Organopalladium Catalysis as a Proving Ground for Data-Rich Approaches to Reaction Development and Quantitative Predictions

Jingru Lu and David C. Leitch

Department of Chemistry, University of Victoria, 3800 Finnerty Rd., Victoria, BC V8P 5C2, Canada. *dcleitch@uvic.ca

KEYWORDS: palladium catalysis; data science; high-throughput experimentation; molecular descriptors; machine learning; multivariate analysis

ABSTRACT. With the advent of high-throughput methods for both computation and experimentation, data-rich approaches to discovering and understanding chemical reactions are becoming ever more central to catalysis research. Organopalladium catalysis is at the forefront of these new approaches, providing a rich proving ground for method development and validation. This critical Perspective discusses a number of recent case studies from academic and industrial laboratories that illustrate how to generate, analyze, and correlate large data sets for quantitative predictions of reactivity and selectivity. Both the power and potential pitfalls of these approaches are discussed, as are the opportunities for both practical predictions and fundamental mechanistic insights.

INTRODUCTION

Organopalladium catalysis, and in particular palladium-catalyzed coupling, continues to be indispensable for fragment coupling in organic synthesis. Its success as a catalytic method for forming all manner of carbon-element bonds is unparalleled, especially in pursuit of complex molecule targets. Pd-catalyzed reactions are (and continue to be) central to the synthesis of myriad natural products,[1–3] active pharmaceutical ingredients,[4–6] agrochemicals,[7] and organic materials.[8] While the data was collected nearly 10 years ago, Brown and Boström's analysis of reaction classes most often used in medicinal chemistry has 3 of the top 20 as Pd-catalyzed: Suzuki-Miyaura, Sonogashira, and Buchwald-Hartwig.[6] No other examples of homogeneous organometallic catalysis are in this set. Certainly, the plethora of excellent research into non-precious metal catalysis for cross-coupling and related transformations means that these systems are emerging as viable alternatives.[9–13] However, the sheer number and variety of documented examples of Pd catalysis in complex molecule synthesis attests to its enduring importance as a reliable and effective means to access new chemical matter.

As a direct result of its fairly unique combination of wide applicability, highly variable reaction conditions, and massive amounts of published data, organopalladium catalysis is also one of the most frequently studied reaction classes in emerging data-rich methods for reaction discovery, optimization, and prediction. A 2018 survey of practitioners reveals catalytic C–C and C–N coupling as 2 of the top 3 most frequently screened reaction types by high-throughput experimentation (HTE) groups within (or affiliated with) pharmaceutical R&D.[14] And a recent comprehensive review of HTE in organometallic chemistry and catalysis from 2006-2020 reveals cross-coupling (sum of all varieties) as the most frequently reported reaction type undergoing high-throughput screening.[15] This is a stark contrast to the prior 10 years, where HTE

for asymmetric hydrogenation and alkene polymerization were overwhelmingly more common.[16] Finally, due to its clear dominance as a preferred method in medicinal chemistry, the patent literature contains millions of unique examples of Pd-catalyzed transformations across a wide swath of chemical space.

Thus, organopalladium catalysis is an ideal proving ground for emerging data-rich approaches to solving chemical synthesis problems. The recent ultrafast growth of computing and data analysis power combined with significant advancements in laboratory automation, miniaturization, and rapid chemical analysis have paved new ways to study and understand chemical systems. In organic synthesis specifically, data-driven methods have made a significant impact on accelerating the process of reaction optimization[17–19] and expanding the applicability of multivariate quantitative structure-reactivity/selectivity prediction models for synthesis planning.[20–25]

Large, accurate, consistent, and comprehensive datasets from reliable sources are the foundation of any data-driven methodology, and often the determining factor for overall success. The data foundation for studies of organopalladium catalysis is built from a combination of three major sources (Figure 1):

1) Previously reported reaction conditions/outcomes obtained directly from the academic/patent literature, and/or from open-source/proprietary reaction databases;

2) Calculated reaction-based parameters, including structurally and/or mechanistically-relevant molecular descriptors and computed reaction energy barriers;

3) "Do-it-yourself" experimentally determined reaction data, ideally collected via automated and/or HTE approaches.

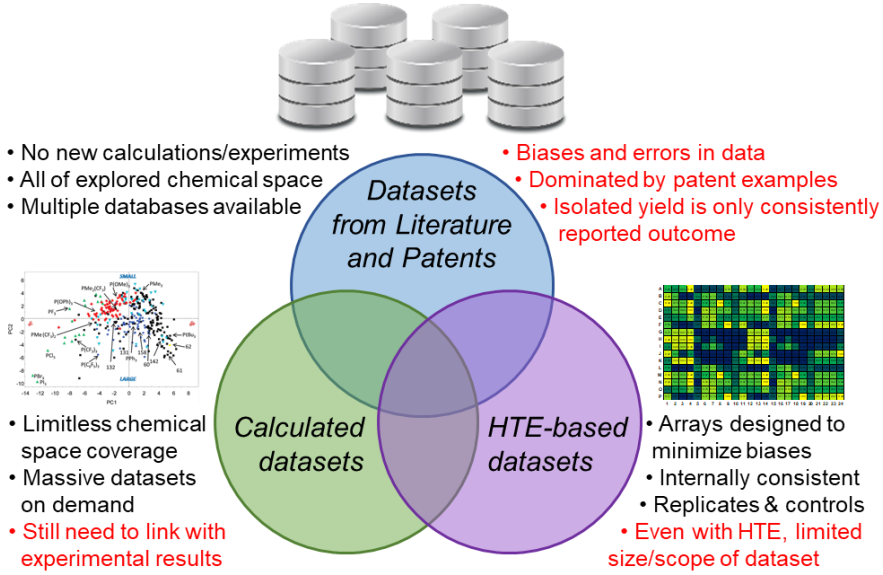

Figure 1. Data foundations for building quantitative models in catalyst reactivity and selectivity predictions.

This critical Perspective will cover key aspects of these three approaches to data collection and analysis, and the resulting outcomes. Illustrative examples from organopalladium catalysis show how these approaches are continuing to evolve. There are, of course, advantages and challenges specific to each approach; it is clear from the success stories thus far that a combination of approaches is the ideal strategy. In addition to accelerating catalyst and/or reaction discovery and optimization, taking a data-rich approach to studying chemical synthesis has the potential to significantly expand our understanding of the interplay between molecular structure, experimental conditions, and reaction mechanisms.

DATASETS FROM EXISTING SOURCES

The power (and pitfalls) of training on the literature. Due to its major role in pharmaceutical synthesis for both discovery and manufacturing activities, there is an abundance of reported examples for Pd-catalyzed cross-coupling across an extremely broad range of chemical space. The details of these reactions are spread throughout the academic and (especially) patent literature, and are readily accessible from major chemical databases and platforms. From a data science and statistical modeling perspective, this appears to be an ideal situation: a large, diverse, and available dataset containing "real-world" reaction systems to feed data-hungry but powerful machine learning algorithms.

Unfortunately, as noted by several experts in the field,[26–29] machine learning predictive models trained exclusively by literature/patent-derived data can suffer from several issues, including low prediction accuracy and lack of generalizability. Reaction outcome data from the literature, despite its abundance, is skewed by the objectives of each individual practitioner. A medicinal chemist making structural analogues during a lead-optimization campaign is not concerned with achieving comprehensive reaction condition coverage or collecting highly accurate yield/rate data; their job is to rapidly prepare target compounds by any means necessary. Likewise, an academic chemist developing new synthetic methods is not interested in reporting a large number of "failed" reactions; their job is to demonstrate the utility and scope of the new method. In other words, the experiments that produced the existing data were not designed or executed with an eye toward predictive statistical modeling, and thus great care must be taken before using these inputs. Glorius and coworkers have nicely addressed the common types of errors and biases in literature-based reaction datasets, which include experimental noise due to reporting errors, selection biases caused by researchers' preference for specific reagents over

others, and reporting biases toward successful results. A comparison of reported yield frequency between the *Reaxys* database and data obtained from HTE clearly reveals this latter point, with literature-derived data skewed heavily toward higher yields (illustrated in Figure 2 for Buchwald-Hartwig couplings).[27]

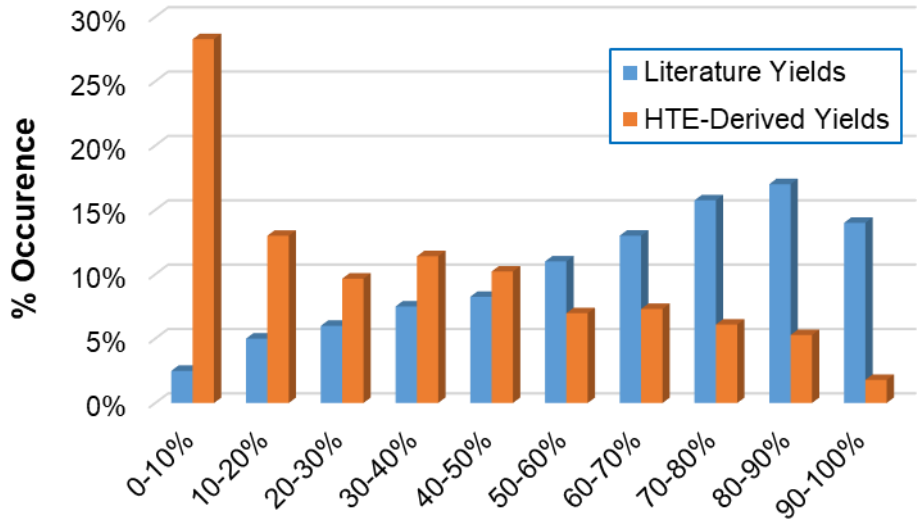

Figure 2. Comparison of frequency of reported yield values for Buchwald-Hartwig coupling reactions taken from the literature (*Reaxys* database) or from an HTE-based dataset. Values from ref. 27.

Selection bias is particularly problematic in synthetic chemistry data. This is because synthetic chemists have a high tendency to select specific reagents/catalysts/conditions based on familiarity, past success on similar systems, and ease of implementation. Such practices are very successful in pursuit of new molecules, but gives rise to a highly unbalanced distribution of reaction conditions in the reported datasets. Specifically for organopalladium catalysis, the

reported cross-coupling reaction conditions reveal a strong selection bias toward specific catalysts. Burke, Grzybowski and coworkers highlighted this issue when attempting to generate machine learning based models for Suzuki-Miyaura reactions.[28] They trained the models using >10,000 literature reported Suzuki couplings with an aim to predict conditions that would lead to success in a hypothetical reaction. Instead of making any chemically-meaningful predictions with respect to catalyst, solvent, or base choice, the models simply captured popularity trends[30,31] in the literature data. For example, >50% of reported Suzuki couplings use $Pd(PPh_3)_4$ as the palladium source, and therefore $Pd(PPh_3)_4$ was the top recommendation in >80% of the test cases.

Selection bias also extends to the specific target reactions being studied. Pd-catalyzed coupling reactions benefit from the large number of commercially available substrates: organohalides, boronic acids, amines, etc. However, even with myriad readily accessible and relatively complex substrates, reported chemical space coverage of the resulting products is actually rather limited. Krska, Dreher, and coworkers at Merck revealed a large disconnect in molecular properties between the reported products of Pd-catalyzed cross-couplings and many approved small molecule drugs.[32] The latter class of molecules generally possess more atoms, more H-bond acceptors, lower lipophilicities, and an overall higher degree of molecular complexity. Thus, prediction models trained exclusively on literature data may not be generalizable to late-stage coupling reactions on advanced pharmaceutical intermediates, where additional functional groups and molecular topologies can dramatically affect catalytic performance.

Reporting bias is another prevalent issue in the synthetic chemistry literature. Methodology studies frequently report univariate optimization within a relatively limited range

of conditions, and/or a series of "deviations from standard" conditions as control studies. Furthermore, the highest achievable yields of the target products are the key outcome reported, with little/no discussion of unsuccessful reactions (though this is changing for the better more recently). In addition, other information that is equally important to guide reaction prediction is infrequently disclosed. This includes aspects such as extent of reaction, rate of reaction, mass balance, and specific byproducts observed. For example, formation of homocoupling-derived byproducts is a common concern in Suzuki cross-couplings,[33] but the presence and/or quantity of these byproducts often goes unreported. This reporting bias results in literature-based datasets being skewed and/or incomplete, which is not ideal when applying them as training/test sets for statistical modeling.

Glorius and coworkers investigated the impact of insufficient low yield data on prediction accuracy for machine learning models.[27] In a case study on Buchwald-Hartwig reaction modeling, they established that the literature dataset was heavily skewed toward high yielding examples (>50% of data points with >70% yield, Figure 2); the resulting ML model performed relatively poorly, with a mean absolute error (MAE) of ±15%. If an HTE-generated dataset was instead used, which has a more realistic distribution of yields (>25% of data points with <10% yield), the resulting model was significantly more accurate (MAE = ±10%). Similarly, the authors simulated the effect of introducing additional experimental data to cover the low yielding reaction space, which again led to improved model performance.

Fitzner, Wuitschik, and coworkers have conducted large meta-analyses on more than 62,000 Buchwald-Hartwig coupling reactions reported in three databases (CAS, *Reaxys*, and USPTO).[29,34] Their aims were to identify and address the common issues in using literature data to guide reaction condition optimization, and to generate useful and broadly generalizable

predictions. They pointed out that many reactions are reported with very limited reaction information. For example, only 75% of the C–N couplings analyzed have reaction yield reported. In addition, essential reaction conditions such as temperature, reaction time, reagent scale and catalyst loading are significantly underreported. As already discussed in the context of Glorius's study, yields that are reported are skewed toward higher values in both the patent and non-patent literature (median yield >60% in recent years).

Importantly, this meta-analysis revealed reporting biases extend beyond (high) yield and (popular) catalyst/ligand to include solvent – nearly 80% of reactions use toluene or dioxane – and base – nearly 80% of reactions use NaO*t*Bu or Cs_2CO_3 (Figure 3). Reporting bias can also result from specific target applications becoming more prominent: the authors noted a large increase in use of P(*t*Bu)$_3$ and diarylamine nucleophiles beginning around 2014, due to an explosion of patent claims around OLED-relevant materials. Thus, while 62,000 reactions may seem like an ideal "big data" set, there are huge swaths of homogeneity in reaction conditions used, as well as overrepresented substrate/product classes based on commercial need. Thus, simply increasing the number of reactions in a dataset may not necessarily expand the chemical/reaction space coverage.

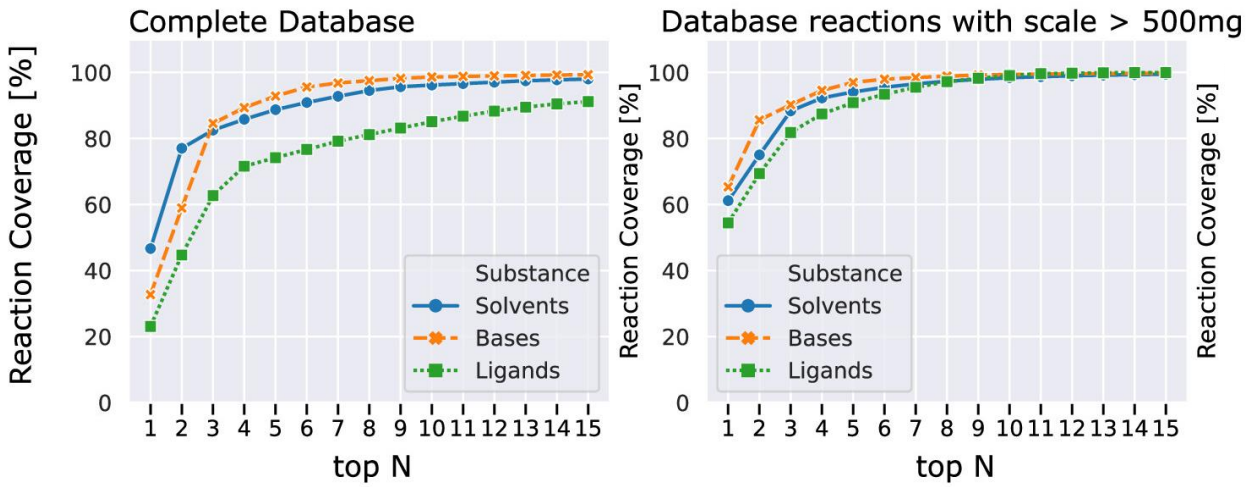

Figure 3. Diversity analysis of reported reaction conditions for C–N couplings from a combination of three databases (CAS, *Reaxys*, and USPTO). Plots represent % of reported conditions (*y*-axis) that contain the top N settings (*x*-axis). For example, ~80% of reactions in the complete database use just the top two solvents (top N = 2). Reproduced from Figure 4 in ref. 29, which is licensed under CC BY-NC-ND 4.0.

These data issues have an obvious impact on model accuracy and applicability. Fitzner and Wuitschik's initial meta-analysis did result in a set of qualitative predictive tools to aid condition selection for Buchwald-Hartwig couplings based on substrate type.[34] Subsequent work from this group on quantitative machine learning models trained on this large dataset revealed how such models can be misleading when applied to new synthetic cases beyond the initial training set.[29] Although their model performs well in training/test splits within the literature data, it fails to extend its predictive power to a new reaction dataset obtained by experiment (external predictions). They attribute this lack of generality to all of the aforementioned biases in the literature data. These hidden biases retained in the literature data make it impossible to account for all the important reaction parameters that must be considered and explored when discovering

or optimizing a new reaction. By presenting their failed attempt to achieve a general-purpose predictive model trained solely on literature data, they have identified a potentially general problem with such models: evaluating performance using test data taken from the overall literature data set may lead to overestimation of accuracy/generality. Thus, the authors suggest that it is important to use lab-generated reaction data (e.g. obtained by HTE) for reaction predictive model training and assessment.

Importantly, expert curation can be a powerful approach in leveraging literature data for predictive models. A study by Schleinitz *et al.* on predicting reaction yield highlights the importance of the quality of the data, rather than the quantity, in assembling a robust training dataset.[35] While this is not specifically organopalladium catalysis, the approach and lessons from this work are highly relevant. Among an initial dataset of >2000 Ni-catalyzed C–O couplings from the literature, the authors carefully selected a small subset of reactions (~200) that best represent the reaction information and chemical space covered by the original dataset. They then trained machine learning models using the large and curated datasets for comparison. The model trained with the small subset achieved the same predictive performance compared to the model trained on the original large dataset. Their study also reveals the importance of including failed experiments and low yield reactions extracted from optimization tables, such as those contained in Supporting Information sections. This will help to overcome reporting biases and lead to reaction datasets with a greater potential for generalization.

Perspectives on the Available Data. The above examples highlight the importance of knowing the limitations and biases of literature databases as a source of training data for predictive models. Using expert chemical knowledge in dataset assembly and curation prior to model training is critical to ensure accurate and useful predictions. The success of any statistical

modeling approach to reactivity predictions will rely on the quality and diversity of the training dataset. Even a relatively small dataset can be suitable if it covers a significant chemical/reaction space. The above examples also highlight how synthetic chemists can help improve the applicability of literature data toward predictive modeling. For any new method, evaluating a diverse and balanced chemical space when exploring reaction scope is enormously useful, such as with an informer library.[32] And evaluating/reporting diverse reaction conditions that result in a range of reaction yields is more useful than reporting only the "champion" conditions and yield. Recent initiatives such as the Open Reaction Database seek to standardize and streamline dataset curation to ensure each entry is complete and in machine-readable format, which should greatly improve the quality and accuracy of resulting models.[36]

One final perspective on existing literature datasets is whether the synthetic chemists' main metric (yield) is the best response variable for predictive modeling. Whether in academic papers or patent examples, the isolated yield is the universal measure of reactivity in organic synthesis. Accordingly, all of the aforementioned models are trained to build connections between molecular structure, reaction conditions, and yield values, with the key predicted outcome being the yield.[37] However, the reliability and consistency of reported yields data from different literature sources give rise to concerns for its use in predictive model training. Due to lack of consistency in our reporting practices, a literature yield could refer to several different measurements, such as solution or assay yield by NMR, GC, or HPLC analysis (common during optimization), or isolated yield after workup/purification (for reaction scope exploration). Importantly, isolated yields are not only a measure of reactivity, but the efficiency of the isolation procedure. Thus, there must be systematic errors in isolated yield values due to product loss during workup and purification.[38] This would be especially prevalent in cases where high

purity is desired over maximum yield – such as in the pharmaceutical syntheses that populate the USPTO database!

Another concern with reaction yield as the major response variable is the small range of values it can take. Yields are bounded between 0-100%, and nearly always reported to the nearest percent. The small product masses (10-100 mg) isolated from many reactions also mean that isolated yields may have significant measurement errors. The narrow reporting range and low precision of isolated yields makes it difficult to capture and model significant reactivity differences, especially in lower yielding reactions. For example, the difference between 85% and 95% yield and between 0% and 10% yield are both 10%; however, from a reactivity standpoint the former is barely significant, while the latter is very significant. Other measures of reaction outcome, such as rate or selectivity, can span multiple orders of magnitude and have distinct advantages in reactivity prediction; unfortunately, they are also much less frequently reported.

DATA-RICH COMPUTATIONAL APPROACHES

A crucial aspect of any quantitative structure-reactivity dataset is the collection of descriptors used to define molecular structures and reaction conditions. Thanks to tremendous developments in computer performance and theoretical methods, computational chemistry has made impressive progress in performing fast and reliable calculations for a wide variety of chemical systems. Especially germane to reactivity prediction, computational chemistry is extensively used to create molecular structural representations and corresponding numerical descriptors, and to compute reaction coordinates, transition states, and reaction barriers. It is a

powerful complement to experiments in validating mechanistic hypotheses and accelerating reaction optimization in synthetic chemistry.[39]

Whether by univariate regression or unsupervised machine learning, data-driven prediction models are developed by establishing quantitative relationships between experimentally obtained or computed molecular descriptors and measured reactivity outcome values. The application of powerful computational tools and mechanistic insights has vastly increased the number of relevant physicochemical descriptors to quantify molecular properties. Mechanistically relevant descriptors that directly impact chemical reactivity have found great value in reactivity prediction, such as the computed notations representing the electronic,[40,41] steric,[42–44] and vibrational[45] properties of a substrate or catalyst. Activation barriers determined from calculated transition state energies have also contributed to quantitatively accurate reaction outcome predictions.[46] Finally, computational methods combined with statistical algorithms are used to develop automated workflows for tasks ranging from training data selection to error assessment of the output numbers for improved efficiency and reliability in reaction predictions.[47,48]

Mapping Catalyst Chemical Space. A key aspect of organopalladium catalysis where computationally-derived datasets are prevalent is in ligand parameterization. Understanding and predicting the effect that a given ligand will have on reactivity and selectivity is central to Pd-catalyzed coupling. Ligand choice is often the most consequential factor for catalyst performance (both rate and productivity), especially with less reactive but more abundant substrates (e.g. unactivated Ar–Cl).[49–52] In addition, ligand identity is known to be a crucial factor to control site selectivity for Pd-catalyzed couplings of multihalogenated substrates.[53,54] Accordingly, ligand screening from large and diverse ligand libraries is a frequent activity in industrial HTE labs.

However, to adequately cover the available chemical space even within a subclass of supporting ligand is time and resource intensive. Computational tools have been invaluable in mapping this chemical space with large sets of relevant numerical descriptors.

Fey and coworkers have pioneered this approach through curation of the Ligand Knowledge Bases, a series of focused ligand descriptor datasets for specific ligand classes relevant to catalysis.[24,55] These sets include mono and bidentate ligands with P (Figure 4A), C (carbene), N, and O donor atoms, with more than 1,300 entries containing at least 20 descriptors each. This focused approach allows each ligand class to be described by a relevant set of descriptors; for example, descriptors for LKB-PP (bidentate P,P and P,N ligands) were adapted to account for multiple donor atoms and specific steric aspects of chelate bite angle.[56] Likewise, LKB-C (carbene ligands) was designed to take into account differences in electronic configuration between Schrock, Fischer, and NHC type carbene ligands.[57] To highlight the LKB's relevance to organopalladium catalysis, Fey and coworkers mapped HTE-derived Pd-catalyzed C–N coupling data against the principal component analysis (PCA) map of LKB-P (monodentate P).[58] While a quantitative model was not reported, the data visualization revealed a clear structure-reactivity trend where large, electron-rich phosphines are clearly optimal. In fact, adjacent to the most reactive ligands in the PCA map are several Buchwald-type biaryl ligands (not tested in the experimental data set) that are known to be very effective in this chemistry. Thus, LKB descriptors can potentially help to guide subsequent ligand screening.

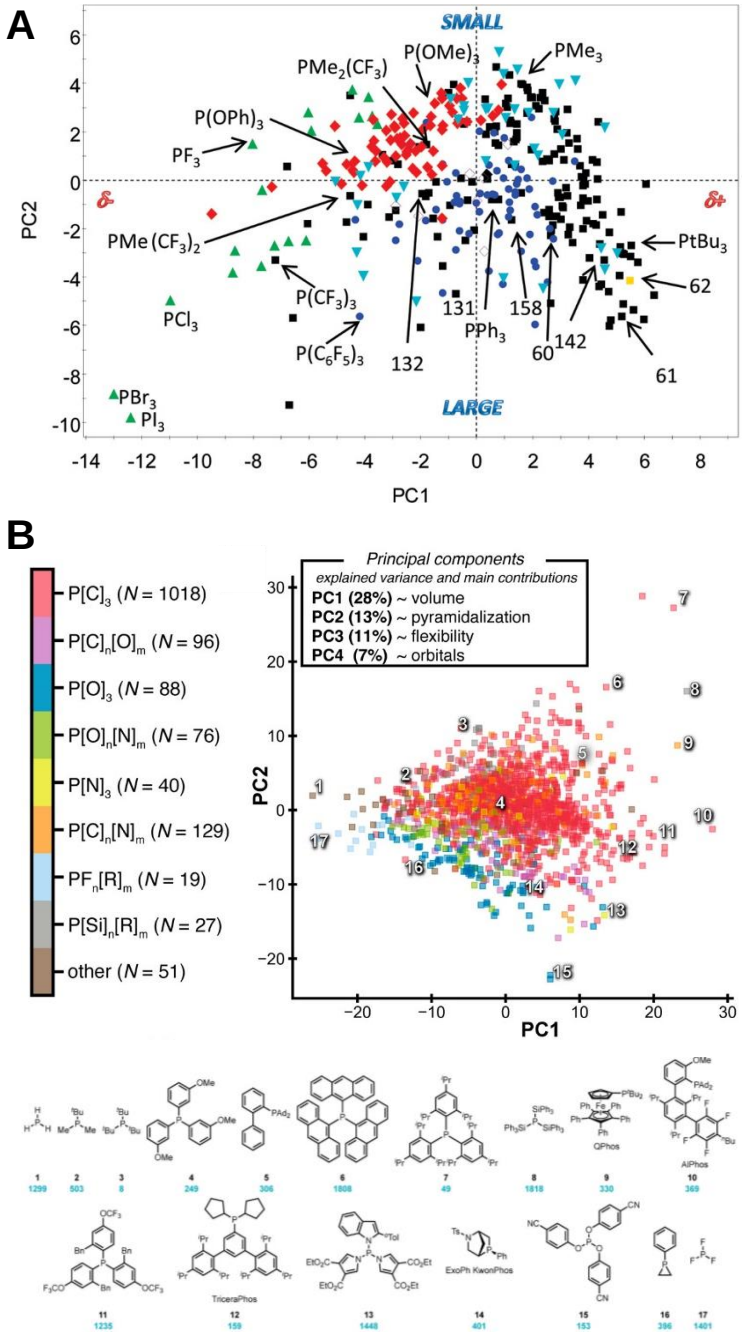

Figure 4. A comparison of two monodentate phosphine databases: A) LKB-P (reproduced with permission from ref. 58) and B) Kraken, with representative phosphines shown (reproduced with permission from ref. 60). Each database is represented by two-dimensional principal component analysis (PCA) plot. In the LKB-P, PC1 is predominantly electronic in nature, while PC2 is predominantly steric. In the Kraken visualization, PC1-PC4 have the indicated contributors.

Schoenebeck and coworkers recently reported using the LKB descriptors as a basis for ligand clustering analysis.[59] Their group's interest in dimeric Pd(I) species as precatalysts led them to investigate the ligand properties that stabilize these dimers versus those that fail to generate stable dimers. Based on the initial LKB PCA map, there is no clear trend, with successful (e.g. P(*t*-Bu)$_3$) and unsuccessful (e.g. PCy$_3$) phosphines residing in the same PCA space. Clustering analysis using *k*-means reveals subtle but key differences between the ligand types, which were reinforced using specific calculated descriptors related to Pd(I) dimer formation. This targeted ligand mapping led to identification of new ligands that support Pd(I) dimers, which were validated experimentally by isolating and characterizing eight previously unknown derivatives. This example shows how effective data analysis is just as important as efficient collection and collation of computational data: the initial PCA plot of LKB descriptors fails to capture the subtle but significant differences within the descriptor map, but alternative analyses (here, *via k*-means clustering) can reveal these "hidden" features.

More recently, Gensch, Sigman, Aspuru-Guzik and coworkers created an online descriptor database for monodentate phosphines (Figure 4B).[60] This database significantly expands the chemical and descriptor space applied to this important class of ligand, and incorporates conformational effects. For an initial set of 1,558 ligands (including the 200 most commonly cited phosphines in the literature), the authors calculated 190 individual descriptors using DFT methods. From this set, they then were able to train machine learning models to predict properties for a further 300,000 potential phosphine ligands. Thus, combining DFT and ML methods is a powerful approach to rapidly generate "big data" for molecular descriptor sets.

In a follow-up study, Gensch, Sigman, and coworkers applied this phosphine descriptor database to guide the development of a diverse monodentate phosphine screening set to

maximize chemical space coverage.[61] Unlike continuous reaction variables, such as temperature and reaction time, the electronic and steric properties of the PR_3 ligands are generally treated as categorical/discrete variables in experimental screening. In a typical HTE-based optimization approach, researchers will design an array using a combination of literature precedent and chemical intuition.[15] They will then determine which ligands to test next by analogy/perceived similarity to the most successful candidates in an iterative approach. Inevitably, this leads to selection biases when the search space for the discrete parameters is not well defined. To explore the discrete ligand space systematically, the authors sampled all of the commercially available PR_3 ligands from the virtual library and generated a four-dimensional principal component (PC) space from the descriptor set. Using a computational data-driven workflow, they then winnowed down to a diverse set of 32 monophosphines intended to maximally and uniformly cover the 4D PC space. Application of this screening set to several Pd-catalyzed C–C and C–N coupling reactions revealed the expected spread of catalytic results, with a clear positive hit (or several) observed in each case. Furthermore, given the logical design of the screening set, follow-up investigations can focus on ligands in close PC space proximity to "hit" ligands, reminiscent of the LKB-P example described previously.

High-Throughput Computation of Reaction Energy Profiles. In the absence of large experimental data sets of catalyst activity and/or selectivity for a given transformation (or set of transformations), computational studies can provide an invaluable source of data. Comparing relative energies of putative intermediates and/or transition states between different systems gives a molecular-level view of catalyst reactivity that is highly tunable. This is especially powerful where obtaining sufficient experimental data is cost and/or time prohibitive. It also

provides a method to interrogate aspects of the reaction mechanism that would be difficult or impossible using purely experimental means.

Corminboeuf and coworkers have successfully applied high-throughput computational tools to construct linear free energy scaling relationships for catalytic reactions, and represent these using "volcano plots".[62,63] Such plots are commonly used in heterogeneous catalysis and electrochemistry, and originate from Sabatier's principle that an ideal catalyst-reactant (or catalyst-product) interaction should be neither too strong nor too weak.[64,65] By extending this concept to homogeneous catalysis, Corminboeuf's group is able to gain new insights and make predictions of catalyst performance for transition metal-catalyzed cross-coupling reactions.

They first validated the concept of molecular volcano plots using Suzuki coupling reactions as a key case study (Figure 5).[66] For this initial analysis, catalytic behavior was modeled using thermodynamic aspects of the catalytic cycle: namely, the free energies of calculated catalytic intermediates for a variety of transition metals, and the linear relationship between the free energies of those intermediates. This analysis effectively reproduced the experimentally validated fact that Pd-based catalysts are optimal for Suzuki cross-coupling.

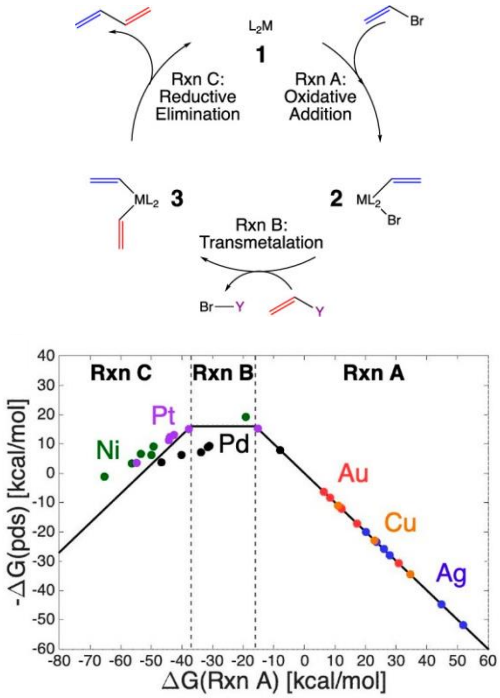

Figure 5. Molecular volcano plot of the calculated thermodynamics of oxidative addition (Rxn A), transmetallation (Rxn B) and reductive elimination (Rxn C) for a series of Ni, Pd, Pt, Cu, Ag, and Au catalysts Optimal performance is achieved by Pd-based catalysts, which have intermediate exergonicities of oxidative addition as compared to other transition metals Reproduced from ref. 62, which is licensed under CC BY-NC-ND 4.0.

In follow-up work, Corminboeuf and coworkers significantly expanded the catalyst pool under investigation to include more than 25,000 potential complexes.[67] Using DFT-obtained thermodynamic data on the free energy of oxidative addition for 7,000 of these as a training set, they were able to use machine learning analysis to predict those values for the remaining 18,000 potential catalysts. Constructing a molecular volcano plot using these thermodynamic parameters and filtering candidate catalysts by cost led to identification of 37 ideal candidates based on Pd or Cu. The authors also exploited this large dataset for further analysis through dimensionality

reduction *via* clustering analysis, extracting ligand effect trends across multiple cross-coupling reaction classes.[68] Finally, in addition to studying catalyst effects on reaction performance, they have also applied the volcano plot analysis to studying how substrate structures and substituents affect the outcome.[69]

The use of calculated transition state energies is another powerful indicator of catalyst performance, though collecting sufficient data can be challenging. A recent example by Yu, Fu, and coworkers showcases this approach to study the activating effect of Brønsted acids on the Pd-catalyzed C–O bond cleavage of allyl alcohols.[70] They created a computationally derived reaction database containing 393 DFT calculated activation barriers, followed by multivariate linear regression (MLR) model training using calculated molecular descriptors. The prediction accuracy was confirmed by a small experimental test set (taken from the literature[71]), with good agreement between predicted $\Delta G^{\ddagger}$ and the experimental product yield.

Despite gains in computing power, generating large data sets by traditional computational methods can be time and resource intensive. This is especially true when assessing a large pool of candidate ligands, catalyst species, and/or substrates at a high level of theory. To streamline computational approaches to studying large numbers of transition metal catalysis, Maeda and coworkers have presented a virtual ligand-assisted screening strategy to overcome the typical difficulties in transition state (TS) calculations for complex catalyst structures.[72] A virtual ligand is designed to approximate the structural features of a genuine ligand. This means that a virtual ligand reproduces the electronic and steric effects of the real system, while being simpler and faster to calculate. The initial case study from Maeda and coworkers successfully demonstrated how this virtual ligand approach can improve the efficiency of TS structure searches, and parameter-based optimization for simple monodentate phosphines (PR_3). The authors point out

that the future direction of virtual ligand screening development is to extend its scope to more complex ligand structures, which requires invention of new computational methods to approximate several other important ligand properties.

Currently, accurate *and* fast mechanistic modeling of organometallic transformations remains challenging. Schoenebeck has highlighted that computational studies of large, complex systems with myriad mechanistic possibilities lead to exponentially growing search directions (and thus exponentially growing computational resources).[39] An additional challenge is that calculation outcomes often depend on the theoretical method employed,[73] leading to another layer of complexity in designing high-throughput computational studies. Therefore, amassing computational data for a large array of Pd-catalyzed reactions needs to be performed with caution, balancing both calculation speed with accuracy.

HIGH-THROUGHPUT EXPERIMENTATION

Big data on demand *via* miniaturized array-based experiments. High throughput experimentation (HTE) is an increasingly common approach in both academic and industrial labs. HTE is especially well-suited to studying catalytic reactions, where the number of factors affecting reaction outcome are generally greater than in non-catalytic systems. While HTE is often used with the aim of either discovery (does this transformation work?) or optimization (what is the best set of conditions to maximize yield/selectivity?), it is also an invaluable approach to generating large internally-consistent datasets on demand. This Perspective will focus only on recent, illustrative applications of HTE in organopalladium chemistry; readers interested in HTE more generally are directed to several recent reviews on the topic.[14,15,74–80]

As HTE enables a data-rich approach to the discovery of chemical reactions – many more factors and their interactions can be studied simultaneously – it is particularly well-suited to the study of Pd-catalyzed processes. A typical cross-coupling reaction system includes numerous categorical factors (Pd source, ligand, base, solvent) and continuous factors (time, temperature, concentration, stoichiometries) that are impossible to fully assess "one factor at a time" (OFAT). Led by industrial R&D groups, particularly in pharmaceutical discovery and process chemistry, as well as academic HTE centers, data-rich HTE approaches to multifactor exploration and optimization have been demonstrated on many Pd-catalyzed reactions.[15] In addition to the direct applicability of these studies to active pharmaceutical ingredient (API) synthesis, organopalladium catalysis is again a general proving ground for data-rich techniques.

Developing new data-rich capabilities often goes hand-in-hand with applications in organopalladium chemistry. One set of examples is the development of new methods for high-throughput reaction set-up and analysis for dense reaction arrays. Researchers at Merck used a combination of advanced liquid handling automation with rapid UPLC-MS or MALDI-TOF MS analysis to interrogate thousands of Pd-catalyzed C–N coupling reactions in plate-based formats.[81,82] Researchers at Pfizer similarly demonstrated a new flow-based HTE platform using Pd-catalyzed Suzuki-Miyaura couplings to run and analyze thousands of examples.[83] In academia, the Cernak group is continuing to push the envelope of ultra-high throughput methods in array design and execution.[84–86] In the area of laboratory automation, the Hein group, in collaboration with Merck and the Aspuru-Guzik and Sigman groups, demonstrated the power of data-dense experimentation in an autonomous optimization of a Pd-catalyzed Suzuki-Miyaura reaction.[18] The Newman group published a user-centered "how-to-HTE" guide on multiwell screening, using a Pd-catalyzed C–N coupling as the prototype reaction.[87] And finally, even

undergraduate teaching lab experiments on microscale HTE employ Pd-catalyzed reactions as the exemplars.[88]

Our research group focuses not only using HTE and data-rich techniques to develop new reactions,[89–93] but also to better understand how reactivity and mechanism change as a function of the reaction system.[41,94,95] To enable these efforts in the realm of organopalladium chemistry, we identified a need for Pd precursor compounds that are specifically suited to HTE studies. While many options exist for Pd(II) precursors, the only versatile Pd(0) source for *in situ* catalyst formation is Pd_2dba_3 and its crystalline solvates.[96–99] Its (mostly) desirable reactivity is offset by known issues with stability, quality, and solubility; the latter point is crucial for conducting HTE that relies on liquid dispensing. In 2021 we reported an easily prepared and air-stable Pd(0) precursor, ^{DMP}DAB−Pd−MAH, which was designed specifically for HTE applications (Figure 6).[100] Inspired by prior work from Cavell, Stufkens, and Vrieze,[101] as well as from Elsevier,[102,103] we chose a glyoxal-derived α-diimine (or diazabutadiene, DAB) supporting ligand with an electron-deficient alkene (maleic anhydride) to stabilize the Pd(0) center. While we initially investigated the known ^{tBu}DAB–Pd–MAH complex as a precatalyst for HTE, the challenges of working with ^{tBu}DAB (volatility and stench) combined with sluggish ligand substitution rates with bulky phosphines led us to explore *N,N'*-diaryl diimines. With 2,6-dimethylphenyl groups at nitrogen (Figure 6A), ^{DMP}DAB–Pd–MAH is easy to prepare and isolate on multigram scale (Figure 6B), is easy to analyze for purity by NMR spectroscopy, is indefinitely bench stable under air, is soluble and stable in a variety of common organic solvents, and – most importantly – rapidly and quantitatively undergoes ligand substitution with every phosphine tested to date Figure 6C). This includes the exceedingly large AdBippyPhos and AlPhos.[104]

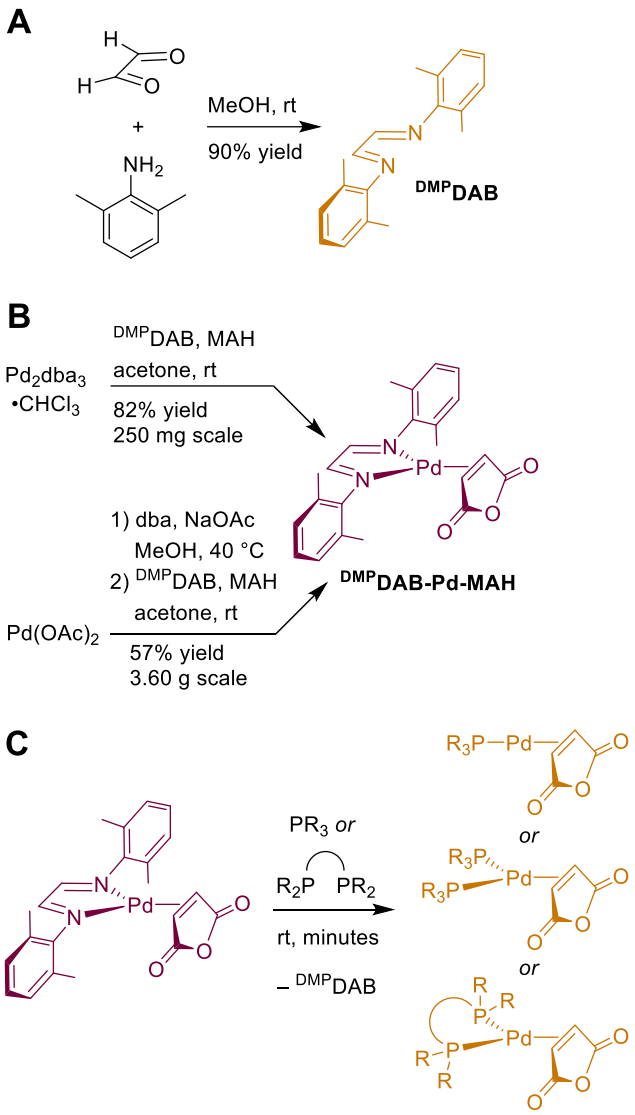

Figure 6. Synthesis of A) ^{DMP}DAB and B) ^{DMP}DAB–Pd–MAH. C) Ligand substitution of ^{DMP}DAB–Pd–MAH to rapidly generate mono or bis(phosphine) complexes, critical for *in situ* catalyst formation during HTE plate setup.

We have confirmed the efficacy of ^{DMP}DAB–Pd–MAH as a precursor for *in situ* catalyst formation during HTE for a number of Pd-catalyzed reactions. Our initial evaluations focused on cross-coupling chemistry, with C–N, C–C, and C–O bond formations among the reactions tested (Figure 7).[100] In subsequent work, we used ^{DMP}DAB–Pd–MAH to aid in reaction discovery for Pd-catalyzed tandem C–O/CH activation, where we found it to be the best Pd precursor for *in situ* catalyst formation.[92] These results indicate not only that ^{DMP}DAB–Pd–MAH is effective in catalysis, but also highlights a broader aspect of applying HTE to Pd-catalyzed reactions: the choice of Pd source is often critical to reaction success, even with all other factors held constant. There is, as yet, no universal Pd precursor that is effective for *in situ* catalyst formation regardless of reaction type or conditions used. This must be taken into consideration when designing and collecting training sets *via* HTE, especially where a variety of ligands and/or reaction conditions are being explored. Work is underway in our laboratory to generate additional Pd(0) and Pd(II) precursors that will approach the ideal of a universal precursor for HTE.

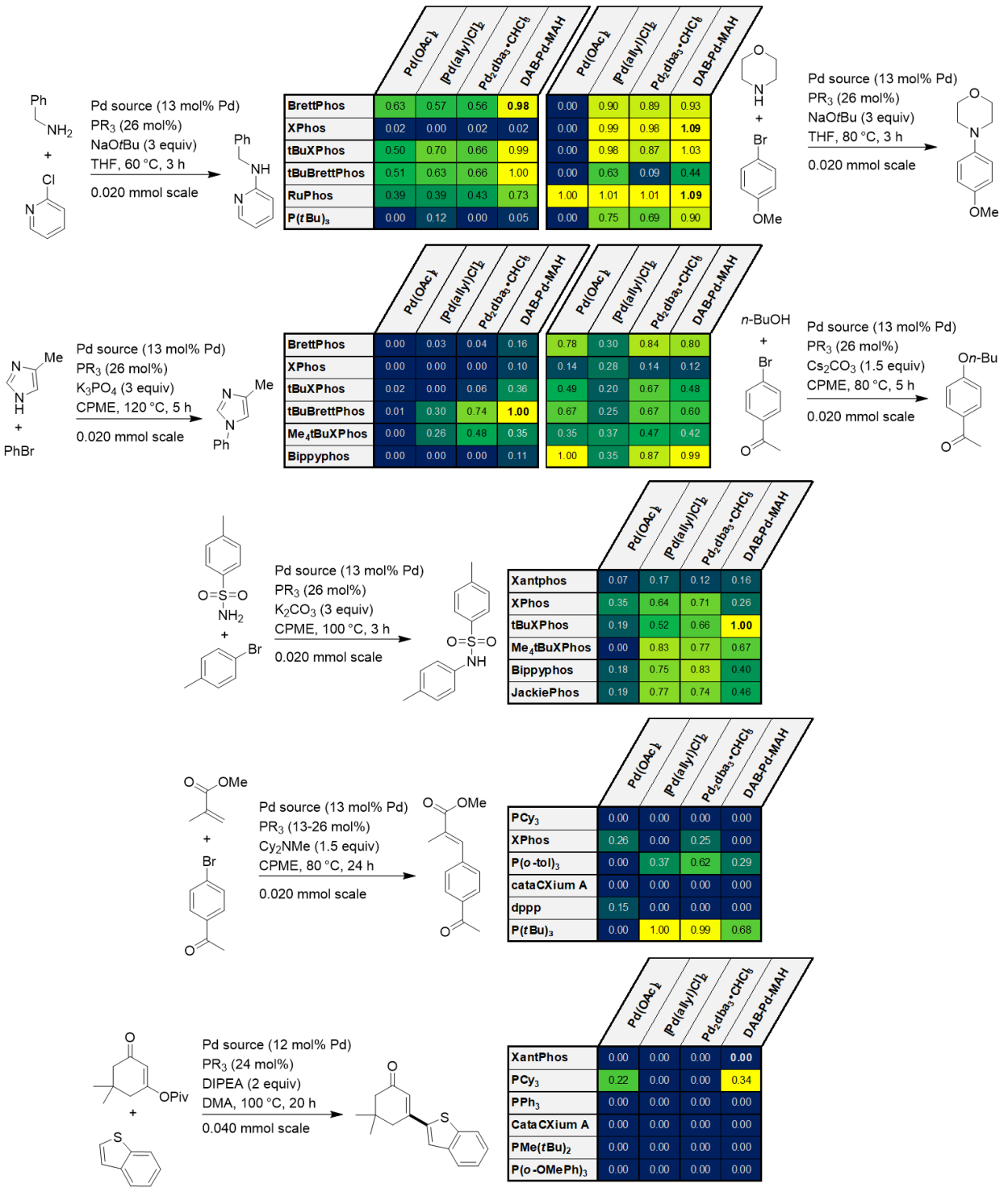

Figure 7. Comparison of ^{DMP}DAB–Pd–MAH (denoted DAB–Pd–MAH) with three other Pd sources in microscale HTE screening for cross-coupling reactions, including a newly developed C–O/C–H activation reaction for direct C–H alkenylation. Data from refs. 92 and 100.

While HTE has been widely used to conduct catalyst screening exercises, as well as multifactor optimization studies like those shown in Figure 7, the identities of the substrates themselves are increasingly being explored *via* HTE. Sather and Martinot used a data-rich approach to find the best reaction conditions for fusing piperidine-based nucleophiles with five-membered heteroaromatic bromides, a difficult C–N coupling type that has been rarely studied before.[105] Extensive HTE revealed Pd-PEPPSI IHept^{Cl} – an advanced precatalyst from the Organ group[106] – as the best catalyst in their model pyrazole-based system, as well as specific solvent/base combinations that prevent substrate decomposition. They subsequently tested the generality of this catalyst system in a panel of 48 unique heteroaryl bromides, with a "hit-rate" of ~50% (products observable by LCMS). This evaluation revealed several important features of the reaction, including incompatibility with certain ester functional groups, as well as protic functional groups including alcohols and amides.

In 2019 we, along with collaborators at GSK and Temple University, conducted a similar study of a challenging C–N coupling involving sulfonamide arylation.[89] Multi-substrate HTE revealed the BippyPhos class of ligands as superior, with AdBippyPhos as the optimal candidate. To evaluate the generality of this reaction, we tested an array of 288 substrate combinations: 24 heteroaryl halides and 12 primary and secondary sulfonamides. In this case, our "hit-rate" was ~25% (where coupled products were observed by LCMS in >10% area). Several substrates were ineffective when paired with any coupling partner, while others were effective with nearly all coupling partners. These kinds of substrate-focused HTE studies not only provide a map of reaction scope and highlight gaps in applicability, they are crucial as training sets for statistical analysis and reaction prediction.

Combining high-throughput experimental and computational approaches. HTE screening in chemistry is often aimed at achieving the ideal solution to a synthetic task. Beyond the optimization goal, the ability to generate hundreds-to-thousands of data points in a deliberate way means HTE is ideally suited as a means to build training datasets for reaction prediction. Multivariate HTE designs create reaction data libraries that cover a broader range of chemical structure and reaction space, and have more representative outcomes than literature data (*vide supra*, Figure 2). Such high-density coverage of reaction space is ideal when building quantitative statistical models for reactivity and/or selectivity, providing the experimental link to large computational datasets that are increasingly accessible (*vide supra*, Figure 4).

Catalytic reactions are known to be acutely sensitive to seemingly minor changes, where subtle alterations to catalyst structure, solvent identity, or even reaction concentration can have a dramatic impact on reaction outcomes. Understanding the underlying connections between reaction conditions, molecular structure, and chemical reactivity is both a fundamental goal of catalysis research, and is crucial to making accurate predictions. Because of the flexibility in experimental design and efficiency in data collection, HTE can be used to study how specific aspects of a reaction system impact reactivity, providing sufficient experimental data to train predictive models. Notably, the following three illustrative examples are all academic/industry collaborations.

A seminal example of combining high-throughput experimentation and computation for predictive modeling is the work of Dreher, Doyle, and coworkers on Pd-catalyzed C–N coupling.[107] Their study was focused on the effect of a specific type of inhibitory additive, isoxazole heterocycles, on the performance of various C–N couplings. Using Merck's nanoscale high-throughput platform, they collected a multidimensional experimental dataset of ~4000

reactions by combining 23 isoxazoles with different aryl halides, solvents and bases. From the computational side, a large set of molecular descriptors was generated without biasing selection with (possibly incorrect) mechanistic hypotheses. A variety of modeling approaches were evaluated, from multivariate linear regression (MLR) to machine learning (ML) algorithms, with a random forest algorithm providing the most accurate model. Even though ML-based models can be difficult to interpret mechanistically, the authors used a sensitivity analysis for key descriptors (i.e. how much is the model error increased by randomizing a given descriptor) to derive mechanistic insight. The two most sensitive descriptors are isoxazole-based, and are linked to electrophilicity (calculated ^{13}C chemical shift of the C_3 site, and LUMO energy), suggesting the most potent isoxazole inhibitors undergo competitive oxidative addition to Pd(0). This was verified experimentally by spectroscopic observation of the competitive N–O oxidative addition to $Pd(PPh_3)_4$. Notably, this work did generate commentary about the use of different feature sets,[108,109] whether chemical-based (as in Dreher and Doyle's approach[107]) or random-valued (as discussed by Chuang and Keiser[108]). While the chemical insights and out-of-sample prediction accuracy of the initial random forest model validate the use of chemical-based descriptors in this case,[109] incorporating control procedures and best practices into ML data analysis is crucial for those looking to use this powerful technique across the chemical sciences.[26,110]

Making predictions for multiple outcomes – such as high activity *and* selectivity – poses additional challenges for reaction modelling. This is particularly the case for regio- and/or stereoselective catalysis, where both high chemical yield and high selectivities are requirements for a successful process; thus, catalyst/ligand optimization requires a multi-objective approach. Mack, Sigman, and coworkers reported a data-driven workflow using high-throughput

computation and experimentation for multi-objective ligand optimization. This approach was demonstrated for two key enantioselective steps toward an investigational API from Genentech.[111] These steps are a regio- and enantioselective Pd-catalyzed Hayashi-Heck coupling, following by regio- and enantioselective Rh-catalyzed hydroformylation; the discussion here will focus on the first step (Figure 8A). In this case, the data foundations for this study are a computational database consisting of >550 bisphosphine ligands with DFT derived descriptors, and an experimental HTE screening database using a group of selected ligands based on previous research findings. These then informed the multi-objective analysis (Figure 8B). The ligands screened were classified based on a reactivity threshold analysis, where the phosphorus lone pair occupancy was an effective single descriptor. This revealed that phosphines below a threshold value were generally active, enabling a more focused ligand set to be taken forward. Then, regioselectivity as a function of ligand structure was assessed by MLR, revealing only two descriptors – anisotropic ^{31}P NMR shielding value and P–C σ* orbital occupancy – were sufficient to build a robust correlation. Finally, a virtual screen of the entire bisphosphine database identified ligands predicted to be even more regioselective than those in the HTE dataset, leading to experimental validation of (*S*)-HexaMeO-BIPHEP as optimal from a combined activity, regioselectivity, and enantioselectivity standpoint (Figure 8C).

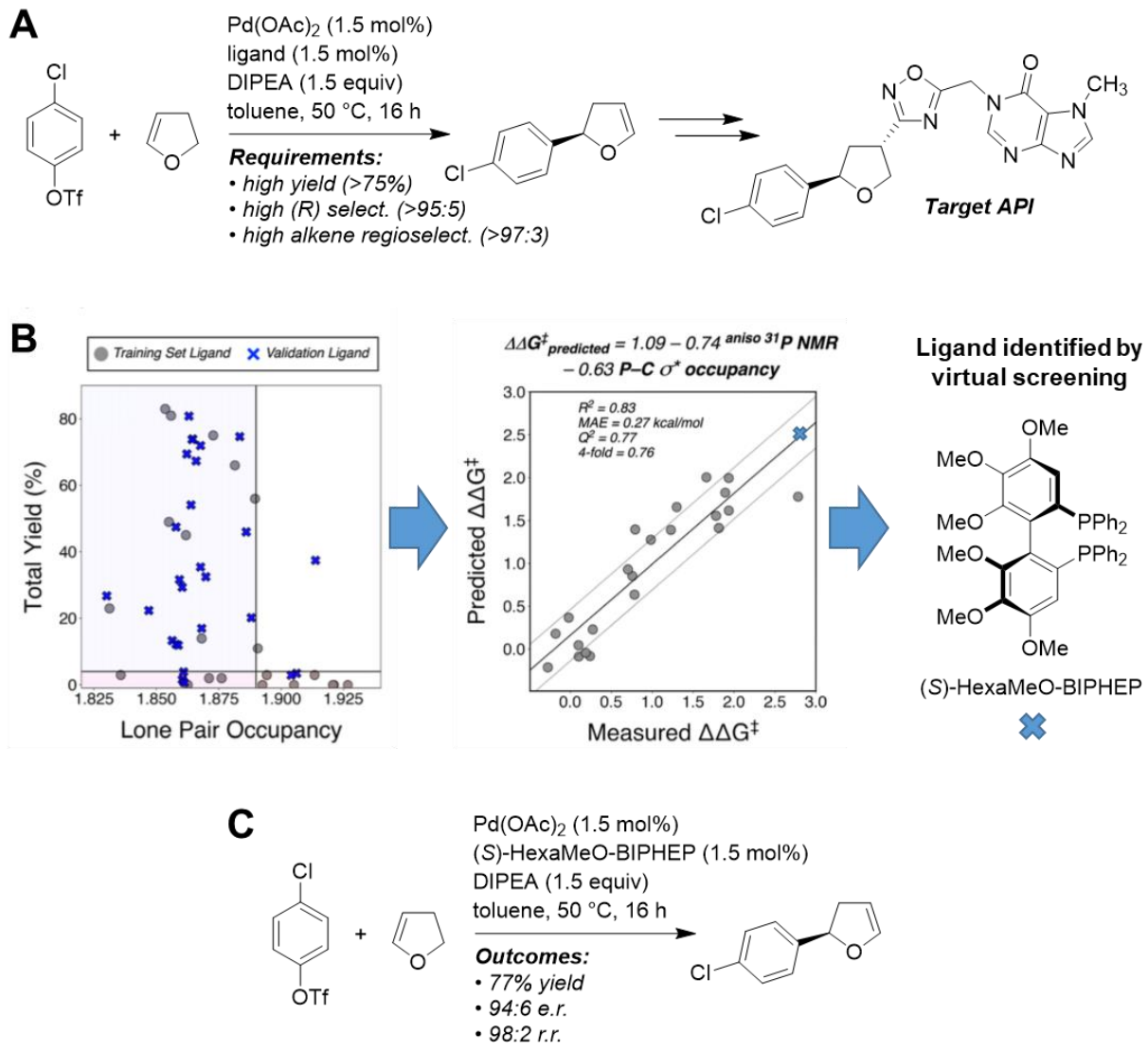

Figure 8. A) Hayashi-Heck coupling required in high yield, enantioselectivity, and regioselectivity. B) Step-wise multi-objective optimization *via* a data-rich computational / HTE approach, with univariate threshold analysis for activity, followed by MLR analysis for regioselectivity, and finally virtual screening for extrapolation to superior catalyst. C) Improved outcome with new catalyst system. Plots reproduced with permission from ref. 111.

In another example from the Sigman/Genentech collaboration, Xu *et al.* reported the use of MLR modeling to optimize an atroposelective Pd-catalyzed Negishi reaction using chiral

bisphosphine ligands. This step is key for the synthesis of a potent KRAS G12C covalent inhibitor, GDC-6036.[112] Using Genentech's HTE capabilities, the authors rapidly collected a focused set of data for a variety of chiral ligands, with only 3 out of 24 giving yields >20%, and none with e.r. >70:30. One Walphos-type ligand was identified as superior in this initial set, leading to a more focused HTE campaign to generate a Walphos training set for MLR analysis. The resulting linear model facilitated virtual ligand screening and experimental verification, akin to that described above, leading to high stereoselectivity for the desired Negishi reaction using W057-2 as the optimal ligand.

Notably in this case, the excellent selectivity exhibited by W057-2 is not transferable to other coupling partners in model Negishi reactions. Even small structural changes to the substrates result in significantly diminished yield and selectivity. This lack of generality is likely due to the focused HTE-based training set, where substrate variations were not taken into account. This is obviously not an issue for the Genentech researchers, who are primarily concerned with synthesis of GDC-6036 rather than simpler biaryls; however, it does reveal the broader challenges and complexities in discovering general catalytic systems, where substantially more data is required to account for interactions between substrate and catalyst.

Quantitative structure-rate relationships (QSRRs). Overall, combining HTE with MLR/ML analysis to link experimental outcomes to calculated descriptors is an excellent approach to building robust, accurate, and potentially generalizable predictive models. However, as discussed previously, using reaction yields as the outcome variable is not optimal for quantitative modeling. Kinetic parameters (rates, rate constants, $\Delta G^{\ddagger}$ values) are superior to yields as an outcome variable, especially for mechanism-based approaches. Many of the best-performing models – such as the two Sigman/Genentech examples above – use selectivity

measurements as the outcome metric. If under kinetic control, selectivity represents a ratio of reaction rates for the formation of one product or the other, and therefore can be used to construct linear free energy relationships (LFERs). This is why selectivity values are often represented as $\Delta\Delta G^{\ddagger}$ – i.e. free energies – in the resulting models.

A classic example of LFERs in physical organic chemistry is the Hammett equation,[113] which has been widely used to study mechanistic aspects of organic and organometallic reactions.[114–116] A relevant example in organopalladium chemistry was reported by Maes, Jutand, and coworkers, who studied the oxidative addition of 2-halopyridines to $Pd(PPh_3)_4$ using Hammett analyses of experimentally measured rates.[117] The corresponding Hammett plots and DFT calculations of transition states reveals that the oxidative addition mechanism for 2-halopyridines changes depending on the identity of the halide. A polarized nucleophilic displacement mechanism was proposed for 2-chloro and 2-bromopyridines, whereas a classic 3-centered mechanism was proposed for 2-iodopyridines.

Kinetic values can outperform yields in providing chemically-meaningful predictions because the major objective that most models are trained to predict – whether explicitly or implicitly – is kinetic in nature. Higher reaction yields are often correlated to favorable reaction kinetics. However, unlike selectivity or yield measurements – which are easily performed in HTE-type formats – determining large numbers of kinetic parameters *via* rate analysis is much more challenging. Rate data is also not as widely reported in the literature, limiting the usefulness of reaction database mining. A recent successful example of using literature rate data to train an ML QSRR model was reported by Jorner, Brinck, Norrby, and Buttar (yet another academic/industry collaboration!).[46] While this study focused on S_NAr chemistry and not organopalladium chemistry, their approach is likely to be generally applicable provided

sufficient data is available. The authors trained an ML model using a combination of >400 experimental rate constants from the literature, automated DFT transition state modeling, and calculated molecular descriptors for each component of the reaction system. This model has excellent prediction accuracy across a wide range of reported S_NAr reactions, though the literature dataset is necessarily skewed toward certain substrate classes (e.g. fluoronitroarenes as electrophiles). Importantly, the hybrid DFT/ML approach is superior when fewer experimental datapoints are used, which is a likely scenario when using kinetic parameters. Finally, this model performed extremely well in site selectivity predictions for out-of-sample multihalogenated electrophiles, despite not being trained for that purpose.

Jorner *et al.*'s work highlights the power of using kinetic parameters in quantitative modeling; however, data collection for new systems is non-trivial. In a collaboration between Pfizer and the Sigman group, independently measured rate constants were used to build a QSRR model for amide coupling.[118] Rate constants for 44 individual amide couplings – chosen *via* PCA of the chemical space for ~5,100 carboxylic acids and ~3,500 primary amines – were determined from concentration versus time profiles. Even for a relatively modest set of rate constants, this is a substantial experimental burden; however, the authors explicitly point out the superiority of using rate constants as opposed to yields.

To efficiently assemble rate-based datasets as a foundation for predictive modeling, alternative techniques are required. The concept of "one-pot multi-substrate screening" was introduced by Kagan for its application in fast optimization of asymmetric catalysts for various enantioselective reactions.[119,120] In 2007-2008, Plenio and coworkers extended this concept to realize high-throughput kinetic analysis of Pd-catalyzed Sonogashira couplings.[121,122] The authors of these two studies were able to simultaneously collect initial rates for up to 25

substrates in a single reaction flask. This significantly reduces the number of measurements needed to generate large kinetic datasets. Their QSRR studies combined 29 aryl bromide substrates with 17 monophosphine ligands for a total of 410 individual reactions. This pioneering study demonstrates the power of creative experimental design in high-throughput experimentation, where more individual experiments is not necessarily required to amass sufficient data. While the quantitative analysis provided by the authors does not attempt to create a unified multivariate model, this dataset was invaluable to our group as an external test set for catalytic predictions (*vide infra)*.

Plenio and coworkers do note some limitations of the one-pot, multi-substrate kinetic analysis approach. One challenge is measuring initial rates for very fast reactions, due to the need for offline analysis by (in their case) GC. Another challenge specific to this method is to accurately measure rates for the slowest reactions in a multi-substrate set, as the effective catalyst concentration per substrate will increase once the faster reactions finish. Thus, care must be taken to validate the results from these multi-substrate studies. Nevertheless, this approach is likely applicable to many reaction types, and should prove valuable for high-throughput kinetic analysis.

Our group has also considered the challenges of collecting large kinetic parameter datasets, and have taken a different approach to reducing the experimental burden. Rather than collect rate data from multiple measurements of concentration over time, we opted to measure *relative* rates *via* competition experiments. This effectively converts rate measurements to selectivity measurements. It also enables data collection for extremely fast reactions, since a concentration-time profile is not required. Using this approach, we recently reported two QSRR models for accurate prediction of S_NAr[94] and Pd-catalyzed cross-coupling reactivity.[41] By

correlating calculated molecular descriptors to relative free energies of activation ($\Delta\Delta G^{\ddagger}$), we obtained MLR models that have excellent performance to predict rate and selectivity for many external datasets. A discussion of the Pd-based work is illustrative of this approach.

Rather than study a specific class of Pd-catalyzed reaction, we focused our QSRR study on a key step in the catalytic mechanism: oxidative addition. This fundamental transformation is common to myriad catalytic reactions, and is often turnover and/or selectivity determining in Pd-catalyzed transformations (Figure 9A). We therefore hypothesized that a QSRR model for Ar–X oxidative addition to Pd(0) would be applicable to many different Pd-catalyzed cross-coupling reactions. To collect the required rate data, we assembled a diverse set of (hetero)aryl halides with a variety of substitution patterns, and ensuring heterocyclic substrates were well-represented. Using a model Pd(0) complex – $Pd(PCy_3)_2$ – we then performed a series of competition experiments under *pseudo* first-order conditions by having two Ar–X electrophiles in excess but equal amount compete for oxidative addition (Figure 9B). The product ratio, and therefore relative rate, is easily measured by quantitative ^{31}P NMR spectroscopy, and control experiments confirmed that the product ratio is kinetically controlled. This technique enables rapid assembly of a reactivity scale for substrates that spans many orders of magnitude in rate (Figure 9C).

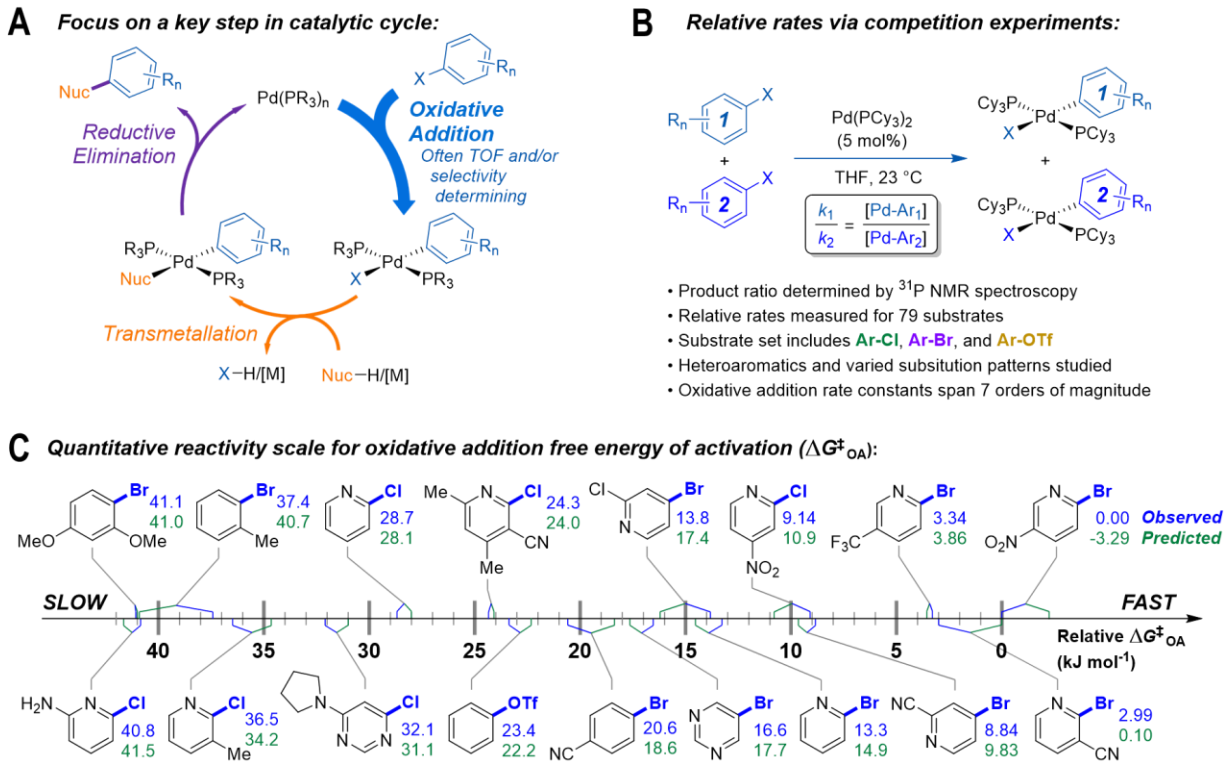

Figure 9. A) Generic catalytic cycle for Pd-catalyzed cross-coupling, highlighting the importance of oxidative addition. B) Experimental approach to rapid relative rate collection *via* competition experiments, with analysis of $L_2Pd(Ar)(X)$ complexes by ^{31}P qNMR spectroscopy. C) Reactivity scale for representative substrates. Figure adapted from ref. 41, which is licensed under CC BY-NC 3.0.

During feature selection for our QSRR models, we quickly determined that the average molecular electrostatic potential (*ESP*) is a particularly useful electronic descriptor. It is the average *ESP* energy across the surface area of an atom in the molecule,[123] and is easily calculated using the wavefunction analysis application Multiwfn.[124] Molecular *ESP* has been previously used to interpret the chemical properties and reactivity of various molecular systems.[125] Germane to this Perspective, Suresh and Koga[126] correlated minimum molecular *ESP*

values at P to a variety of phosphine molecular properties, demonstrating *ESP* is an excellent descriptor for the electronic nature of phosphine ligands. Subsequent work from Anjali and Suresh correlated molecular *ESP* at Pd to computationally-determined activation barriers for oxidative addition of various Ph–X substrates to L–Pd(0) fragments.[127]

In our oxidative addition QSRR model, five descriptors are sufficient to describe the reactivity for Ar–Cl, Ar–Br, and Ar–OTf substrates (Figure 10). The most significant contributors are the average molecular *ESP* at the reactive carbon (ESP_1) and at the adjacent atom (C or N, ESP_2); combined, these values contribute >60% to the model output. These local descriptors not only indicate the extent of electron-deficiency, but also the polarization of the C=C or C=N bond. Steric effects were initially accounted for using simple tabulated *A* values,[128] and work is underway to incorporate calculated steric descriptors.[43] Finally, the identity of the leaving group is linked to two additional descriptors: the C–X bond strength, given by the intrinsic bond strength index (*IBSI*),[129] and the pK_a of the leaving group conjugate acid. Both of these descriptors are necessary to unify the three classes of electrophile, and subsequent work indicates this set of features enables incorporating Ar–I substrates as a fourth substrate class.[104]

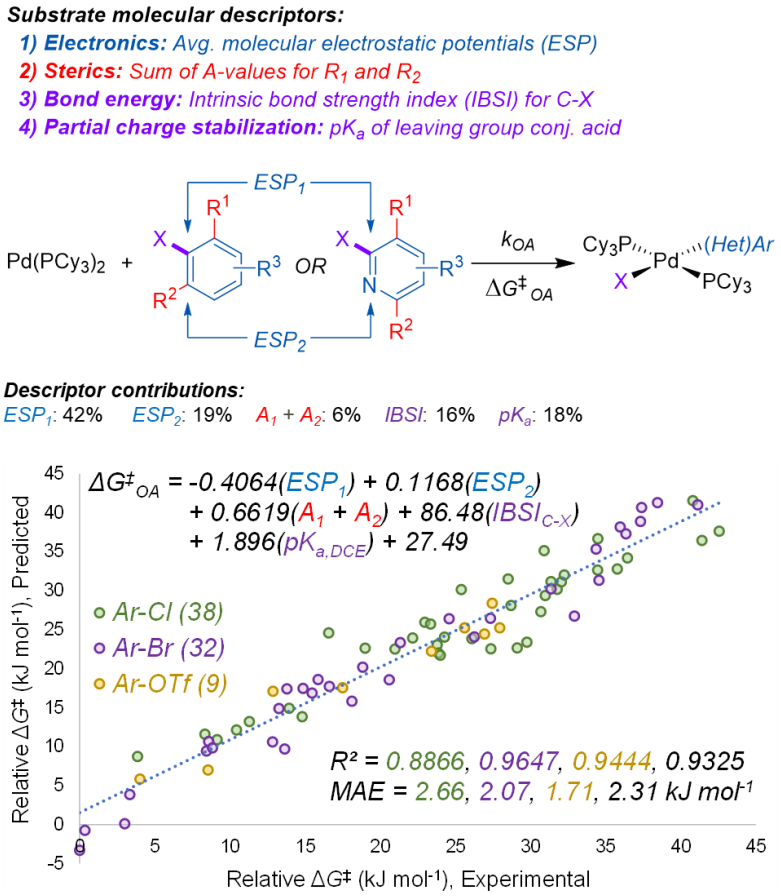

Figure 10. Multivariate QSRR model for Ar–X oxidative addition to $Pd(PCy_3)_2$. Figure adapted from ref. 41, which is licensed under CC BY-NC 3.0.

To test our hypothesis that a QSRR model for oxidative addition would be applicable to multiple classes of catalytic reactions, we used predicted $\Delta G^{\ddagger}_{OA}$ values from the model to make predictions about site selectivity in multihalogenated electrophiles. There is a plethora of literature available on cross-coupling site selectivity,[130] and our analysis indicated that the MLR model in Figure 10 is able to predict "conventional" site selectivity for a wide range of substrates in both Suzuki and Buchwald-Hartwig reactions. This is the expected site selectivity when using "simple" catalyst systems, such as those based on PPh_3, dppf, or other common phosphines.

Given we used $Pd(PCy_3)_2$ as the Pd(0) source to generate our model, this consistency with simple phosphines makes sense. Neufeldt's group have assessed factors that lead to unconventional site selectivity in Pd-catalyzed coupling, many of which are related to ligand and/or solvent identity, and their impacts on Pd speciation.[53,54,131–135] Future work in QSRR modeling for oxidative addition must therefore endeavor to capture these ligand/solvent/speciation effects.

Finally, we tested our oxidative addition QSRR model for predicting external catalytic rate data. The aforementioned Sonogashira dataset from Plenio and coworkers was an ideal and comprehensive case study, especially since the data could be cleanly separated into two distinct sets (Figure 11). In one paper, the authors focused on electronic effects (*meta* and *para* substitution), with 20 different Ar–Br substrates (Substrate set #1).[122] In another report, only steric effects were studied (*ortho* substitution), with 9 additional Ar–Br substrates (Substrate set #2, which also contains Ph–Br from Substrate set #1).[121] Importantly, most of the substrates from both sets are not included in our $\Delta G^{\ddagger}_{OA}$ training set. The inclusion of 17 different monophosphine ligands also provided an opportunity to test our model's applicability beyond PCy_3-ligated catalysts.

We exclusively used Substrate set #1 as our training and test set to build an initial MLR model. For each substrate, we calculated the predicted $\Delta G^{\ddagger}_{OA}$ value from the equation in Figure 10; effectively, this is a pre-weighted "super-descriptor" that takes into account the steric and electronic effects of the electrophile. For the phosphine ligands, we used two descriptors – average molecular *ESP* at P (ESP_P) and % buried volume ($\%V_{bur}$)[42] – to account for electronic and steric effects, respectively. These three descriptors – $\Delta G^{\ddagger}_{OA}$, ESP_P, and $\%V_{bur}$ – lead to an accurate linear model for ln *k* of the Sonogashira reactions in Substrate set #1, with statistics given in Figure 11.

To challenge this model, we reserved data from Substrate set #2, which contains all *ortho*-substituted substrates, as an external validation set. Even though the MLR model training data had none of these substrates included, predictions for Substrate set #2 are still excellent, with a mean absolute error of 0.732 (compared to 0.529-0.542 for the training/test sets). This is possible because steric effects are already accounted for in the predicted $\Delta G^{\ddagger}_{OA}$ values, since that model was trained using a far more diverse range of (Het)Ar–X substrates. In addition, the predicted $\Delta G^{\ddagger}_{OA}$ model appears to function equally well for all 17 phosphine ligands. Only two outlier points are observed, corresponding to 2,4,6-triisopropylphenyl bromide as a substrate with P(*t*Bu)$_3$ and P(Ad)$_2$(*t*Bu); in other words, the most sterically hindered substrate with the two largest phosphines. Individual univariate correlations of $\Delta G^{\ddagger}_{OA}$ with ln *k* for all 17 phosphines confirm this generality. Thus, even though our initial oxidative addition QSRR model was built using only Pd(PCy$_3$)$_2$, it is clearly able to make accurate predictions for many additional catalysts.

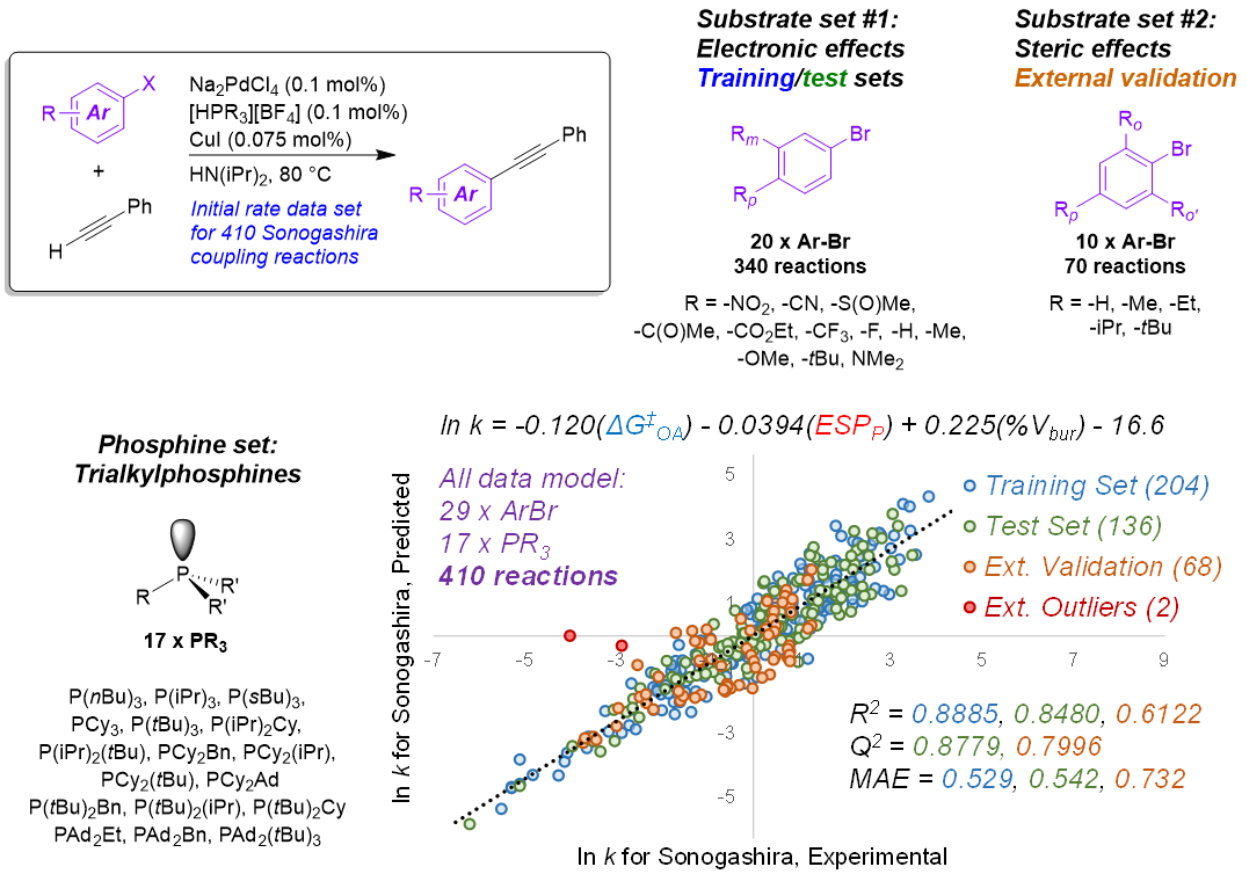

Figure 11. Applying predicted $\Delta G^{\ddagger}_{OA}$ values to external catalytic reaction datasets: unified QSRR of Plenio and coworkers' Sonogashira initial rates for all substrate/catalyst combinations. Figure adapted from ref. 41, which is licensed under CC BY-NC 3.0.

CONCLUSIONS AND OUTLOOK

The art and practice of catalyst and reaction development in synthetic chemistry is in a state of change. The advent of rapid tools to collect, visualize, and analyze the plethora of existing data from the academic and patent literatures is enabling chemists to use these data like never before. Increases in computing power and speed are greatly expanding the utility of high-throughput computational chemistry approaches to mapping chemical space with molecular descriptors, as well as rapid and accurate calculation of reaction intermediates and transition states for ever more complex molecules. And advances in laboratory automation and miniaturization are changing the way chemists design and conduct experiments, moving from traditional "one factor at a time" iteration to more holistic, multifactor approaches. Organopalladium catalysis is central to each of these three aspects, providing large existing datasets, rich catalyst structure space and mechanistic diversity, and a plethora of opportunities for high-throughput studies.

With these powerful new tools and approaches becoming more mainstream in catalysis research, we must be cognizant of the challenges and drawbacks of each individual approach. Literature/patent data is known to be skewed by selection and reporting biases; computational results are not acquired instantaneously, nor are they error-free; and HTE is still bottlenecked by reagent/catalyst availability, and by analysis speed/accuracy. The ideal data-rich approach to catalysis research must combine all three sources of data and the corresponding analysis methods. Importantly, massively large datasets are not required to make accurate and actionable reactivity/selectivity predictions, especially if mechanistic aspects are taken into account. As an exemplar of this fact, our work on quantitative predictions for Pd oxidative addition reactivity is

broadly applicable across multiple reaction classes and accurate when presented with new data, despite being built from only ~80 relative rate constants.

While this Perspective showed how organopalladium catalysis has been the proving ground of choice for many data-rich methods, there is nothing inherent about palladium or its reactivity that makes it uniquely suited for these new approaches. Many other reaction classes are well-represented in the chemical literature, and high-throughput experimentation methods are increasingly accessible to many reaction types.[85,136] The concepts from the case studies presented here, as well as others from the literature, are applicable to any type of synthetic chemistry. Finally, as remarked throughout, many of the seminal works on data-driven catalysis research are the result of academic/industry collaborations. These research partnerships will likely continue to be at the forefront of this emerging area.

AUTHOR INFORMATION

Corresponding Author

*dcleitch@uvic.ca

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

We acknowledge with respect the Lekwungen peoples on whose traditional territory the University of Victoria (UVic) stands, and the Songhees, Esquimalt and WSÁNEĆ peoples whose historical relationships with the land continue to this day. We also acknowledge funding from the New Frontiers in Research Fund – Exploration and the NSERC Discovery Grant programs. D.C.L. thanks the Research Corporation for Science Advancement for a Cottrell Scholar Award.

REFERENCES

1) Trost, B. M.; Crawley, M. L. Asymmetric Transition-Metal-Catalyzed Allylic Alkylations: Applications in Total Synthesis. *Chem. Rev.* **2003**, *103*, 2921–2944. https://doi.org/10.1021/cr020027w

2) Nicolaou, K. C.; Bulger, P. G.; Sarlah, D. Palladium-Catalyzed Cross-Coupling Reactions in Total Synthesis. *Angew. Chem. Int. Ed.* **2005**, *44*, 4442–4489. https://doi.org/10.1002/anie.200500368

3) Bai, Y.; Davis, D. C.; Dai, M. Natural Product Synthesis via Palladium-Catalyzed Carbonylation. *J. Org. Chem.* **2017**, *82*, 2319–2328. https://doi.org/10.1021/acs.joc.7b00009

4) Magano, J.; Dunetz, J. R. Large-Scale Applications of Transition Metal-Catalyzed Couplings for the Synthesis of Pharmaceuticals. *Chem. Rev.* **2011**, *111*, 2177–2250. https://doi.org/10.1021/cr100346g

5) Ruiz-Castillo, P.; Buchwald, S. L. Applications of Palladium-Catalyzed C–N Cross-Coupling Reactions. *Chem. Rev.* **2016**, *116*, 12564–12649. https://doi.org/10.1021/acs.chemrev.6b00512

6) Brown, D. G.; Boström, J. Analysis of Past and Present Synthetic Methodologies on Medicinal Chemistry: Where Have All the New Reactions Gone? *J. Med. Chem.* **2016**, *59*, 4443–4458. https://doi.org/10.1021/acs.jmedchem.5b01409

7) Devendar, P.; Qu, R.-Y.; Kang, W.-M.; He, B.; Yang, G.-F. Palladium-Catalyzed Cross-Coupling Reactions: A Powerful Tool for the Synthesis of Agrochemicals. *J. Agric. Food Chem.* **2018**, *66*, 8914–8934. https://doi.org/10.1021/acs.jafc.8b03792

8) Baker, M. A.; Tsai, C.-H.; Noonan, K. J. T. Diversifying Cross-Coupling Strategies, Catalysts and Monomers for the Controlled Synthesis of Conjugated Polymers. *Chem. – Eur. J.* **2018**, *24*, 13078–13088. https://doi.org/10.1002/chem.201706102

9) Han, F.-S. Transition-Metal-Catalyzed Suzuki–Miyaura Cross-Coupling Reactions: A Remarkable Advance from Palladium to Nickel Catalysts. *Chem. Soc. Rev.* **2013**, *42*, 5270–5298. https://doi.org/10.1039/C3CS35521G

10) Tasker, S. Z.; Standley, E. A.; Jamison, T. F. Recent Advances in Homogeneous Nickel Catalysis. *Nature* **2014**, *509*, 299–309. https://doi.org/10.1038/nature13274

11) Ananikov, V. P. Nickel: The "Spirited Horse" of Transition Metal Catalysis. *ACS Catal.* **2015**, *5*, 1964–1971. https://doi.org/10.1021/acscatal.5b00072

12) Coombs, J. R.; Green, R. A.; Roberts, F.; Simmons, E. M.; Stevens, J. M.; Wisniewski, S. R. Advances in Base-Metal Catalysis: Development of a Screening Platform for Nickel-Catalyzed Borylations of Aryl (Pseudo)Halides with $B_2(OH)_4$. *Organometallics* **2019**, *38*, 157–166. https://doi.org/10.1021/acs.organomet.8b00307

13) Chirik, P. J.; Engle, K. M.; Simmons, E. M.; Wisniewski, S. R. Collaboration as a Key to Advance Capabilities for Earth-Abundant Metal Catalysis. *Org. Process Res. Dev.* **2023**, *27*, 1160-1184. https://doi.org/10.1021/acs.oprd.3c00025

14) Mennen, S. M.; Alhambra, C.; Allen, C. L.; Barberis, M.; Berritt, S.; Brandt, T. A.; Campbell, A. D.; Castañón, J.; Cherney, A. H.; Christensen, M.; Damon, D. B.; Eugenio de Diego, J.; García-Cerrada, S.; García-Losada, P.; Haro, R.; Janey, J.; Leitch, D. C.; Li, L.; Liu, F.; Lobben, P. C.; MacMillan, D. W. C.; Magano, J.; McInturff, E.; Monfette, S.; Post, R. J.; Schultz, D.; Sitter, B. J.; Stevens, J. M.; Strambeanu, I. I.; Twilton, J.; Wang, K.; Zajac, M. A. The Evolution of High-Throughput Experimentation in Pharmaceutical Development and Perspectives on the Future. *Org. Process Res. Dev.* **2019**, *23*, 1213–1242. https://doi.org/10.1021/acs.oprd.9b00140

15) Leitch, D. C.; Becica, J. High-Throughput Experimentation in Organometallic Chemistry and Catalysis. In *Comprehensive Organometallic Chemistry IV*; Parkin, G., Meyer, K., O'Hare, D., Eds.; Elsevier: Oxford, 2022; pp 502–555. https://doi.org/10.1016/B978-0-12-820206-7.00111-6

16) Murphy, V. High-Throughput Organometallic Chemistry: Chemical Approaches, Experimental Methods, and Screening Techniques. In *Comprehensive Organometallic Chemistry III*; Mingos, D. M. P., Crabtree, R. H., Eds.; Elsevier: Oxford, 2007; pp 341–379. https://doi.org/10.1016/B0-08-045047-4/00014-5

17) Clayton, A. D.; Manson, J. A.; Taylor, C. J.; Chamberlain, T. W.; Taylor, B. A.; Clemens, G.; Bourne, R. A. Algorithms for the Self-Optimisation of Chemical Reactions. *React. Chem. Eng.* **2019**, *4*, 1545–1554. https://doi.org/10.1039/C9RE00209J

18) Christensen, M.; Yunker, L. P. E.; Adedeji, F.; Häse, F.; Roch, L. M.; Gensch, T.; dos Passos Gomes, G.; Zepel, T.; Sigman, M. S.; Aspuru-Guzik, A.; Hein, J. E. Data-Science Driven Autonomous Process Optimization. *Commun. Chem.* **2021**, *4*, 112. https://doi.org/10.1038/s42004-021-00550-x

19) Taylor, C. J.; Pomberger, A.; Felton, K. C.; Grainger, R.; Barecka, M.; Chamberlain, T. W.; Bourne, R. A.; Johnson, C. N.; Lapkin, A. A. A Brief Introduction to Chemical Reaction Optimization. *Chem. Rev.* **2023**, *123*, 3089–3126. https://doi.org/10.1021/acs.chemrev.2c00798

20) Sigman, M. S.; Harper, K. C.; Bess, E. N.; Milo, A. The Development of Multidimensional Analysis Tools for Asymmetric Catalysis and Beyond. *Acc. Chem. Res.* **2016**, *49*, 1292–1301. https://doi.org/10.1021/acs.accounts.6b00194

21) Santiago, C. B.; Guo, J.-Y.; Sigman, M. S. Predictive and Mechanistic Multivariate Linear Regression Models for Reaction Development. *Chem. Sci.* **2018**, *9*, 2398–2412. https://doi.org/10.1039/C7SC04679K

22) Muratov, E. N.; Bajorath, J.; Sheridan, R. P.; Tetko, I. V.; Filimonov, D.; Poroikov, V.; Oprea, T. I.; Baskin, I. I.; Varnek, A.; Roitberg, A.; Isayev, O.; Curtalolo, S.; Fourches, D.; Cohen, Y.; Aspuru-Guzik, A.; Winkler, D. A.; Agrafiotis, D.; Cherkasov, A.; Tropsha, A. QSAR without Borders. *Chem. Soc. Rev.* **2020**, *49*, 3525–3564. https://doi.org/10.1039/D0CS00098A

23) Shen, Y.; Borowski, J. E.; Hardy, M. A.; Sarpong, R.; Doyle, A. G.; Cernak, T. Automation and Computer-Assisted Planning for Chemical Synthesis. *Nat. Rev. Methods Primer* **2021**, *1*, 23. https://doi.org/10.1038/s43586-021-00022-5

24) Durand, D. J.; Fey, N. Building a Toolbox for the Analysis and Prediction of Ligand and Catalyst Effects in Organometallic Catalysis. *Acc. Chem. Res.* **2021**, *54*, 837–848. https://doi.org/10.1021/acs.accounts.0c00807

25) Williams, W. L.; Zeng, L.; Gensch, T.; Sigman, M. S.; Doyle, A. G.; Anslyn, E. V. The Evolution of Data-Driven Modeling in Organic Chemistry. *ACS Cent. Sci.* **2021**, *7*, 1622–1637. https://doi.org/10.1021/acscentsci.1c00535

26) Artrith, N.; Butler, K. T.; Coudert, F.-X.; Han, S.; Isayev, O.; Jain, A.; Walsh, A. Best Practices in Machine Learning for Chemistry. *Nat. Chem.* **2021**, *13*, 505–508. https://doi.org/10.1038/s41557-021-00716-z

27) Strieth-Kalthoff, F.; Sandfort, F.; Kühnemund, M.; Schäfer, F. R.; Kuchen, H.; Glorius, F. Machine Learning for Chemical Reactivity: The Importance of Failed Experiments. *Angew. Chem. Int. Ed.* **2022**, *61*, e202204647. https://doi.org/10.1002/anie.202204647

28) Beker, W.; Roszak, R.; Wołos, A.; Angello, N. H.; Rathore, V.; Burke, M. D.; Grzybowski, B. A. Machine Learning May Sometimes Simply Capture Literature Popularity Trends: A Case Study of Heterocyclic Suzuki–Miyaura Coupling. *J. Am. Chem. Soc.* **2022**, *144*, 4819–4827. https://doi.org/10.1021/jacs.1c12005

29) Fitzner, M.; Wuitschik, G.; Koller, R.; Adam, J.-M.; Schindler, T. Machine Learning C–N Couplings: Obstacles for a General-Purpose Reaction Yield Prediction. *ACS Omega* **2023**, *8*, 3017–3025. https://doi.org/10.1021/acsomega.2c05546

30) Kowalczyk, B.; Bishop, K. J. M.; Smoukov, S. K.; Grzybowski, B. A. Synthetic Popularity Reflects Chemical Reactivity. *J. Phys. Org. Chem.* **2009**, *22*, 897–902. https://doi.org/10.1002/poc.1535

31) Soh, S.; Wei, Y.; Kowalczyk, B.; Gothard, C. M.; Baytekin, B.; Gothard, N.; Grzybowski, B. A. Estimating Chemical Reactivity and Cross-Influence from Collective Chemical Knowledge. *Chem. Sci.* **2012**, *3*, 1497–1502. https://doi.org/10.1039/C2SC00011C

32) 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. https://doi.org/10.1039/C5SC04751J

33) Moreno-Mañas, M.; Pérez, M.; Pleixats, R. Palladium-Catalyzed Suzuki-Type Self-Coupling of Arylboronic Acids. A Mechanistic Study. *J. Org. Chem.* **1996**, *61*, 2346–2351. https://doi.org/10.1021/jo9514329

34) Fitzner, M.; Wuitschik, G.; Koller, R. J.; Adam, J.-M.; Schindler, T.; Reymond, J.-L. What Can Reaction Databases Teach Us about Buchwald–Hartwig Cross-Couplings? *Chem. Sci.* **2020**, *11*, 13085–13093. https://doi.org/10.1039/D0SC04074F

35) Schleinitz, J.; Langevin, M.; Smail, Y.; Wehnert, B.; Grimaud, L.; Vuilleumier, R. Machine Learning Yield Prediction from NiCOlit, a Small-Size Literature Data Set of Nickel Catalyzed C–O Couplings. *J. Am. Chem. Soc.* **2022**, *144*, 14722–14730. https://doi.org/10.1021/jacs.2c05302

36) 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. *J. Am. Chem. Soc.* **2021**, *143*, 18820–18826. https://doi.org/10.1021/jacs.1c09820

37) Ertl, P.; Gerebtzoff, G.; Lewis, R.; Muenkler, H.; Schneider, N.; Sirockin, F.; Stiefl, N.; Tosco, P. Chemical Reactivity Prediction: Current Methods and Different Application Areas. *Mol. Inform.* **2022**, *41*, 2100277. https://doi.org/10.1002/minf.202100277

38) Wernerova, M.; Hudlicky, T. On the Practical Limits of Determining Isolated Product Yields and Ratios of Stereoisomers: Reflections, Analysis, and Redemption. *Synlett* **2010**, 2701–2707. https://doi.org/10.1055/s-0030-1259018

39) Sperger, T.; Sanhueza, I. A.; Schoenebeck, F. Computation and Experiment: A Powerful Combination to Understand and Predict Reactivities. *Acc. Chem. Res.* **2016**, *49*, 1311–1319. https://doi.org/10.1021/acs.accounts.6b00068

40) Gadre, S. R.; Suresh, C. H.; Mohan, N. Electrostatic Potential Topology for Probing Molecular Structure, Bonding and Reactivity. *Molecules* **2021**, *26*, 3289. https://doi.org/10.3390/molecules26113289

41) Lu, J.; Donnecke, S.; Paci, I.; Leitch, D. C. A Reactivity Model for Oxidative Addition to Palladium Enables Quantitative Predictions for Catalytic Cross-Coupling Reactions. *Chem. Sci.* **2022**, *13*, 3477–3488. https://doi.org/10.1039/D2SC00174H

42) Clavier, H.; Nolan, S. P. Percent Buried Volume for Phosphine and N-Heterocyclic Carbene Ligands: Steric Properties in Organometallic Chemistry. *Chem. Commun.* **2010**, *46*, 841–861. https://doi.org/10.1039/B922984A

43) Brethomé, A. V.; Fletcher, S. P.; Paton, R. S. Conformational Effects on Physical-Organic Descriptors: The Case of Sterimol Steric Parameters. *ACS Catal.* **2019**, *9*, 2313–2323. https://doi.org/10.1021/acscatal.8b04043

44) McFord, A. W.; Butts, C. P.; Fey, N.; Alder, R. W. 3× Axial vs 3× Equatorial: The ΔG_{GA} Value Is a Robust Computational Measure of Substituent Steric Effects. *J. Am. Chem. Soc.* **2021**, *143*, 13573–13578. https://doi.org/10.1021/jacs.1c04247

45) Milo, A.; Bess, E. N.; Sigman, M. S. Interrogating Selectivity in Catalysis Using Molecular Vibrations. *Nature* **2014**, *507*, 210–214. https://doi.org/10.1038/nature13019

46) Jorner, K.; Brinck, T.; Norrby, P.-O.; Buttar, D. Machine Learning Meets Mechanistic Modelling for Accurate Prediction of Experimental Activation Energies. *Chem. Sci.* **2021**, *12*, 1163–1175. https://doi.org/10.1039/D0SC04896H

47) Burai Patrascu, M.; Pottel, J.; Pinus, S.; Bezanson, M.; Norrby, P.-O.; Moitessier, N. From Desktop to Benchtop with Automated Computational Workflows for Computer-Aided Design in Asymmetric Catalysis. *Nat. Catal.* **2020**, *3*, 574–584. https://doi.org/10.1038/s41929-020-0468-3

48) Zahrt, A. F.; Rose, B. T.; Darrow, W. T.; Henle, J. J.; Denmark, S. E. Computational Methods for Training Set Selection and Error Assessment Applied to Catalyst Design: Guidelines for Deciding Which Reactions to Run First and Which to Run Next. *React. Chem. Eng.* **2021**, *6*, 694–708. https://doi.org/10.1039/D1RE00013F

49) Martin, R.; Buchwald, S. L. Palladium-Catalyzed Suzuki−Miyaura Cross-Coupling Reactions Employing Dialkylbiaryl Phosphine Ligands. *Acc. Chem. Res.* **2008**, *41*, 1461–1473. https://doi.org/10.1021/ar800036s

50) Fu, G. C. The Development of Versatile Methods for Palladium-Catalyzed Coupling Reactions of Aryl Electrophiles through the Use of $P(t\text{-}Bu)_3$ and PCy_3 as Ligands. *Acc. Chem. Res.* **2008**, *41*, 1555–1564. https://doi.org/10.1021/ar800148f

51) Surry, D. S.; Buchwald, S. L. Dialkylbiaryl Phosphines in Pd-Catalyzed Amination: A User's Guide. *Chem. Sci.* **2010**, *2*, 27–50. https://doi.org/10.1039/C0SC00331J

52) Valente, C.; Belowich, M. E.; Hadei, N.; Organ, M. G. Pd-PEPPSI Complexes and the Negishi Reaction. *Eur. J. Org. Chem.* **2010**, 4343–4354. https://doi.org/10.1002/ejoc.201000359

53) Norman, J. P.; Larson, N. G.; Neufeldt, S. R. Different Oxidative Addition Mechanisms for 12- and 14-Electron Palladium(0) Explain Ligand-Controlled Divergent Site Selectivity. *ACS Catal.* **2022**, *12*, 8822–8828. https://doi.org/10.1021/acscatal.2c01698

54) Norman, J. P.; Neufeldt, S. R. The Road Less Traveled: Unconventional Site Selectivity in Palladium-Catalyzed Cross-Couplings of Dihalogenated N-Heteroarenes. *ACS Catal.* **2022**, *12*, 12014–12026. https://doi.org/10.1021/acscatal.2c03743

55) Durand, D. J.; Fey, N. Computational Ligand Descriptors for Catalyst Design. *Chem. Rev.* **2019**, *119*, 6561–6594. https://doi.org/10.1021/acs.chemrev.8b00588

56) Fey, N.; Harvey, J. N.; Lloyd-Jones, G. C.; Murray, P.; Orpen, A. G.; Osborne, R.; Purdie, M. Computational Descriptors for Chelating P,P- and P,N-Donor Ligands. *Organometallics* **2008**, *27*, 1372–1383. https://doi.org/10.1021/om700840h

57) Fey, N.; Haddow, M. F.; Harvey, J. N.; McMullin, C. L.; Orpen, A. G. A Ligand Knowledge Base for Carbenes (LKB-C): Maps of Ligand Space. *Dalton Trans.* **2009**, 8183–8196. https://doi.org/10.1039/B909229C

58) Jover, J.; Fey, N.; Harvey, J. N.; Lloyd-Jones, G. C.; Orpen, A. G.; Owen-Smith, G. J. J.; Murray, P.; Hose, D. R. J.; Osborne, R.; Purdie, M. Expansion of the Ligand Knowledge Base for Monodentate P-Donor Ligands (LKB-P). *Organometallics* **2010**, *29*, 6245–6258. https://doi.org/10.1021/om100648v

59) Hueffel, J. A.; Sperger, T.; Funes-Ardoiz, I.; Ward, J. S.; Rissanen, K.; Schoenebeck, F. Accelerated Dinuclear Palladium Catalyst Identification through Unsupervised Machine Learning. *Science* **2021**, *374*, 1134–1140. https://doi.org/10.1126/science.abj0999

60) Gensch, T.; dos Passos Gomes, G.; Friederich, P.; Peters, E.; Gaudin, T.; Pollice, R.; Jorner, K.; Nigam, A.; Lindner-D'Addario, M.; Sigman, M. S.; Aspuru-Guzik, A. A Comprehensive Discovery Platform for Organophosphorus Ligands for Catalysis. *J. Am. Chem. Soc.* **2022**, *144*, 1205–1217. https://doi.org/10.1021/jacs.1c09718

61) Gensch, T.; Smith, S. R.; Colacot, T. J.; Timsina, Y. N.; Xu, G.; Glasspoole, B. W.; Sigman, M. S. Design and Application of a Screening Set for Monophosphine Ligands in Cross-Coupling. *ACS Catal.* **2022**, *12*, 7773–7780. https://doi.org/10.1021/acscatal.2c01970

62) Wodrich, M. D.; Sawatlon, B.; Busch, M.; Corminboeuf, C. The Genesis of Molecular Volcano Plots. *Acc. Chem. Res.* **2021**, *54*, 1107–1117. https://doi.org/10.1021/acs.accounts.0c00857

63) Laplaza, R.; Das, S.; Wodrich, M. D.; Corminboeuf, C. Constructing and Interpreting Volcano Plots and Activity Maps to Navigate Homogeneous Catalyst Landscapes. *Nat. Protoc.* **2022**, *17*, 2550–2569. https://doi.org/10.1038/s41596-022-00726-2

64) Sabatier, P. Hydrogénations et Déshydrogénations Par Catalyse. *Ber. Dtsch. Chem. Ges.* **1911**, *44*, 1984-2001. https://doi.org/10.1002/cber.19110440303

65) Sabatier, P. *La Catalyse en Chimie Organique,* 2nd ed.; Béranger, 1913

66) Busch, M.; Wodrich, M. D.; Corminboeuf, C. Linear Scaling Relationships and Volcano Plots in Homogeneous Catalysis – Revisiting the Suzuki Reaction. *Chem. Sci.* **2015**, *6*, 6754–6761. https://doi.org/10.1039/C5SC02910D

67) Meyer, B.; Sawatlon, B.; Heinen, S.; Lilienfeld, O. A. von; Corminboeuf, C. Machine Learning Meets Volcano Plots: Computational Discovery of Cross-Coupling Catalysts. *Chem. Sci.* **2018**, *9*, 7069–7077. https://doi.org/10.1039/C8SC01949E

68) Sawatlon, B.; Wodrich, M. D.; Meyer, B.; Fabrizio, A.; Corminboeuf, C. Data Mining the C−C Cross-Coupling Genome. *ChemCatChem* **2019**, *11*, 4096–4107. https://doi.org/10.1002/cctc.201900597

69) Sawatlon, B.; Wodrich, M. D.; Corminboeuf, C. Probing Substrate Scope with Molecular Volcanoes. *Org. Lett.* **2020**, *22*, 7936–7941. https://doi.org/10.1021/acs.orglett.0c02862

70) Liu, D.; Xu, Z.; Lu, X.; Yu, H.; Fu, Y. Linear Regression Model for Predicting Allyl Alcohol C–O Bond Activity under Palladium Catalysis. *ACS Catal.* **2022**, *12*, 13921–13929. https://doi.org/10.1021/acscatal.2c03847

71) Jing, J.; Huo, X.; Shen, J.; Fu, J.; Meng, Q.; Zhang, W. Direct Use of Allylic Alcohols and Allylic Amines in Palladium-Catalyzed Allylic Amination. *Chem. Commun.* **2017**, *53*, 5151–5154. https://doi.org/10.1039/C7CC01069A

72) Matsuoka, W.; Harabuchi, Y.; Maeda, S. Virtual Ligand Strategy in Transition Metal Catalysis Toward Highly Efficient Elucidation of Reaction Mechanisms and Computational Catalyst Design. *ACS Catal.* **2023**, *13*, 5697–5711. https://doi.org/10.1021/acscatal.3c00576

73) Lyngvi, E.; Sanhueza, I. A.; Schoenebeck, F. Dispersion Makes the Difference: Bisligated Transition States Found for the Oxidative Addition of $Pd(PtBu_3)_2$ to Ar-OSO_2R and Dispersion-Controlled Chemoselectivity in Reactions with $Pd[P(IPr)(tBu_2)]_2$. *Organometallics* **2015**, *34*, 805–812. https://doi.org/10.1021/om501199t

74) Shultz, C. S.; Krska, S. W. Unlocking the Potential of Asymmetric Hydrogenation at Merck. *Acc. Chem. Res.* **2007**, *40*, 1320–1326. https://doi.org/10.1021/ar700141v

75) Krska, S. W.; DiRocco, D. A.; Dreher, S. D.; Shevlin, M. The Evolution of Chemical High-Throughput Experimentation to Address Challenging Problems in Pharmaceutical Synthesis. *Acc. Chem. Res.* **2017**, *50*, 2976–2985. https://doi.org/10.1021/acs.accounts.7b00428

76) Welch, C. J. High Throughput Analysis Enables High Throughput Experimentation in Pharmaceutical Process Research. *React. Chem. Eng.* **2019**, *4*, 1895–1911. https://doi.org/10.1039/C9RE00234K

77) Allen, C. L.; Leitch, D. C.; Anson, M. S.; Zajac, M. A. The Power and Accessibility of High-Throughput Methods for Catalysis Research. *Nat. Catal.* **2019**, *2*, 2–4. https://doi.org/10.1038/s41929-018-0220-4

78) Isbrandt, E. S.; Sullivan, R. J.; Newman, S. G. High Throughput Strategies for the Discovery and Optimization of Catalytic Reactions. *Angew. Chem. Int. Ed.* **2019**, *58*, 7180–7191. https://doi.org/10.1002/anie.201812534

79) *The Power of High-Throughput Experimentation: General Topics and Enabling Technologies for Synthesis and Catalysis (Volume 1)*. Emmert, M. H.; Jouffroy, M.; Leitch, D. C., Eds.; ACS Symposium Series, Vol. 1419; American Chemical Society, 2022. DOI: 10.1021/bk-2022-1419. https://doi.org/10.1021/bk-2022-1419

80) *The Power of High-Throughput Experimentation: Case Studies from Drug Discovery, Drug Development, and Catalyst Discovery (Volume 2)*; Emmert, M. H.; Jouffroy, M.; Leitch, D. C., Eds.; ACS Symposium Series, Vol. 1420; American Chemical Society, 2022. DOI: 10.1021/bk-2022-1420. https://doi.org/10.1021/bk-2022-1420

81) 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. Nanomole-Scale High-Throughput Chemistry for the Synthesis of Complex Molecules. *Science* **2015**, *347* (6217), 49–53. https://doi.org/10.1126/science.1259203

82) Lin, S.; Dikler, S.; Blincoe, W. D.; Ferguson, R. D.; Sheridan, R. P.; Peng, Z.; Conway, D. V.; Zawatzky, K.; Wang, H.; Cernak, T.; Davies, I. W.; DiRocco, D. A.; Sheng, H.; Welch, C.

J.; Dreher, S. D. Mapping the Dark Space of Chemical Reactions with Extended Nanomole Synthesis and MALDI-TOF MS. *Science* **2018**, *361*, eaar6236. https://doi.org/10.1126/science.aar6236

83) Perera, D.; Tucker, J. W.; Brahmbhatt, S.; Helal, C. J.; Chong, A.; Farrell, W.; Richardson, P.; Sach, N. W. A Platform for Automated Nanomole-Scale Reaction Screening and Micromole-Scale Synthesis in Flow. *Science* **2018**, *359*, 429–434. https://doi.org/10.1126/science.aap9112

84) Mahjour, B.; Shen, Y.; Cernak, T. Ultrahigh-Throughput Experimentation for Information-Rich Chemical Synthesis. *Acc. Chem. Res.* **2021**, *54*, 2337–2346. https://doi.org/10.1021/acs.accounts.1c00119

85) Gesmundo, N.; Dykstra, K.; Douthwaite, J. L.; Kao, Y.-T.; Zhao, R.; Mahjour, B.; Ferguson, R.; Dreher, S.; Sauvagnat, B.; Saurí, J.; Cernak, T. Miniaturization of Popular Reactions from the Medicinal Chemists' Toolbox for Ultrahigh-Throughput Experimentation. *Nat. Synth.* **2023**. https://doi.org/10.1038/s44160-023-00351-1

86) 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. *Nat. Commun.* **2023**, *14*, 3924. https://doi.org/10.1038/s41467-023-39531-0

87) Cook, A.; Clément, R.; Newman, S. G. Reaction Screening in Multiwell Plates: High-Throughput Optimization of a Buchwald–Hartwig Amination. *Nat. Protoc.* **2021**, *16*, 1152–1169. https://doi.org/10.1038/s41596-020-00452-7

88) Lee, J.; Schmink, J. R.; Berritt, S. Introduction of Low-Barrier High-Throughput Experimentation in the Undergraduate Laboratory: Suzuki—Miyaura Reaction. *J. Chem. Educ.* **2020**, *97*, 538–542. https://doi.org/10.1021/acs.jchemed.9b00794

89) Becica, J.; Hruszkewycz, D. P.; Steves, J. E.; Elward, J. M.; Leitch, D. C.; Dobereiner, G. E. High-Throughput Discovery and Evaluation of a General Catalytic Method for N-Arylation of Weakly Nucleophilic Sulfonamides. *Org. Lett.* **2019**, *21*, 8981–8986. https://doi.org/10.1021/acs.orglett.9b03380

90) Becica, J.; Heath, O. R. J.; Zheng, C. H. M.; Leitch, D. C. Palladium-Catalyzed Cross-Coupling of Alkenyl Carboxylates. *Angew. Chem. Int. Ed.* **2020**, *59*, 17277–17281. https://doi.org/10.1002/anie.202006586

91) Gaube, G.; Fernandez, N. P.; C. Leitch, D. An Evaluation of Palladium-Based Catalysts for the Base-Free Borylation of Alkenyl Carboxylates. *New J. Chem.* **2021**, *45*, 20095–20098. https://doi.org/10.1039/D1NJ04008A

92) Pipaón Fernández, N.; Gaube, G.; Woelk, K. J.; Burns, M.; Hruszkewycz, D. P.; Leitch, D. C. Palladium-Catalyzed Direct C–H Alkenylation with Enol Pivalates Proceeds via Reversible C–O Oxidative Addition to Pd(0). *ACS Catal.* **2022**, *12*, 6997–7003. https://doi.org/10.1021/acscatal.2c01305

93) Dhake, K.; Woelk, K. J.; Becica, J.; Un, A.; Jenny, S. E.; Leitch, D. C. Beyond Bioisosteres: Divergent Synthesis of Azabicyclohexanes and Cyclobutenyl Amines from Bicyclobutanes. *Angew. Chem. Int. Ed.* **2022**, *61*, e202204719. https://doi.org/10.1002/anie.202204719

94) Lu, J.; Paci, I.; Leitch, D. C. A Broadly Applicable Quantitative Relative Reactivity Model for Nucleophilic Aromatic Substitution (S_NAr) Using Simple Descriptors. *Chem. Sci.* **2022**, *13*, 12681–12695. https://doi.org/10.1039/D2SC04041G

95) Becica, J.; Glaze, O. D.; Hruszkewycz, D. P.; Dobereiner, G. E.; Leitch, D. C. The Influence of Additives on Orthogonal Reaction Pathways in the Mizoroki–Heck Arylation of Vinyl Ethers. *React. Chem. Eng.* **2021**, *6*, 1212–1219. https://doi.org/10.1039/D1RE00124H

96) Takahashi, Y.; Ito, T.; Sakai, S.; Ishii, Y. A Novel Palladium(0) Complex; Bis(dibenzylideneacetone)Palladium(0). *J. Chem. Soc. Chem. Commun.* **1970**, 1065–1066. https://doi.org/10.1039/C29700001065

97) Amatore, C.; Jutand, A. Role of dba in the Reactivity of Palladium(0) Complexes Generated in situ from Mixtures of Pd(dba)2 and Phosphines. *Coord. Chem. Rev.* **1998**, *178–180*, 511–528. https://doi.org/10.1016/S0010-8545(98)00073-3

98) Zalesskiy, S. S.; Ananikov, V. P. $Pd_2(dba)_3$ as a Precursor of Soluble Metal Complexes and Nanoparticles: Determination of Palladium Active Species for Catalysis and Synthesis. *Organometallics* **2012**, *31*, 2302–2309. https://doi.org/10.1021/om201217r

99) Weber, P.; Biafora, A.; Doppiu, A.; Bongard, H.-J.; Kelm, H.; Gooßen, L. J. A Comparative Study of Dibenzylideneacetone Palladium Complexes in Catalysis. *Org. Process Res. Dev.* **2019**, *23*, 1462–1470. https://doi.org/10.1021/acs.oprd.9b00214

100) Huang, J.; Isaac, M.; Watt, R.; Becica, J.; Dennis, E.; Saidaminov, M. I.; Sabbers, W. A.; Leitch, D. C. ^{DMP}DAB–Pd–MAH: A Versatile Pd(0) Source for Precatalyst Formation, Reaction

Screening, and Preparative-Scale Synthesis. *ACS Catal.* **2021**, *11*, 5636–5646. https://doi.org/10.1021/acscatal.1c00288

101) Cavell, K. J.; Stufkens, D. J.; Vrieze, K. 1,4-Diazabutadiene Olefin Complexes of Zerovalent Palladium: Preparation and Characterization. *Inorg. Chim. Acta* **1981**, *47*, 67–76. https://doi.org/10.1016/S0020-1693(00)89309-3

102) van Asselt, R.; Elsevier, C. J. New Palladium Complexes of Cis-Fixed Bidentate Nitrogen Ligands as Catalysts for Carbon-Carbon Bond Formation. *Organometallics* **1992**, *11*, 1999–2001. https://doi.org/10.1021/om00042a010

103) de Pater, J. J. M.; Tromp, D. S.; Tooke, D. M.; Spek, A. L.; Deelman, B.-J.; van Koten, G.; Elsevier, C. J. Palladium(0)-Alkene Bis(Triarylphosphine) Complexes as Catalyst Precursors for the Methoxycarbonylation of Styrene. *Organometallics* **2005**, *24*, 6411–6419. https://doi.org/10.1021/om0506419

104) Leitch, D. C. Unpublished Results.

105) Sather, A. C.; Martinot, T. A. Data-Rich Experimentation Enables Palladium-Catalyzed Couplings of Piperidines and Five-Membered (Hetero)Aromatic Electrophiles. *Org. Process Res. Dev.* **2019**, *23*, 1725–1739. https://doi.org/10.1021/acs.oprd.9b00233

106) Atwater, B.; Chandrasoma, N.; Mitchell, D.; Rodriguez, M. J.; Organ, M. G. Pd-PEPPSI-IHeptCl: A General-Purpose, Highly Reactive Catalyst for the Selective Coupling of Secondary Alkyl Organozincs. *Chem. – Eur. J.* **2016**, *22*, 14531–14534. https://doi.org/10.1002/chem.201603603

107) Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning. *Science* **2018**, *360*, 186–190. https://doi.org/10.1126/science.aar5169

108) Chuang, K. V.; Keiser, M. J. Comment on "Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning." *Science* **2018**, *362*, eaat8603. https://doi.org/10.1126/science.aat8603

109) Estrada, J. G.; Ahneman, D. T.; Sheridan, R. P.; Dreher, S. D.; Doyle, A. G. Response to Comment on "Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning." *Science* **2018**, *362*, eaat8763. https://doi.org/10.1126/science.aat8763

110) Lones, M. A. How to Avoid Machine Learning Pitfalls: A Guide for Academic Researchers. *arXiv* **2023**. https://doi.org/10.48550/arXiv.2108.02497

111) Dotson, J. J.; van Dijk, L.; Timmerman, J. C.; Grosslight, S.; Walroth, R. C.; Gosselin, F.; Püntener, K.; Mack, K. A.; Sigman, M. S. Data-Driven Multi-Objective Optimization Tactics for Catalytic Asymmetric Reactions Using Bisphosphine Ligands. *J. Am. Chem. Soc.* **2023**, *145*, 110–121. https://doi.org/10.1021/jacs.2c08513

112) Xu, J.; Grosslight, S.; Mack, K. A.; Nguyen, S. C.; Clagg, K.; Lim, N.-K.; Timmerman, J. C.; Shen, J.; White, N. A.; Sirois, L. E.; Han, C.; Zhang, H.; Sigman, M. S.; Gosselin, F. Atroposelective Negishi Coupling Optimization Guided by Multivariate Linear Regression Analysis: Asymmetric Synthesis of KRAS G12C Covalent Inhibitor GDC-6036. *J. Am. Chem. Soc.* **2022**, *144*, 20955–20963. https://doi.org/10.1021/jacs.2c09917

113) Hammett, L. P. The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives. *J. Am. Chem. Soc.* **1937**, *59*, 96–103. https://doi.org/10.1021/ja01280a022

114) Hansch, Corwin.; Leo, A.; Taft, R. W. A Survey of Hammett Substituent Constants and Resonance and Field Parameters. *Chem. Rev.* **1991**, *91*, 165–195. https://doi.org/10.1021/cr00002a004

115) Yau, H. M.; Haines, R. S.; Harper, J. B. A Robust, "One-Pot" Method for Acquiring Kinetic Data for Hammett Plots Used To Demonstrate Transmission of Substituent Effects in Reactions of Aromatic Ethyl Esters. *J. Chem. Educ.* **2015**, *92*, 538–542. https://doi.org/10.1021/ed400701m

116) Bain, R. M.; Pulliam, C. J.; Yan, X.; Moore, K. F.; Müller, T.; Cooks, R. G. Mass Spectrometry in Organic Synthesis: Claisen–Schmidt Base-Catalyzed Condensation and Hammett Correlation of Substituent Effects. *J. Chem. Educ.* **2014**, *91*, 1985–1989. https://doi.org/10.1021/ed500288m

117) Maes, B. U. W.; Verbeeck, S.; Verhelst, T.; Ekomié, A.; von Wolff, N.; Lefèvre, G.; Mitchell, E. A.; Jutand, A. Oxidative Addition of Haloheteroarenes to Palladium(0): Concerted versus S_NAr-Type Mechanism. *Chem. – Eur. J.* **2015**, *21*, 7858–7865. https://doi.org/10.1002/chem.201406210

118) Haas, B. C.; Goetz, A. E.; Bahamonde, A.; McWilliams, J. C.; Sigman, M. S. Predicting relative efficiency of amide bond formation using multivariate linear regression. *Proc. Nat. Acad. Sci.* **2022**, *119*, e2118451119. https://doi.org/10.1073/pnas.2118451119

119) Gao, X.; Kagan, H. B. One-Pot Multi-Substrate Screening in Asymmetric Catalysis. *Chirality* **1998**, *10*, 120–124. https://doi.org/10.1002/chir.19

120) Satyanarayana, T.; Kagan, H. B. The Multi-Substrate Screening of Asymmetric Catalysts. *Adv. Synth. Catal.* **2005**, *347*, 737–748. https://doi.org/10.1002/adsc.200505057

121) an der Heiden, M. R.; Plenio, H. The Effect of Steric Bulk in Sonogashira Coupling Reactions. *Chem. Commun.* **2007**, 972–974. https://doi.org/10.1039/B616608C

122) an der Heiden, M. R.; Plenio, H.; Immel, S.; Burello, E.; Rothenberg, G.; Hoefsloot, H. C. J. Insights into Sonogashira Cross-Coupling by High-Throughput Kinetics and Descriptor Modeling. *Chem. – Eur. J.* **2008**, *14*, 2857–2866. https://doi.org/10.1002/chem.200701418

123) Lu, T.; Chen, F. Quantitative Analysis of Molecular Surface Based on Improved Marching Tetrahedra Algorithm. *J. Mol. Graph. Model.* **2012**, *38*, 314–323. https://doi.org/10.1016/j.jmgm.2012.07.004

124) Lu, T.; Chen, F. Multiwfn: A Multifunctional Wavefunction Analyzer. *J. Comput. Chem.* **2012**, *33*, 580–592. https://doi.org/10.1002/jcc.22885

125) Suresh, C. H.; Remya, G. S.; Anjalikrishna, P. K. Molecular electrostatic potential analysis: A powerful tool to interpret and predict chemical reactivity. *WIREs Comp. Mol. Sci.* **2022**, *12*, e1601. https://doi.org/10.1002/wcms.1601

126) Suresh, C. H.; Koga, N. Quantifying the Electronic Effect of Substituted Phosphine Ligands via Molecular Electrostatic Potential. *Inorg. Chem.* **2002**, *41*, 1573–1578. https://doi.org/10.1021/ic0109400

127) Anjali, B. A.; Suresh, C. H. Interpreting Oxidative Addition of Ph–X (X = CH_3, F, Cl, and Br) to Monoligated Pd(0) Catalysts Using Molecular Electrostatic Potential. *ACS Omega* **2017**, *2*, 4196–4206. https://doi.org/10.1021/acsomega.7b00745

128) Hirsch, J. A. Table of Conformational Energies—1967. In *Topics in Stereochemistry*; John Wiley & Sons, Ltd, 1967; pp 199–222. https://doi.org/10.1002/9780470147108.ch4

129) Klein, J.; Khartabil, H.; Boisson, J.-C.; Contreras-García, J.; Piquemal, J.-P.; Hénon, E. New Way for Probing Bond Strength. *J. Phys. Chem. A* **2020**, *124*, 1850–1860. https://doi.org/10.1021/acs.jpca.9b09845

130) Almond-Thynne, J.; Blakemore, D. C.; Pryde, D. C.; Spivey, A. C. Site-Selective Suzuki–Miyaura Coupling of Heteroaryl Halides – Understanding the Trends for Pharmaceutically Important Classes. *Chem. Sci.* **2017**, *8*, 40–62. https://doi.org/10.1039/C6SC02118B

131) Reeves, E. K.; Bauman, O. R.; Mitchem, G. B.; Neufeldt, S. R. Solvent Effects on the Selectivity of Palladium-Catalyzed Suzuki-Miyaura Couplings. *Isr. J. Chem.* **2020**, *60*, 406–409. https://doi.org/10.1002/ijch.201900082

132) Reeves, E. K.; Entz, E. D.; Neufeldt, S. R. Chemodivergence between Electrophiles in Cross-Coupling Reactions. *Chem. – Eur. J.* **2021**, *27*, 6161–6177. https://doi.org/10.1002/chem.202004437

133) Elias, E. K.; Rehbein, S. M.; Neufeldt, S. R. Solvent Coordination to Palladium Can Invert the Selectivity of Oxidative Addition. *Chem. Sci.* **2022**, *13*, 1618–1628. https://doi.org/10.1039/D1SC05862B

134) Norman, J. P.; Larson, N. G.; Entz, E. D.; Neufeldt, S. R. Unconventional Site Selectivity in Palladium-Catalyzed Cross-Couplings of Dichloroheteroarenes under Ligand-Controlled and Ligand-Free Systems. *J. Org. Chem.* **2022**, *87*, 7414–7421. https://doi.org/10.1021/acs.joc.2c00665

135) Ibsen, G. M.; Menezes da Silva, V. H.; Pettigrew, J. C.; Neufeldt, S. R. Triflate-Selective Suzuki Cross-Coupling of Chloro- and Bromoaryl Triflates Under Ligand-Free Conditions. *Chem. – Asian J.* **2023**, *18*, e202300036. https://doi.org/10.1002/asia.202300036

136) Dombrowski, A. W.; Gesmundo, N. J.; Aguirre, A. L.; Sarris, K. A.; Young, J. M.; Bogdan, A. R.; Martin, M. C.; Gedeon, S.; Wang, Y. Expanding the Medicinal Chemist Toolbox: Comparing Seven C(sp^2)–C(sp^3) Cross-Coupling Methods by Library Synthesis. *ACS Med. Chem. Lett.* **2020**, *11*, 597-604. https://doi.org/10.1021/acsmedchemlett.0c00093