum electrostatic potential is most effective at distinguishing high and low-yielding ligands. Experimental validations are also presented. Using the two physicochemical descriptors from the binary classifier to represent the chemical space of monophosphine ligands leads to a more predictive guide for structure-reactivity relationships compared with classic chemical space representations.

Introduction

One of the most interesting physical organic questions in the development of catalysts for a particular reaction is why one ligand is effective and another one that appears structurally similar is not. As a step toward deconvoluting these subtleties, our teams recently reported the integration of simple classification algorithms with an extensive feature database of monodentate phosphine descriptors to uncover a "threshold" reactivity relationship between ligand properties and effectiveness in cross-coupling catalysis (Figure 1A).[1] In particular, the descriptor $\%V_{bur}$(min)[2] was found through a set of computational and experimental studies to be related to the ligation state of a given metal, which was ultimately tied to catalytic function. While initially developed for late transition metal catalyzed cross-coupling reactions, both the classification tool and the parameter have found use in a wide range of applications since the initial report.[3–14]

A nuance of the algorithm (a single-node decision tree) is that we purposefully penalize false negatives to ensure that effective examples are not missed for downstream applications, such as virtual screening or secondary statistical modeling campaigns. In contrast, the algorithm does not penalize false positives to the same extent, ultimately resulting in many more examples of inactive catalysts in this regime of the confusion matrix. It is not surprising that many false positives are found as it suggests that another ligand feature is responsible for poor performance. Therefore, we were inspired to reinvestigate the original datasets to determine if another descriptor can be identified to provide a second reactivity "threshold" (Figure 1B). Herein we report the application of this strategy to five Ni-catalyzed Suzuki–Miyaura coupling (SMC) reactions. Our investigation revealed that in addition to the threshold in $\%V_{bur}$(min), ligand activity is also gated by the Boltzmann averaged minimum electrostatic potential at the phosphorus atom [V_{min}-(Boltz)]. This parameter is correlated to the original experimental Tolman Electronic Parameter (TEP)[15] that is defined as the CO stretching frequency of the resultant $LNi(CO)_3$ (L = monodentate phosphine) complex.[16]

[a] A. LeSueur, M. Sigman
Department of Chemistry, University of Utah, Salt Lake City UT 84112, USA
E-mail: sigman@chem.utah.edu

[b] N. Tao, A. Doyle
Department of Chemistry and Biochemistry, University of California, Los Angeles CA 90095, USA
E-mail: agdoyle@chem.ucla.edu

Supporting information for this article is available on the WWW under https://doi.org/10.1002/ejoc.202400428

Results and Discussion

Design of the Multi-Threshold Algorithm

The multi-threshold algorithm presented here expands on the previously reported threshold analysis method. As the first step, the chemist identifies a cutoff in the dataset's output (y-cut) that will be used to classify ligands as active or inactive. The cutoff is selected by the chemist to explore a given hypothesis. As examples, one could select a certain conversion value with a given catalyst loading as a possible y cutoff to interrogate turnover or simply a natural division in the dataset such as the conversion for the ligand-less control in a given reaction to eliminate low-performing ligands. Each ligand in the dataset is then classed as positive (active) or negative (inactive) based on its experimental output and the y cutoff.

With the ligands sorted into their classes, the chemist also must identify the relative importance of each class. For applications in reaction optimization or ligand evaluation, a simple ratio of ligands classified correctly by the algorithm to total number of ligands (accuracy) is an insufficient metric as it is generally more important to identify the maximum number of active ligands (recall) than to exclude inactive ligands (precision). Specifying the relative importance of these two tasks is left to the user by allowing them to define the relative importance or weight of each class in the algorithm (Figure 1C).

Research Article
doi.org/10.1002/ejoc.202400428

Chemistry Europe
European Chemical Societies Publishing

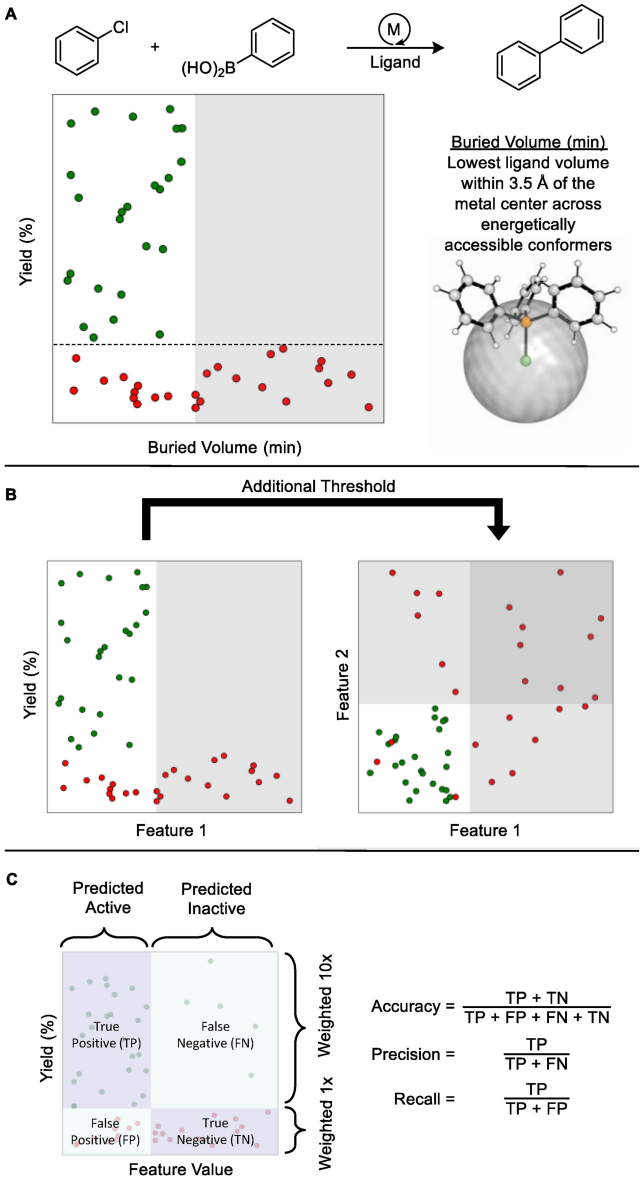


Figure 1. (A) Previous work identifying a reactivity threshold in minimum buried volume of phosphine ligands for Ni-catalyzed Suzuki–Miyaura Coupling reactions. (B) This work expanding the algorithm to identify additional thresholds to further narrow down the active chemical space. (C) A confusion matrix assists in the evaluation of classification algorithms. Ligands are sorted into bins for those that are classified correctly as active (true positives), classified incorrectly as active (false positives), classified correctly as inactive (true negatives) and classified incorrectly as inactive (false negatives). Accuracy is the ratio of those classified correctly to those classified incorrectly. Precision represents the ratio of ligands predicted to be active that are truly active. Recall is the ratio of ligands that are experimentally active that were identified as active by the algorithm.

The algorithm scans through all descriptors using single-node decision trees, identifying descriptors and their values that best separate the defined ligand classes. By default, the algorithm prioritizes classifying active ligands correctly at the expense of inactive ligands through a weighting function of ten to one. This increases the resulting threshold's recall score (active points classified correctly/total number of points) and minimizes the number of active ligands excluded from down-stream processes. One method to visualize the results is a confusion matrix wherein you can determine "true" positives, "true" negatives, "false" positives, and "false" negatives as well as the common metrics we evaluate (Figure 1C).

As the next step, we have expanded the algorithm to build multi-node decision trees to determine combined reactivity thresholds. In addition to the same user-controlled y-cutoff and class weighting, the user can also specify how many sequential nodes to evaluate. The hypothesis is that each sequential node will provide more specificity as to which ligand features are most important for catalytic performance. Similarly, it can also increase classification accuracy, resulting in a more defined search space for a ligand optimization campaign. Below we analyze five previously reported[1] SMC reactions (Figure 2) as case studies to demonstrate the utility of this method.

Examination of the General 2D Classification and Validation

Using the *kraken* monophosphine descriptor library[17] as the feature source and yield as the experimental output, we deployed the two-dimensional threshold or two-node decision tree workflow on Ni-catalyzed SMC reaction datasets that comprised different aryl chloride and aryl boronic acid pairings, as shown in Figure 2. As a first step, we combined all of the data from each reaction (90 ligands repeated over five reactions, 450 reactions total) (Tables S1–S2) and deployed the algorithm. As ligand-less reactivity is different for each reaction,

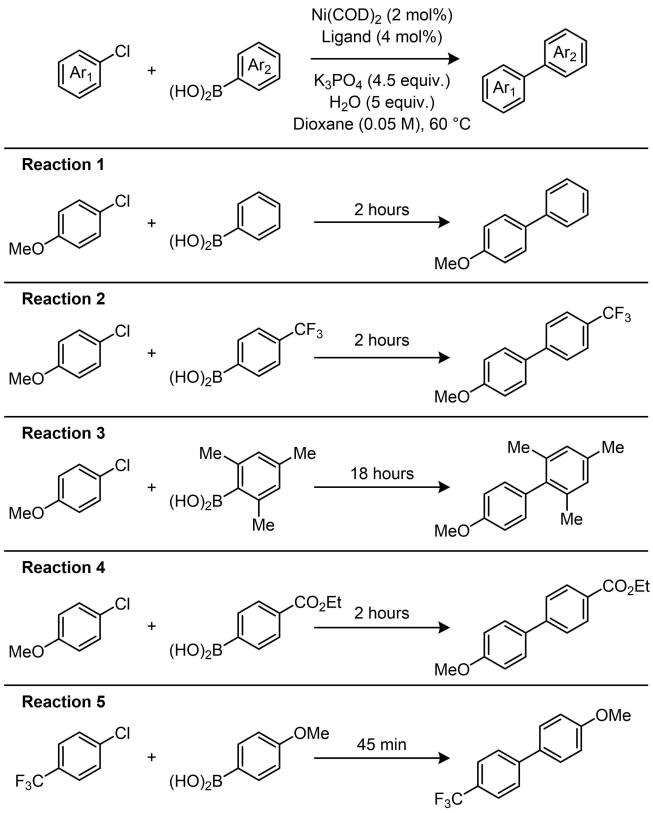


Figure 2. SMC reactions used as case studies for additional threshold analysis.

we set the y-cut at 10% after subtracting the ligand-less yield from the reported yield. An example of a high performing model for this dataset included the same V_{bur}(min) threshold previously detected with the original single-node decision tree in combination with a second threshold using V_{min}(Boltz), which is the minimum of the molecular electrostatic potential (MESP) on the phosphorus atom that is Boltzmann averaged across all energetically accessible conformers (Figure 3). The significance of these features is supported by recent studies that also identified them or related features as relevant to mono-phosphine catalysis.[14,18]

Statistically, the two-node decision tree has a combined classification accuracy of 81%, precision of 66%, and recall of 97% as compared to an accuracy of 66%, precision of 51%, and recall of 98% with V_{bur}(min) alone (Figure S2). The double threshold's ability to exclude an additional subsection of inactive ligands, shown by the increased accuracy and precision, while maintaining strong identification of active ligands is of particular note, as the primary drawback of the single threshold alone was the high number of false positive ligand classifications.

We have previously investigated the implications of a V_{bur}(min) threshold in terms of ligation state of the metal.[1] The discovery that a V_{min}(Boltz) threshold further classifies inactive ligands is consistent with the original work of Tolman in defining electronic features of phosphines as this parameter is inversely correlated to the TEP. In the context of Ni-catalyzed cross coupling reactions of aryl chlorides, the electronic character of a phosphine ligand has been shown to impact the reaction outcome in two primary ways: (1) more electron-rich ligands displace cyclooctadiene from the Ni(0) precatalyst more effectively than electron-deficient ligands, affording higher concentrations of an active catalyst in situ.[19] Alternatively, (2) more electron-rich phosphines accelerate oxidative addition and prevent generation of catalytically inactive Ni(I) intermediates, which occurs via comproportionation if both Ni(0) and Ni(II) are present in high concentration.[20] Previous mechanistic studies have also implicated the formation of inactive μ_2-OH Ni(II) dimers as responsible for poor performance in Ni-catalyzed SMC reactions.[21,22] Phosphine ligand electronics could also impact the equilibrium generation of this inactive species. Identification of the V_{min} threshold is consistent with these mechanistic scenarios.

While these two features appear to be orthogonal for the purpose of classification, they are not entirely independent of one another. This can be seen in the relative lack of ligands in the top right quadrant representing ligands with a high V_{bur}(min) and high V_{min}(Boltz) in both a common screening set (Figure 3) and in the full *kraken* library (Figure S11). One possible reason for this observation is that the additional steric bulk required for a high V_{bur}(min) generally includes more donating alkyl/aryl groups, raising the V_{min}(Boltz) as well.

To validate this general double threshold classification, we screened additional ligands in reactions 2 and 4 (Figure 4). The validation ligands were selected from a distribution of accessible ligands predicted to contain both active and inactive examples. An effort was made to distribute validation ligands across the double-threshold space, although, as previously mentioned, the upper-right quadrant contains fewer easily accessible ligands than the rest and therefore received less

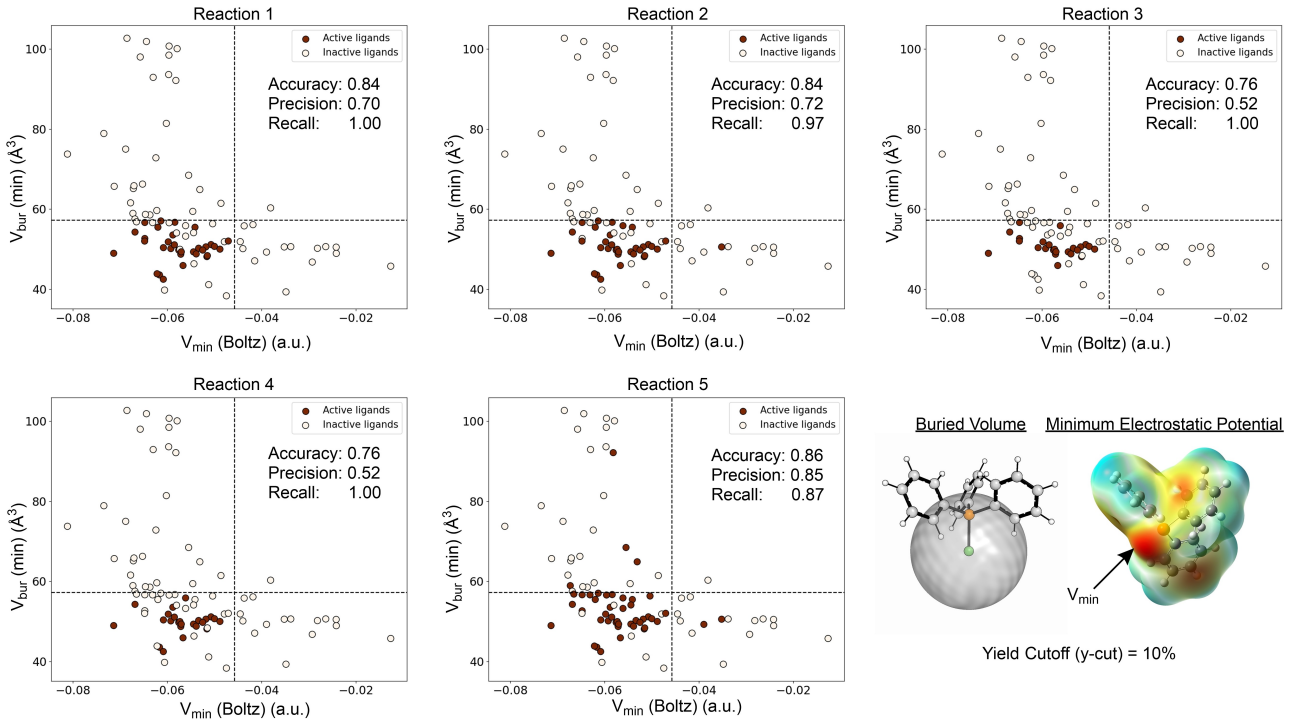


Figure 3. A general two-parameter threshold shown for each SMC reaction. Ligands are classified as active with yields over 10%. Ligands with a minimum buried volume below 57.277 Å^3 and an average minimum electrostatic potential below −0.046 a.u. are predicted to be active.

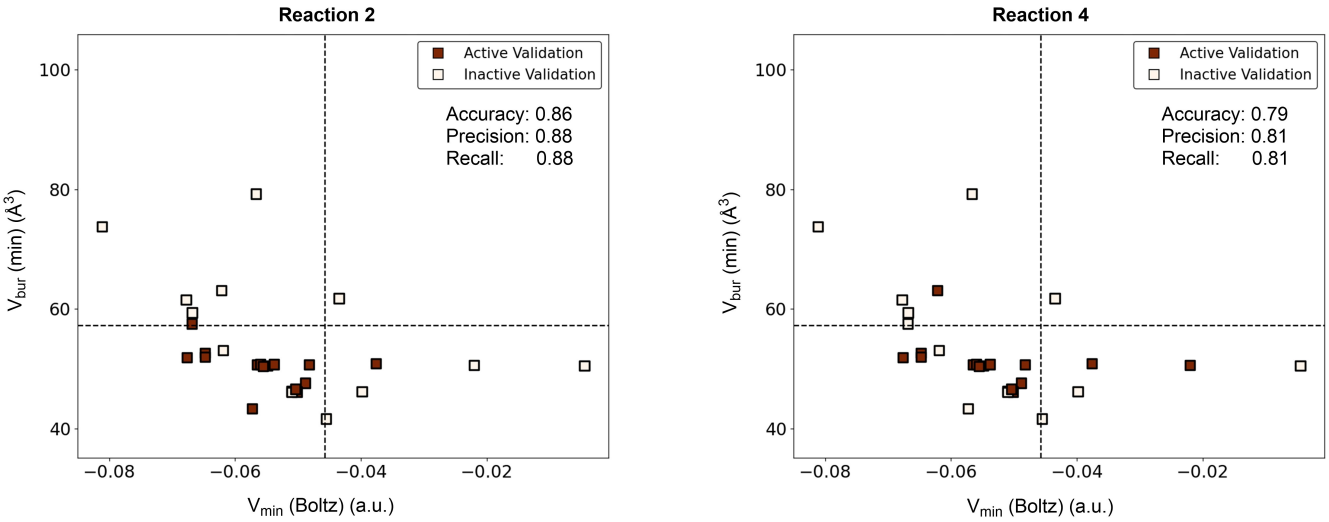


Figure 4. Double threshold validations of reactions 2 and 4.

focus in our validation set. A total of 28 ligands were evaluated for each reaction (Tables S8–S9).

The double thresholds validated well, with an average classification accuracy of 83%, compared to training set accuracy of 81%. Precision was improved in the validation sets, with an average of 85%, 19% greater than the training set. This may be in part due to the fact that the validation set was collected independent of the HTE studies, on a benchtop scale and with a slightly modified experimental procedure. For phosphines repeated under both conditions, the outcomes were well correlated but the benchtop conditions were ~15% higher-yielding than the HTE conditions (see supporting information).

Conversely, recall was somewhat worse in the validation of reactions 2 and 4 at an average of 84.5%, with one active ligand being misclassified by the V_{min}(Boltz) threshold in reaction 2 and two in reaction 4. This may be similarly impacted by the overall higher yields of the validation set, but also demonstrates the limitations of this classification method. Interestingly, P(4-FPh)$_3$ appears to be an outlier in the V_{min} threshold across multiple substrate pairings, perhaps suggesting a distinct role for the fluoro substituent rather than simply serving as an inductive electron withdrawing group to the phosphorous donor.

It is important to note that this classification model is not universally applicable. As the model was trained on Ni-catalyzed SMC reactions using simple monophosphine ligands, it cannot be assumed to classify other types of ligands or ligands used in other reactions. For example, application of this model to a series of previously reported Pd-catalyzed couplings yielded poor predictive accuracy (Figures S4–S9). Similarly, all SMC reactions used to train the model were performed under constant conditions and departure from these conditions may impact the chemistry in a manner that the model has not seen and therefore is unequipped to predict.

Comparison to Traditional Chemical Space Representations

This multi-threshold analysis workflow is based on the concept of classifying regions of chemical space as either active or inactive as a function of multiple descriptors. Chemical space can be defined as the subset of molecules that are relevant to a certain application and it is interesting to consider the best way to represent these chemical spaces.[23,24] For example, a chemical space comprising all commercially available or easily synthesizable monophosphine ligands is a relevant chemical space for a ligand screening campaign. Chemical spaces like these are often represented by molecular descriptor libraries (i.e., the *kraken* database) where the Euclidean distance measured in these descriptors can be used to define similarity of molecules, with those most similar closer to one another and diverse molecules further apart. The utility of this theoretical space is in the assumption that molecules near each other will have similar reactivity if the molecular descriptors capture the structure function relationships appropriately.

An excellent and early example is the chemical space maps for monophosphines originally reported by Tolman and later elaborated by Fey wherein the ligand's cone angle is plotted against its TEP, thereby grouping ligands with similar steric and electronic characteristics.[15,25] This can be converted into a reactivity space by subsequently overlaying reaction output (e.g., yield or another measurable) onto the map in the third dimension, thereby identifying the characteristics of high-performing ligands. With the development of more modern physicochemical descriptors and computational methods, chemical spaces can now be developed using a range of descriptors. For example, with the *kraken* library, a chemical space for monophosphine ligands can be defined by using the 190 parameters given for each ligand. While there is far more information in a chemical space defined by the *kraken* library than in the Tolman plots, it requires dimensionality reduction techniques to visualize and interpret the high dimensional space. Most commonly, algorithms like principal component

Research Article
doi.org/10.1002/ejoc.202400428

analysis (PCA)[26] or uniform manifold approximation and projection (UMAP)[27] are utilized to build representations while retaining the maximum amount of information about molecule diversity and relative positioning (Figure 5A and 5B). PCA maps of the *kraken* space have been used successfully for designing diverse training sets[28] and defining the domain of applicability of a statistical model.[7]

The primary drawback to dimensionality reduction of *kraken* is the possible loss of specificity. While principal components of a high-dimensional space are useful for generalizing the space, many chemical applications of the space require the precision that comes from the individual descriptors. Perhaps more importantly, not every descriptor in the library will be relevant to every reaction. The use of multi-threshold analysis provides a bridge between the interpretability of a Tolman map and the quantity of information in modern descriptor libraries. Instead of trying to combine all available descriptors into a comprehensive space, this workflow identifies a finite number of descriptors most relevant to a given reaction based on their ability to divide data points into active and inactive categories. In essence it produces a Tolman-like map tailored to a given reaction – and perhaps will inspire designer screening sets for future applications in late metal catalysis.

To illustrate these differences, plots of the Tolman map, PCA and UMAP *kraken* space, and double threshold are shown side by side as a function of yield in reaction 1 (Figure 5). In both dimensionality reduction representations (5 A and 5B), the potentially active ligands are spread across the chemical space in a way that would make intuitive selection of high-activity areas difficult for the chemist when screening ligands for a

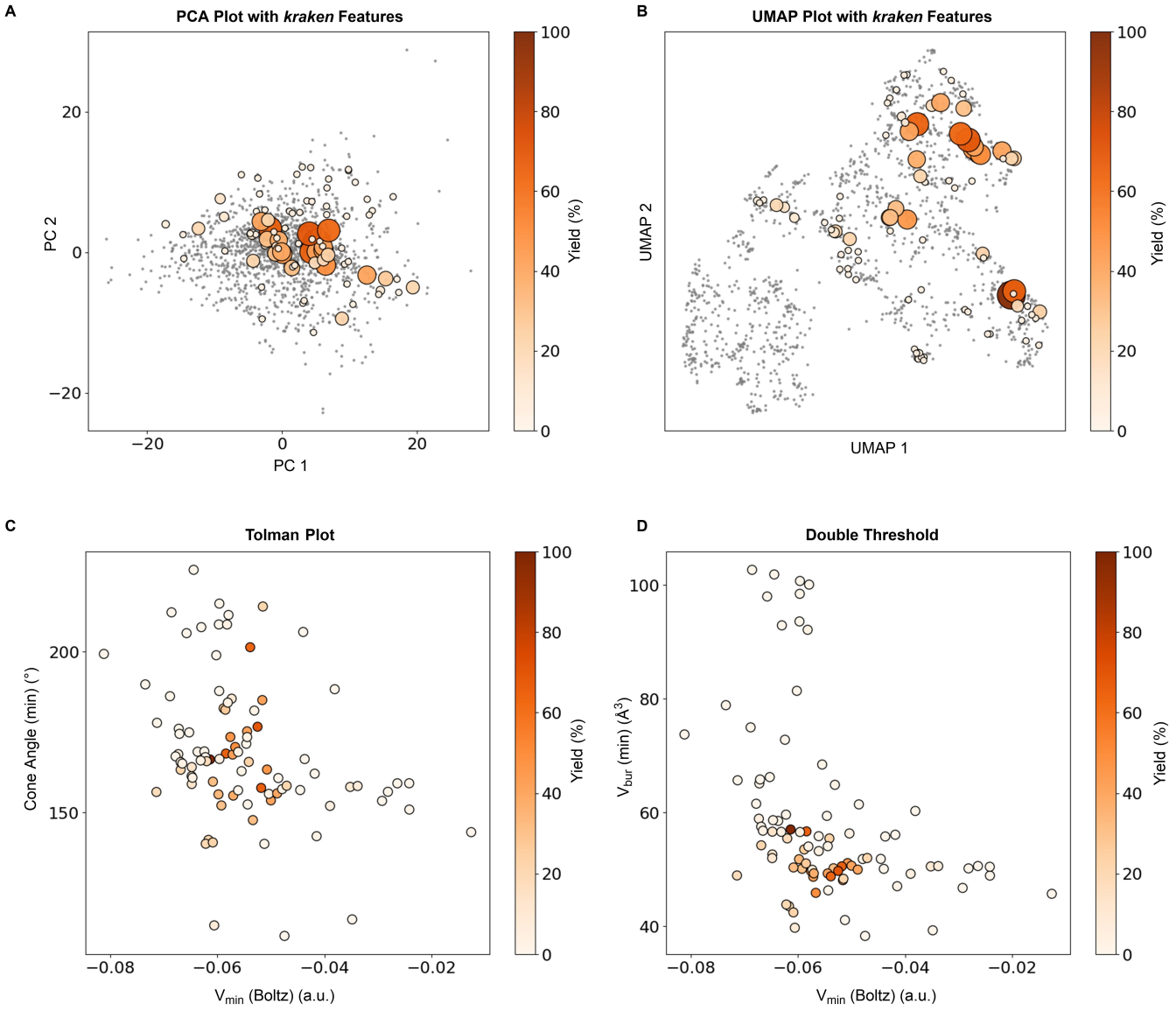


Figure 5. Chemical space representation comparisons. (A) Reaction 1 plotted on PCA space of the *kraken* library with yields represented as bubble size and color. Ligands in the *kraken* library not used in this study are shown as gray points. (B) Reaction 1 plotted on the UMAP representation of the *kraken* library. (C) Traditional Tolman map of Reaction 1. Active ligands are grouped in the center, but not clearly separable from inactive ligands. (D) Two-dimensional threshold space of Reaction 1. Active ligands are more separable from inactive ligands.

reaction. In the UMAP particularly, the active ligands are distributed among many clusters so that it would be difficult to use the UMAP alone to identify areas of activity. By selecting a few features that are most relevant to our chemistry of interest though, we can build a chemical space representation that more accurately reflects relevant ligand space for the SMC reactions.

Conclusions

The multi-threshold analysis strategy presented here is a natural expansion of the previously published threshold analysis. By extending the number of descriptors in the classification method, more information can be learned about both the ligand effects on a given reaction and the suitability of untested ligands for that reaction. This work specifically shows the impact of phosphine electronics on the SMC reaction, in addition to the ligand's V_{bur}(min). Additionally, the double threshold delineated the active chemical space for this reaction in a way that traditional chemical space representations like PCA and UMAP were unable to. By crafting a chemical space tailored to the SMC reaction, we were able to identify additional ligands that were likely to be active and increase the ratio of hits in subsequent validation.

Supporting Information Summary

The authors have cited additional references within the Supporting Information.[29–34]

Acknowledgements

M.S.S., A.G.D., N.T., and A.L. acknowledge generous financial support from the NSF under the CCI Center for Computer Assisted Synthesis (CHE-1925607 and CHE-2202693).

Conflict of Interests

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

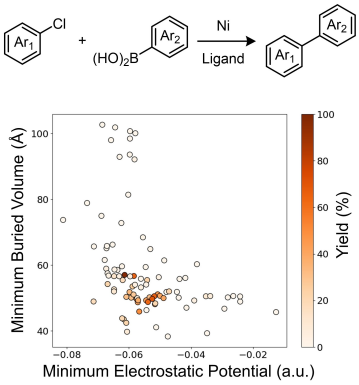
Keywords: Data science · Homogeneous catalysis · Ligand optimization · P ligands

[1] S. H. Newman-Stonebraker, S. R. Smith, J. E. Borowski, E. Peters, T. Gensch, H. C. Johnson, M. S. Sigman, A. G. Doyle, *Science* **2021**, *374*, 301–308.
[2] H. Clavier, S. P. Nolan, *Chem. Commun.* **2010**, *46*, 841–861.
[3] M. H. Samha, L. J. Karas, D. B. Vogt, E. C. Odogwu, J. Elward, J. M. Crawford, J. E. Steves, M. S. Sigman, *Sci. Adv.* **2024**, *10*, eadn3478.
[4] J. Rein, S. D. Rozema, O. C. Langner, S. B. Zacate, M. A. Hardy, J. C. Siu, B. Q. Mercado, M. S. Sigman, S. J. Miller, S. Lin, *Science* **2023**, *380*, 706–712.
[5] J. P. Liles, C. Rouget-Virbel, J. L. H. Wahlman, R. Rahimoff, J. M. Crawford, A. Medlin, V. S. O'Connor, J. Li, V. A. Roytman, F. D. Toste, M. S. Sigman, *Chem* **2023**, *9*, 1518–1537.
[6] J. Rein, J. M. Meinhardt, J. L. Hofstra Wahlman, M. S. Sigman, S. Lin, *Angew. Chem. Int. Ed.* **2023**, *62*, e202218213.
[7] J. J. Dotson, L. van Dijk, J. C. Timmerman, S. Grosslight, R. C. Walroth, F. Gosselin, K. Püntener, K. A. Mack, M. S. Sigman, *J. Am. Chem. Soc.* **2023**, *145*, 110–121.
[8] S. H. Newman-Stonebraker, J. Y. Wang, P. D. Jeffrey, A. G. Doyle, *J. Am. Chem. Soc.* **2022**, *144*, 19635–19648.
[9] R. C. Cammarota, W. Liu, J. Bacsa, H. M. L. Davies, M. S. Sigman, *J. Am. Chem. Soc.* **2022**, *144*, 1881–1898.
[10] D. M. Lustosa, A. Milo, *ACS Catal.* **2022**, *12*, 7886–7906.
[11] J. Chen, R. Zhang, *J. Phys. Chem. Lett.* **2022**, *13*, 520–526.
[12] A. A. Barboza, J. A. Dantas, G. A. M. Jardim, A. C. Neto, W. X. C. Oliveira, T. Gensch, M. A. B. Ferreira, *ChemCatChem* **2022**, *14*, e202200894.
[13] A. Modak, J. V. Alegre-Requena, L. de Lescure, K. J. Rynders, R. S. Paton, N. J. Race, *J. Am. Chem. Soc.* **2022**, *144*, 86–92.
[14] D. Zell, C. Kingston, J. Jermaks, S. R. Smith, N. Seeger, J. Wassmer, L. E. Sirois, C. Han, H. Zhang, M. S. Sigman, F. Gosselin, *J. Am. Chem. Soc.* **2021**, *143*, 19078–19090.
[15] C. A. Tolman, *Chem. Rev.* **1977**, *77*, 313–348.
[16] C. H. Suresh, N. Koga, *Inorg. Chem.* **2002**, *41*, 1573–1578.
[17] T. Gensch, G. Dos Passos Gomes, P. Friederich, E. Peters, T. Gaudin, R. Pollice, K. Jorner, A. Nigam, M. Lindner-D'Addario, M. S. Sigman, A. Aspuru-Guzik, *J. Am. Chem. Soc.* **2022**, *144*, 1205–1217.
[18] J. M. Stevens, J. Li, E. M. Simmons, S. R. Wisniewski, S. DiSomma, K. J. Fraunhoffer, P. Geng, B. Hao, E. W. Jackson, *Organometallics* **2022**, *41*, 1847–1864.
[19] S. Bajo, G. Laidlaw, A. R. Kennedy, S. Sproules, D. J. Nelson, *Organometallics* **2017**, *36*, 1662–1672.
[20] M. Mohadjer Beromi, A. Nova, D. Balcells, A. M. Brasacchio, G. W. Brudvig, L. M. Guard, N. Hazari, D. J. Vinyard, *J. Am. Chem. Soc.* **2017**, *139*, 922–936.
[21] P.-A. Payard, L. A. Perego, I. Ciofini, L. Grimaud, *ACS Catal.* **2018**, *8*, 4812–4823.
[22] A. H. Christian, P. Müller, S. Monfette, *Organometallics* **2014**, *33*, 2134–2137.
[23] C. W. Coley, *Trends Chem.* **2021**, *3*, 133–145.
[24] N. Fey, *Chem. Cent. J.* **2015**, *9*, 38.
[25] N. Fey, A. G. Orpen, J. N. Harvey, *Coord. Chem. Rev.* **2009**, *253*, 704–722.
[26] K. P. F. R. S . *Philos. Mag.* **1901**, *2*, 559–572.
[27] L. McInnes, J. Healy, J. Melville, **2018**, arXiv preprint arXiv:1802.03426.
[28] T. Gensch, S. R. Smith, T. J. Colacot, Y. N. Timsina, G. Xu, B. W. Glasspoole, M. S. Sigman, *ACS Catal.* **2022**, *12*, 7773–7780.
[29] R. W. Kennard, L. A. Stone, *Technometrics* **1969**, *11*, 137–148.
[30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
[31] S. Zhao, T. Gensch, B. Murray, Z. L. Niemeyer, M. S. Sigman, M. R. Biscoe, *Science* **2018**, *362*, 670–674.
[32] J. P. Stambuli, S. R. Stauffer, K. H. Shaughnessy, J. F. Hartwig, *J. Am. Chem. Soc.* **2001**, *123*, 2677–2678.
[33] J. Zhang, A. Bellomo, A. D. Creamer, S. D. Dreher, P. J. Walsh, *J. Am. Chem. Soc.* **2012**, *134*, 13765–13772.
[34] A. Zapf, M. Beller, *Chemistry* **2001**, *7*, 2908–2915.

Manuscript received: April 18, 2024
Revised manuscript received: July 14, 2024
Version of record online: ■■, ■■

RESEARCH ARTICLE

By mapping Ni-catalyzed Suzuki-Miyaura coupling yields onto a plot of ligand buried volume and minimum electrostatic potential, we identified the area of monophosphine ligand space best suited for this reaction. Reported here is a generalizable workflow for identifying physicochemical descriptors that best classify the chemical space for a given reaction.



A. LeSueur, N. Tao, A. Doyle, M. Sigman**