A reactivity map for oxidative addition enables quantitative predictions for multiple catalytic reaction classes

Jingru Lu, Sofia Donnecke, Irina Paci,[*] David C. Leitch[*]

Department of Chemistry, University of Victoria, 3800 Finnerty Rd. Victoria BC, CANADA, V8P 5C2.

*ipaci@uvic.ca; dcleitch@uvic.ca.

ABSTRACT: Making accurate, quantitative predictions of chemical reactivity based on molecular structure is an unsolved problem in chemical synthesis, particularly for complex molecules. We report a generally applicable, mechanistically based structure-reactivity approach to create a quantitative model for the oxidative addition of (hetero)aryl electrophiles to palladium(0), which is a key step in myriad catalytic processes. This model links simple molecular descriptors to relative rates of oxidative addition for 79 substrates, including chloride, bromide and triflate leaving groups. Because oxidative addition often controls the rate and/or selectivity of palladium-catalyzed reactions, this model can be used to make quantitative predictions about catalytic reaction outcomes. Demonstrated applications include a multivariate linear model for the initial rate of Sonogashira coupling reactions, and successful site-selectivity predictions for a series of multihalogenated substrates relevant to the synthesis of pharmaceuticals and natural products.

The synthesis of structurally complex organic molecules relies on forging new chemical bonds between diverse molecular building blocks. Catalytic cross-coupling is one of the most versatile and widely-used methods to link these molecular fragments,[1] with applications ranging from the manufacture of active pharmaceutical ingredients,[2] to the selective modification of biomolecules,[3] to the creation of new functional materials.[4] While much has been done to develop and understand new cross-coupling reactions and catalysts, less is known about how the specific molecular structures of the building blocks affect the likelihood of successful coupling. As a result, time- and resource-intensive reaction screening and optimization campaigns are often required for each new synthetic target. All too often these efforts fail, impeding access to potentially promising new medicines and materials. Emerging approaches in reactivity prediction that combine high-throughput experimentation[5–8] with molecular descriptor sets[9–11] and multivariate statistical analysis including machine learning[12–16] can accelerate this process and increase success rates; however, the predictions generated by these approaches are often limited to the specific reaction under investigation (Figure 1A). Developing and refining the next generation of organic chemistry tools, including computer-aided synthesis design, automated reaction optimization, and predictive algorithms,[17] requires the development of general and quantitative frameworks linking molecular structure to reactivity for many different reactants and catalysts.

Here, we describe an approach to predict outcomes for a broad range of catalytic processes by focusing on the structure-reactivity relationships for one key mechanistic step. To demonstrate this approach, we have applied it to oxidative addition, a fundamental organometallic transformation common to many catalytic reactions.[18] Oxidative addition is particularly relevant to palladium-catalyzed cross-coupling, where it often controls the reaction rate and/or selectivity[19] (Figure 1B). We hypothesized that a quantitative structure-reactivity model for the oxidative

addition[20] of aryl electrophiles to a catalytically-relevant Pd(0) complex would enable generally-applicable predictions for many palladium-catalyzed reactions under a variety of conditions. This model was assembled by correlating experimental relative rate data with easily-obtained molecular descriptors for a diverse set of substrates, with an emphasis on incorporating pharmaceutically-relevant heterocycles. As hypothesized, the resulting model – a "reactivity map" – is general: it can predict catalytic rate constants as a function of substrate structure even when different catalysts/solvents are used, can correctly identify the most reactive position(s) when multiple reaction sites are available, and can also help to guide synthetic route planning. Importantly from both a fundamental and practical perspective, the predictive ability of our reactivity map extends well beyond the specific molecular structures and reaction conditions included in the initial data set. By revealing how subtle changes to the reacting molecules affect a key step in a catalytic mechanism, this map serves as a powerful predictive tool for multiple reaction classes, and will enable more sophisticated and accurate computer-aided retrosynthetic design.

Figure 1. Approaches to catalytic reaction prediction. (**A**) Approaches to quantitative reactivity predictions. (**B**) Simplified mechanism for Pd-catalyzed cross-coupling, highlighting oxidative addition as the rate and/or selectivity determining step. (**C**) Competition experiment approach to map relative rates of oxidative addition. (**D**) Reactivity scale for oxidative addition to Pd(PCy$_3$)$_2$ with selected substrates; experimental $\Delta G^{\ddagger}_{OA}$ for 2-bromo-5-nitropyridine set to 0 kJ mol^{-1}.

Results and Discussion

Development of the oxidative addition reactivity map. As the basis for a quantitative structure/reactivity model for cross-coupling catalysis, we conducted a series of oxidative addition competition experiments in THF using a library of 79 (hetero)aryl chlorides, bromides, and triflates, reacting with $Pd(PCy_3)_2$ (Figure 1C). Measuring the Pd(II) product ratio by ^{31}P NMR spectroscopy gives relative observed rate constants, and the corresponding relative free energies of activation ($\Delta G^{\ddagger}_{OA}$). As a check on these kinetic data, we constructed Hammett plots for five sets of substrates, obtaining reaction constants consistent with previous reports (Figures S31-37). From these data, we have constructed a unified reactivity scale that spans more than 7 orders of magnitude in rate (Figure 1D), containing substrates with a wide variety of steric and electronic parameters, as well as the three aforementioned leaving groups. We also isolated and characterized six representative Pd(II) oxidative addition complexes to confirm their structures as *trans*-$Pd(PCy_3)_2(Ar)(X)$ (compounds **S1**-**S6**, Figures S7-S30).

Using the mechanistic features of oxidative addition to Pd(0) as a guide, we considered molecular descriptors that would provide mechanistically meaningful correlations between substrate structures and oxidative addition reactivity (Figure 2A).[20] Mechanisms for aryl halide oxidative addition to Pd(0) have been extensively studied both computationally[21–24] and experimentally,[25–29] and are generally proposed to proceed via initial coordination of the aromatic π-system to Pd. Two bonding extremes can be envisioned for the π-complex intermediate, where the degree of polarization of the coordinated C=C or C=N bond in the substrate influences partial charge distribution in the π-complex. From this intermediate, two types of oxidative addition transition state have been proposed: a 3-centered, relatively non-polar transition state involving simultaneous Pd–C and Pd–X bond formation, and a polarized transition state with C–X heterolytic bond cleavage occurring alongside Pd–C bond formation; this latter pathway resembles the proposed mechanism for nucleophilic aromatic substitution (S_NAr).[26,29]

Based on the structural and electronic features of these transition states, we built our reactivity map for oxidative addition from a combination of average molecular electrostatic potentials (*ESP*) as electronic descriptors for specific atoms in the substrate,[30] A-values as steric descriptors,[31] and the intrinsic bond strength index (*IBSI*) as a bond energy descriptor (Figure 2B).[32] Importantly, all of these descriptors are directly calculated from electronic wavefunctions (obtained from density functional theory calculations), or are tabulated in the literature. An initial multivariate linear regression analysis[13,33] of the descriptor sets for the Ar–Cl and Ar–Br substrates versus relative $\Delta G^{\ddagger}_{OA}$ (kJ mol^{-1}) reveals a strong correlation across the entire substrate library (Figure 2C). This model incorporates two *ESP* values: one for the carbon undergoing substitution (ESP_1), and a second for an atom adjacent to the reactive site (ESP_2). Similarly, two substituent A-values for groups R_1 and R_2 account for steric effects on the oxidative addition rate. While the steric effect of groups *ortho* to the reactive C–X bond is intuitive, the effect of R_2 for 2-halopyridine substrates is not initially obvious; however, our experimental results reveal the steric influence of R_2 is approximately equal to that of R_1 throughout the 2-halopyridine substrate set. As these two terms have very similar coefficients when treated separately during linear regression (Table S9), we opted to use the sum ($A_1 + A_2$) as a single descriptor. Finally, including the *IBSI* is necessary to create a model applicable to both (hetero)aryl chlorides and bromides; note that we have multiplied the small-magnitude *IBSI* (0.23-0.40) by 1000 to give a better representation of the relative contribution of each descriptor to the predicted $\Delta G^{\ddagger}_{OA}$.

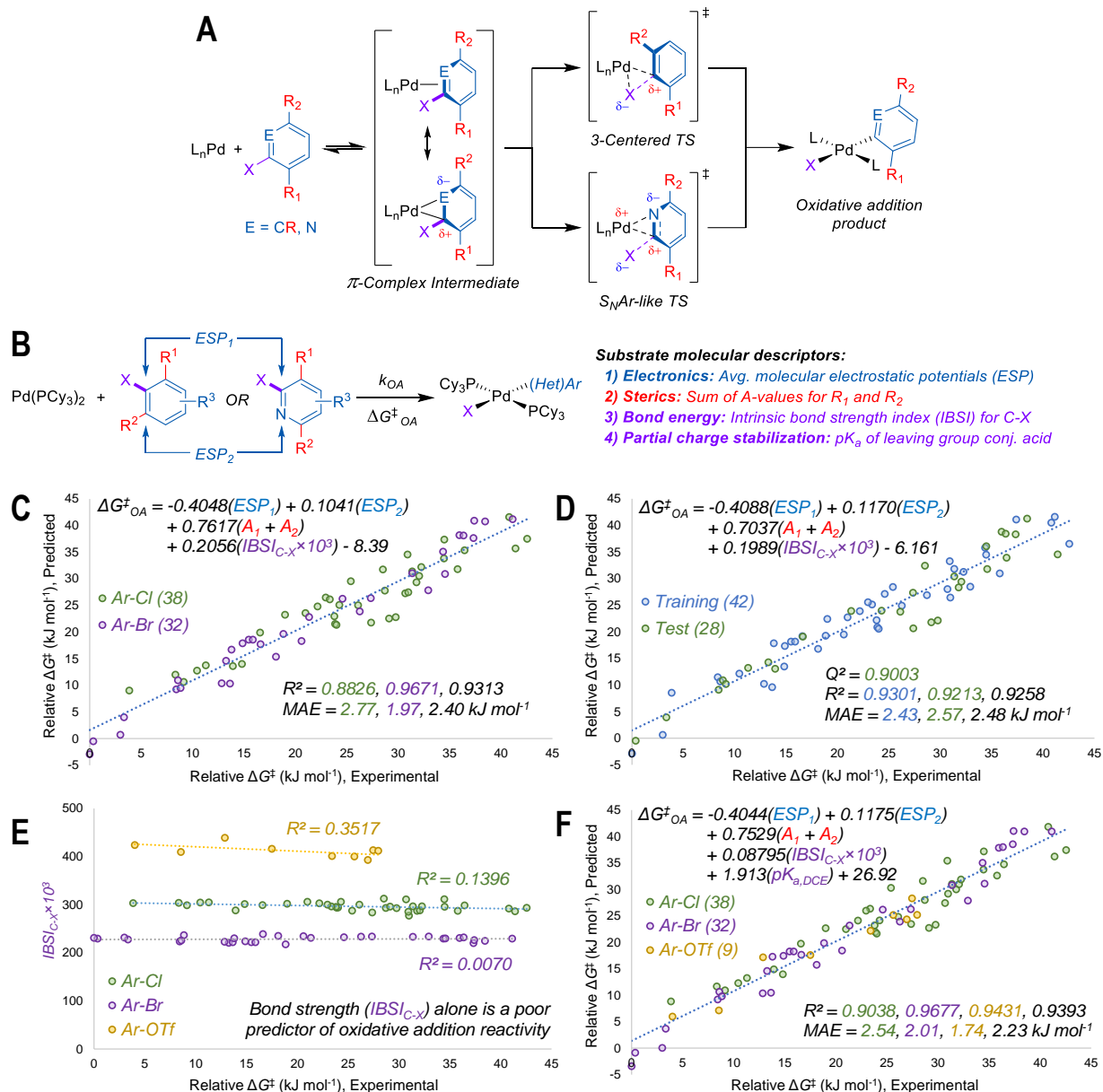**Figure 2. Design and performance of a general and quantitative reactivity map for oxidative addition to Pd(0).** (**A**) General mechanism for oxidative addition to $L_nPd(0)$, with π-complex intermediate preceding either Pd insertion into C–X bond, or an S_NAr-like displacement of X^-. (**B**) Molecular descriptors used to model oxidative addition reactivity as a function of substrate structure. (**C**) Multivariate linear regression model of $\Delta G^{\ddagger}_{OA}$ for 70 Ar–Cl and Ar–Br substrates in THF, including all data points in regression analysis. (**D**) Representative multivariate linear regression model generated using a 60/40 training/test split. (**E**) Univariate plot of $IBSI_{C–X}$ versus $\Delta G^{\ddagger}_{OA}$ for Ar–Cl, Ar–Br, and Ar–OTf, demonstrating that bond strength alone is a poor predictor of oxidative addition reactivity. (**F**) Unified linear regression model of $\Delta G^{\ddagger}_{OA}$ for Ar–Cl, Ar–Br, and Ar–OTf substrates in THF, including all data points in regression analysis.

We have evaluated the robustness of this linear model by regression analysis of five random training (60%) and test (40%) sets, and comparing the mean absolute errors (MAE) and predictive squared correlation coefficients (Q^2);[34] one example is shown in Figure 1D (remainder in Figures S60-63). All of these analyses give similar linear equations, and excellent agreement between predicted and experimental values in the test sets. We also evaluated alternative models (Table S9), and partitioned the data into targeted training and test sets to evaluate out-of-sample prediction accuracy (Figures S64-67). These latter models again give excellent agreement between predicted and experimental values.

Our initial attempts to incorporate aryl triflate electrophiles into this model using the aforementioned descriptor set were unsuccessful, leading to poor correlations and inaccurate predictions. We attribute this to C–X bond strength being an insufficient descriptor to differentiate between leaving groups. Bond strength arguments, often using calculated BDEs, are frequently used to rationalize relative oxidative addition reactivity for different substrates. In our initial model (Figure 2C, D), the *IBSI* term is essentially a step function: there are relatively narrow value distributions within the Ar–Br or Ar–Cl data sets, but a large gap between those data sets that reflects the weaker C–Br bond. Plotting a univariate correlation between $IBSI_{C–X}$ and $\Delta G^{\ddagger}_{OA}$ for all three electrophile classes reveals a similarly tiered structure (Figure 2E). Strikingly, there is essentially zero correlation between *IBSI* and $\Delta G^{\ddagger}_{OA}$ within each electrophile class, and the relative ordering of bond strength (C–O > C–Cl > C–Br) is inconsistent with the fact that aryl triflates react faster than identically-substituted aryl bromides with Pd(PCy$_3$)$_2$ (k_{PhOTf} / k_{PhBr} ~ 100). Clearly, bond strength on its own is a poor predictor of oxidative addition reactivity.

To account for additional transition state stabilization by the leaving group itself, which builds up negative charge as the C–X bond is broken (Figure 2A), we used a simple descriptor of anion stability: the pK_a of the leaving group's conjugate acid. Adding the pK_a values for HOTf, HBr, and HCl (previously measured in DCE as a non-polar solvent[35]) results in a unified predictive model (Figure 2F). Both $IBSI_{C-X}$ and pK_a are required as descriptors, with the unified model re-weighting the *IBSI* term down by a factor of 2. Notably, the relative contributions of the *ESP* and A-value terms remain essentially constant. This model provides, for the first time, a reliable method to quantitatively evaluate the relative reactivity of a hypothetical (hetero)aryl triflate, bromide, or chloride substrate toward oxidative addition with a Pd(0) complex. We again performed cross-validation with a set of five random 60/40 training/test data partitions, which give excellent agreement between experimental and predicted values (Figures S69-S73).

Mechanistic aspects of oxidative addition linked to molecular descriptors. The generality and predictive power of this reactivity map is a direct result of its mechanistic foundation. ESP_1 is related to the electrophilicity of the carbon undergoing oxidative addition, reflecting the degree of partial positive charge in the π-complex intermediate and transition state: a larger positive ESP_1 leads to a smaller $\Delta G^{\ddagger}_{OA}$ and thus faster oxidative addition. In contrast, ESP_2 reflects the degree of negative charge on the adjacent atom (C or N), where a larger negative ESP_2 leads to a faster oxidative addition. Considered together, these *ESP* terms indicate that a more polarized C=C or C=N bond in the substrate leads to faster oxidative addition. It follows from this analysis that the *ESP* at Pd in the intermediate and/or transition state should also affect the oxidative addition rate, with a larger positive ESP_{Pd} correlating with faster oxidative addition. We have confirmed this by determining ESP_{Pd} for a set of 11 calculated π-complex intermediates and 6 transition states (Figures S40-56). Remarkably, there is a linear correlation between ESP_{Pd} for the π-complexes and $\Delta G^{\ddagger}_{OA}$ (Figures 3A, S57), revealing the significant effect that substrate-catalyst bonding has

on the electronic structure of Pd. While the influence of ancillary ligands is often invoked to explain organometallic reactivity, the substrate itself clearly has a profound impact on the reactivity of the metal center.
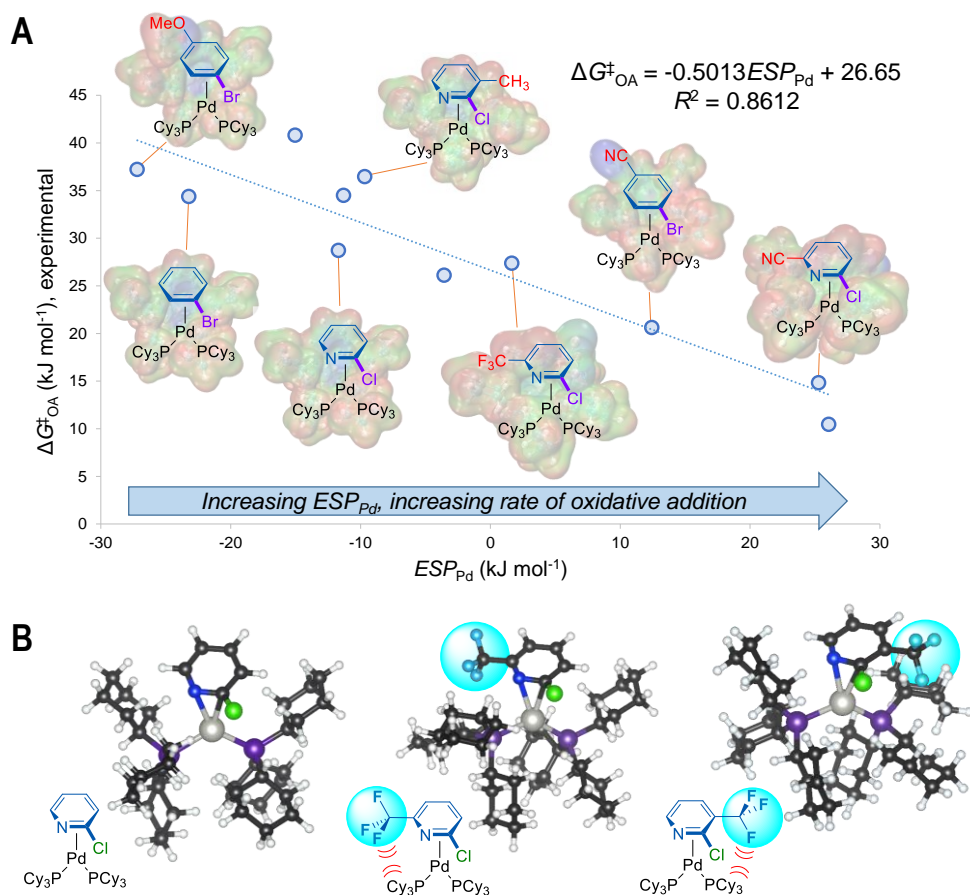

Figure 3. Mechanistic aspects of oxidative addition linked to molecular descriptors. (**A**) ESP_{Pd} for calculated π-complex intermediate structures correlates with oxidative addition rates; structures for 7 of 11 examples shown; electrostatic potential maps for each intermediate are overlayed onto the line structures. (**B**) Calculated structures of π-complex intermediates reveal how steric strain induced by R^1 and R^2 (here, –CF₃ groups) in 2-halopyridines affect oxidative addition reactivity in equal proportions.

Analyzing how the substrate binds to the catalyst also sheds light on the observed equal weights of the steric A-values for R^1 and R^2 in affecting $\Delta G^{\ddagger}_{OA}$ for the 2-halopyridine substrate series. Larger substituents in these positions destabilize the π-complex intermediate and oxidative addition transition state through steric repulsion between the bound substrate and the ancillary phosphine ligands. Because the Pd center coordinates to the C=N bond, substituents adjacent to either C or N will occupy roughly equivalent positions with respect to the phosphines. Comparing the calculated structures of the π-complex intermediates for 2-chloropyridine to its 3- and 6-trifluoromethyl substituted analogues illustrates this feature of the substrate-catalyst interaction

6

(Figure 3B). Importantly, this mechanistic insight into equal steric effects for R^1 and R^2 flows directly from our quantitative reactivity map.

Application case studies. To test our hypothesis that the oxidative addition reactivity map is generally applicable in cross-coupling catalysis, we applied the $\Delta G^{\ddagger}_{OA}$ model to three case studies. The first test case involves modeling the initial rates of Sonogashira coupling reactions, drawn from published data sets containing 410 individual rates (29 substrates and 17 catalysts, Figure 4A).[36,37] We first predicted $\Delta G^{\ddagger}_{OA}$ for each of the 29 aryl bromides in the data sets, using the equation from Figure 2D (the simpler model when considering only halide-based electrophiles); while 9 of these substrates are included in our experimental oxidative addition data set, the other 20 are out-of-sample predictions. Remarkably, the predicted $\Delta G^{\ddagger}_{OA}$ values are linearly-correlated with the corresponding ln k values for all 17 investigated phosphine ligands (Figure S74-75); three of these – for P(nBu)$_3$, PCy$_3$, and P(tBu)$_3$ – are shown in Figure 4B. These correlations hold despite the fact that the Sonogashira reactions are conducted under different conditions (higher temperature, different solvent) than our oxidative addition experiments, and the fact that our predicted $\Delta G^{\ddagger}_{OA}$ is derived from a model using only PCy$_3$. Thus, $\Delta G^{\ddagger}_{OA}$ can be applied to quantitatively predict the outcome of catalytic reactions for out-of-sample substrates, out-of-sample reaction conditions, and even out-of-sample catalysts.

We then expanded the application of $\Delta G^{\ddagger}_{OA}$ by combining it with descriptors for the 17 free phosphines to assemble a single and unified linear model to accurately predict ln k for the entire 410 Sonogashira reaction data set. Two descriptors were calculated for the free phosphines – the average *ESP* at phosphorus, and the percent buried volume (%V_{bur}) at phosphorus[38] – which were combined with $\Delta G^{\ddagger}_{OA}$ in a multivariate regression analysis. As shown in Figure 4C, we obtain excellent linear fit and predictive power with initial rates spanning 10 orders of magnitude. These phosphine descriptors outperform the analogous descriptors calculated for the corresponding mono or bis(phosphine) Pd(0) complexes (Figure S76-77). The training and test sets used to build this model are from a random 60/40 split of substrate set #1, which focuses on electronic effects (Fig 4A). To challenge the model, we reserved substrate set #2, which focuses on steric effects, as an external validation set. Despite the fact that the training set contains no substrates with *ortho*-substituents, and therefore no information about steric effects on reaction rate, the model is still able to predict ln k for substrate set #2 with a mean absolute error of 0.729 ($Q^2 = 0.8009$). Only two reactions are identified as significant outliers (MAE > 2): 2,4,6-triisopropylphenylbromide with P(tBu)$_3$ and PAd$_2$(tBu), representing the most sterically-hindered substrate with the two largest ligands. The robustness of this model in the face of significantly different out-of-sample predictions stems directly from the generality of the oxidative addition reactivity map. The substrate molecular properties – electronics, sterics, and C–X bond strengths – are all encoded within the predicted $\Delta G^{\ddagger}_{OA}$ values by virtue of our diverse oxidative addition training set.

Improved prediction accuracy can be achieved by separating the 410-member data set into two subsets based on phosphine ligand size. One set includes reactions using 13 phosphines with a %$V_{bur} < 75$, and the other includes the 4 largest phosphines (%$V_{bur} > 75$). For the small phosphine set, a slight reweighting of the three descriptors leads to smaller MAEs for training, test, and external data sets (Figure 4D). For the large phosphines, we used a 70/30 training/test split due to the smaller number of data points. We also found that a simpler, two-descriptor model is superior for this subset, with only $\Delta G^{\ddagger}_{OA}$ and %V_{bur} needed to make accurate predictions (Figure 4E). This partition into small and large phosphine sets is consistent with prior studies that describe a change in mechanism, where oxidative addition occurs from either bis(phosphine) or mono(phosphine)

Pd(0) intermediates.[39,40] Notably, our predicted $\Delta G^{\ddagger}_{OA}$ is effective in both cases. This case study not only demonstrates how $\Delta G^{\ddagger}_{OA}$ can be used to predict how substitution patterns on the electrophile will affect the kinetic behaviour of a coupling reaction, but also how it can be used in concert with catalyst-based descriptors to develop a holistic picture of a reaction system.
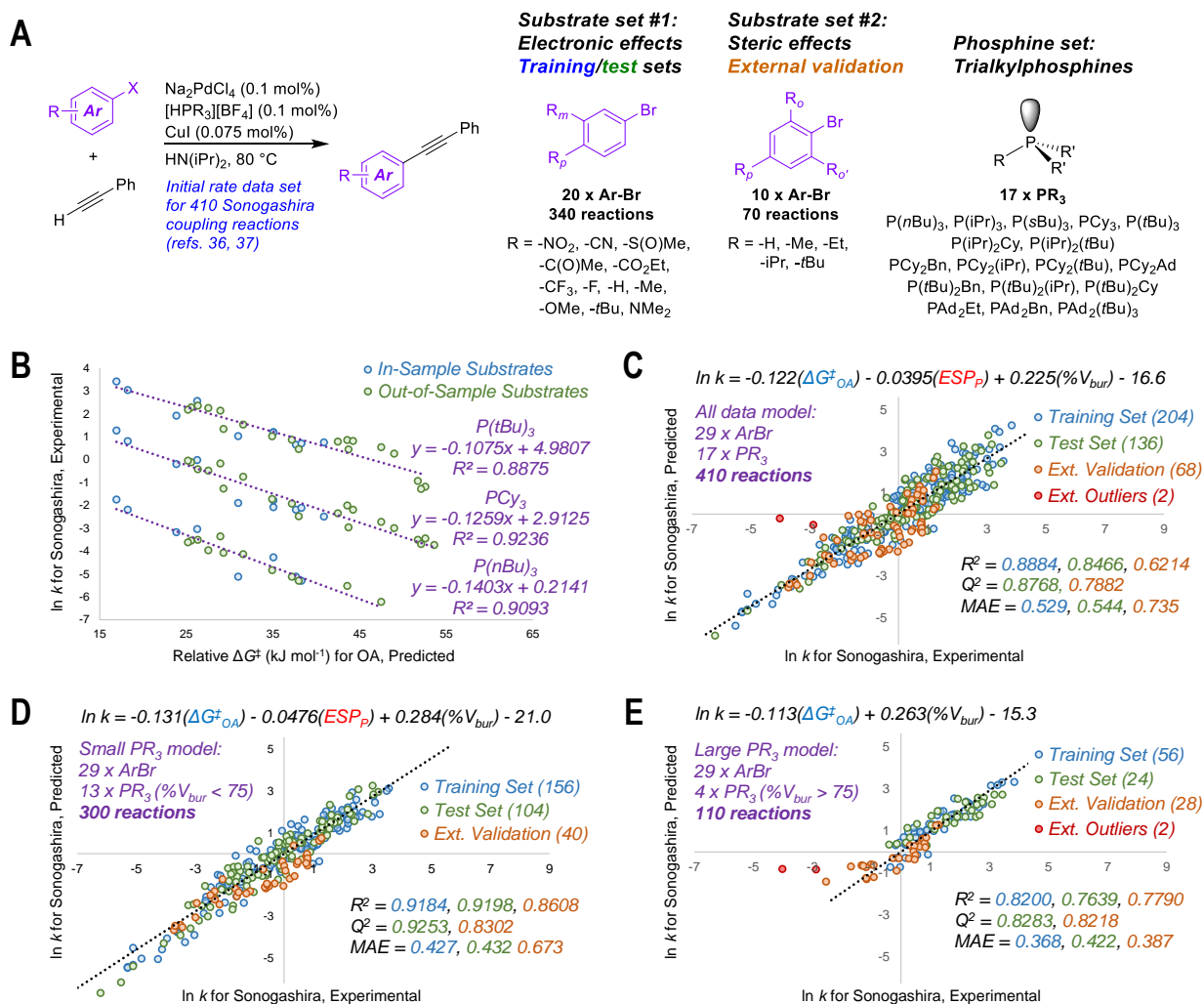

Figure 4. Translating oxidative addition predictions to quantitative models of catalytic reactivity. (**A**) General reaction scheme and chemical space explored for 410 Sonogashira reactions, with two distinct substrate sets; initial rates determined previously.[36,37] (**B**) Univariate linear correlations between predicted $\Delta G^{\ddagger}_{OA}$ for oxidative addition to Pd(PCy$_3$)$_2$ and ln k for Sonogashira coupling with three phosphines; out-of-model substrates are Ar–Br molecules not included in $\Delta G^{\ddagger}_{OA}$ training set. (**C**) Unified three-descriptor model for predicting ln k for the entire set of 410 reactions (29 substrates, 17 ligands), with data partitioned into training (60% of set #1), test (40% of set #1), and external validation (set #2). (**D**) Subset of the model with 13 "small" phosphines (%V$_{bur}$ < 75). (**E**) Subset of the model with 4 "large" phosphines (%V$_{bur}$ > 75).

The second case study tests the use of $\Delta G^{\ddagger}_{OA}$ in predicting site selectivity for cross-coupling when multiple reactive positions are present (Figure 5). Previous approaches to this problem have involved spectroscopic descriptors,[41] the distortion-interaction transition state model,[42] and qualitative arguments based on empirical observations.[19,43] This latter method is most suitable for synthetic planning efforts, but it is largely applied "by analogy" to known systems, and it is not quantitative. Thus, rapid but accurate predictions of the likely coupling site involving a structurally-complex multihalogenated substrate is an unsolved challenge.

Because all the descriptors in the $\Delta G^{\ddagger}_{OA}$ prediction model are local rather than global, one can calculate distinct $\Delta G^{\ddagger}_{OA}$ values for each reactive site in a molecule, with the predicted major site corresponding to the lowest $\Delta G^{\ddagger}_{OA}$. Assessing a series of multihalogenated heterocycles with reported experimental selectivities[43] reveals that these $\Delta G^{\ddagger}_{OA}$ predictions correctly identify the major site of reactivity across a diverse range of substrates, including many heterocycle classes (isoquinolines, diazines, 5-membered rings) not included in our initial training set (Figure 5A). These predictions also correctly identify when a C–Cl position is more reactive than a C–Br position in the same molecule, and vice versa. These predictions are also quantitative. A large difference between the $\Delta G^{\ddagger}_{OA}$ for each site ($\Delta\Delta G^{\ddagger}_{OA}$) indicates very high selectivity for one site over another, while a small $\Delta\Delta G^{\ddagger}_{OA}$ indicates likely poor selectivity. Two specific examples of this are given in Figure 5B. For Suzuki-Miyaura coupling of methyl 2,6-dichloronicotinate, our model predicts $\Delta\Delta G^{\ddagger}_{OA}$ = 5.6 kJ mol^{-1}, favoring C6 by ~7:1 at 65 °C. The observed selectivity using Pd(PPh$_3$)$_4$ as a catalyst at 65 °C is 5:1 C6 to C3 ($\Delta\Delta G^{\ddagger}$ = 4.5 kJ mol^{-1}).[44] For 3,6-dichloro-4-methoxypyridazine, our model predicts $\Delta\Delta G^{\ddagger}_{OA}$ = 1.0 kJ mol^{-1}, favoring C6 by only ~3:1 at room temperature (though this difference is smaller than the model MAE). The observed selectivity using Pd(PPh$_3$)$_4$ as a catalyst at 100 °C is 3:1 C6 to C3, ($\Delta\Delta G^{\ddagger}$ = 3.4 kJ mol^{-1}).[45]

Site selectivity is known to be influenced by reaction conditions, such as catalyst and solvent.[19,43,46] Our prediction model is based on a simple monodentate phosphine using non-polar reaction solvents, which represents a very typical combination used in synthetic applications. To illustrate how our model could guide the development of chemo/regiodivergent coupling reactions, we have applied predictions to substrates known to have tunable selectivity (Figure 5C). First, 3,5-dichloropyridazine is predicted to have good selectivity for C3 over C5 ($\Delta\Delta G^{\ddagger}_{OA}$ = 10.6 kJ mol^{-1}), consistent with the observed selectivity using simple ligands (PPh$_3$, dppf); therefore, inverting this selectivity should require extensive catalyst/solvent screening. Researchers at Merck took this exact approach, discovering that the large QPhos ligand gives high C5 selectivity.[47] Second, the predicted $\Delta\Delta G^{\ddagger}_{OA}$ for 2-chloro-5-bromopyridine is 0.2 kJ mol^{-1}, indicating effectively no selectivity; we confirmed this by oxidative addition of substrate with Pd(PCy$_3$)$_2$, which gives a 1:1 mixture of C2 and C5 products. Thus, our prediction indicates achieving selectivity one way or the other will require more complex systems. Consistent with this, two previous studies reveal ligand-controlled reactivity at either C2 or C5, with high C2 selectivity requiring extensive screening and catalyst parameterization, and use of an unconventional diaminophosphine (dmapf).[48,49] Finally, we predict that 2,4-dibromopyridine should be selective for C2 ($\Delta\Delta G^{\ddagger}_{OA}$ = 8.1 kJ mol^{-1}, C2:C4 ~ 26:1 at room temperature), which is exactly what is observed experimentally using homogeneous Pd catalysis.[50] A recent report revealed that heterogeneous, nanoparticle-based Pd catalysis is capable of inverting the selectivity to ~1:7 C2:C4.[51] As for the examples above, overriding the predicted selectivity requires a dramatically different catalyst system.
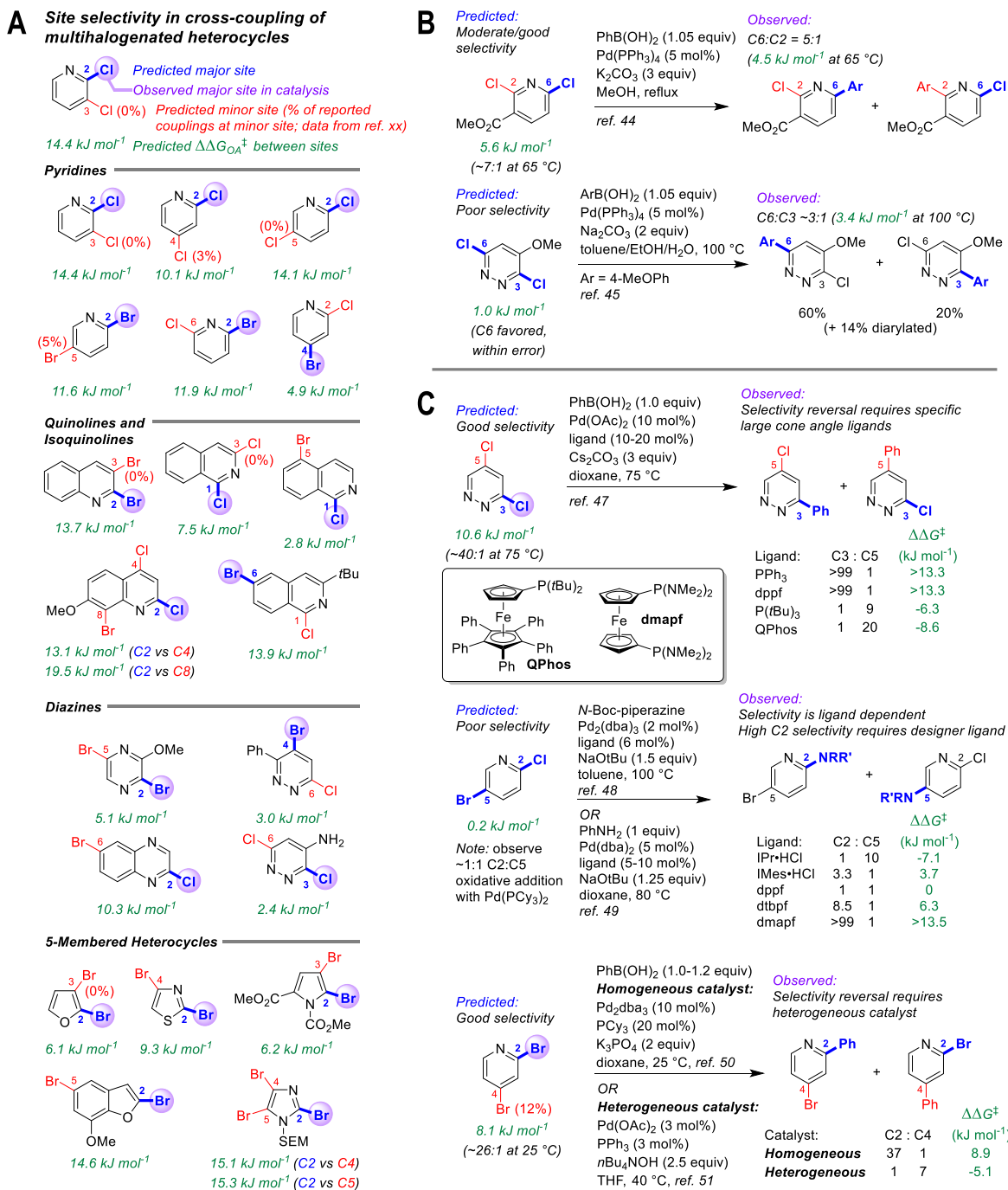
Figure 5. Quantitative site selectivity predictions for cross-coupling reactions. (**A**) Predicted and observed selectivities for multihalogenated heterocycles undergoing Suzuki-Miyaura cross-coupling reactions. (**B**) Selectivity predictions for dihalogenated heterocycles with small $\Delta\Delta G^{\ddagger}_{OA}$ between the two sites, and observed product ratios consistent with predictions. (**C**) Predictions for substrates with observed tunable selectivity, demonstrating that "simple" catalysts are quantitatively consistent with predicted selectivities; overriding predicted reactivity requires targeted screening and/or catalyst design.

As a final case study in synthetic planning, we have retrospectively applied $\Delta G^{\ddagger}_{OA}$ predictions to two reported synthesis toward Dragmacidin D (Figure 6).[52–54] The general sequence relies on two regioselective cross-coupling reactions to a dihalogenated pyrazine core (Figure 6A). Among the key design questions are: how to achieve selective, sequential couplings; and how to ensure compatibility with the existing Ar–Br in indole 1 (which is present in the natural product).[54] Two approaches to this problem have been reported. En route to the first completed synthesis of Dragmacidin D, Garg, Sarpong, and Stoltz (Figure 2B)[52] conducted model studies using 2-iodo-5-chloro-3-methoxypyrazine to maximize site-selectivity, though they observed that the nature of the protecting group on the indole also has a significant effect on reactivity. With an *N*-Ts protecting group, bis(arylation) is readily achieved at elevated temperature, whereas with *N*-TIPS, the intermediate pyrazine is deactivated, preventing a second coupling. The remote electronic effect of the *N*-Ts group, which activates the C5 pyrazine position, is described as a fortuitous discovery. Applying our $\Delta G^{\ddagger}_{OA}$ predictions to these intermediates clearly reveals not only the existence but the magnitude of this effect, leading to ~10-fold higher predicted reactivity between *N*-Ts and *N*-TIPS substrates. Thus, a subtle but important reactivity difference could be anticipated based on predicted $\Delta G^{\ddagger}_{OA}$ prior to experimental work, and help guide protecting group selection.

The second coupling required a switch to the 2-iodo-5-bromo-3-methoxypyrazine to maximize site-selectivity, which again could be predicted based on our model. The $\Delta\Delta G^{\ddagger}_{OA}$ for the two Ar–Br positions in the pyrazine-indole intermediate used in the synthesis is 22.9 kJ mol^{-1}, consistent with the observed high selectivity. Hypothetical use of the corresponding 5-chloropyrazine intermediate gives a predicted $\Delta\Delta G^{\ddagger}_{OA}$ of <5 kJ mol^{-1} (~6:1 at 50 °C), which while still selective for the desired position, would likely lead to overreaction and yield loss. Again, this type of prediction could help to guide synthetic design and subsequent experimental investigations.

Yang, Liu, and Jiang reported a similar approach (Figure 6C)[53] that employed 2,5-dibromo-3-methoxypyrazine as the pyrazine building block; a series of model studies established that regioselective C2-coupling is possible. Despite the fact that both sites appear very similar, our model predicts that C2 is the major site, albeit with moderate selectivity ($\Delta\Delta G^{\ddagger}_{OA}$ = 5.1 kJ mol^{-1}, 6:1 at 80 °C); this lower selectivity may be reflected in the lower isolated yield relative to the differential halogenation strategy. In the subsequent Stille coupling, Yang *et al.* use the unprotected indole derivative, which we predict is ~10-fold less reactive than the corresponding *N*-Ts substrate; nevertheless, the desired reaction site is heavily favored relative to the other two Ar–Br positions, consistent with the observed chemo and regioselectivity.
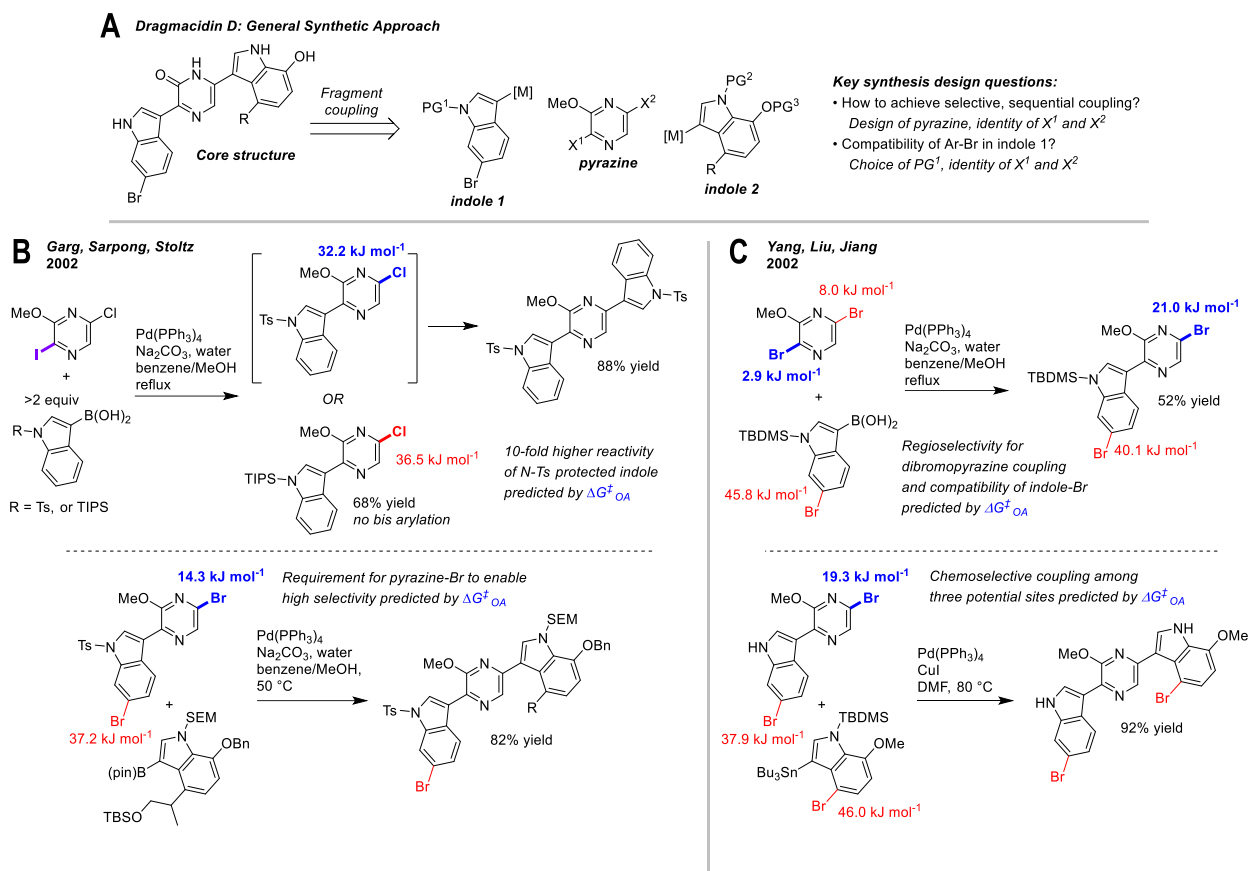
Figure 6. Retrospective analysis of applying oxidative addition predictions to synthesis design: Dragmacidin D. (**A**) Abbreviated retrosynthesis of core structure, involving selective fragment coupling to a dihalogenated pyrazine. (**B**) Approach involving differential halogenation, tosylate protecting group on indole 1, and switch from 5-chloro to 5-bromopyrazine intermediates for selective coupling.[52] (**C**) Approach involving regioselective coupling to dibromopyrazine, TBS protecting group on indole 1, and chemoselective Stille coupling in the presence of two potentially competitive Ar–Br positions.[53]

Conclusions

We have demonstrated that a quantitative structure-reactivity model for oxidative addition, a key step in many catalytic mechanisms, enables accurate predictions for the outcome of various cross-coupling reactions. By correlating relative reaction rates with easily-obtained molecular descriptors, we have mapped the reactivity of diverse (hetero)aryl electrophiles toward oxidative addition to Pd(0). This reactivity map links molecular structure to predicted $\Delta G^{\ddagger}_{OA}$ for any hypothetical substrate, with applicability well beyond the oxidative addition training set. We used these $\Delta G^{\ddagger}_{OA}$ values to predict rates and selectivities for many different catalytic reactions under various conditions, including Sonogashira, Suzuki, Buchwald-Hartwig, and Stille couplings. Given the importance of these reactions in the synthesis of complex organic molecules, we anticipate that quantitative reactivity predictions could be used during synthetic planning to design substrates with high intrinsic selectivity, and/or to identify where achieving the desired selectivity

is likely to be more challenging/resource intensive. Applying $\Delta G^{\ddagger}_{OA}$ predictions to hypothetical synthetic sequences and potential intermediates could therefore be used to design more selective routes and/or prioritize different potential routes prior to commencing experimental investigations. It could also be used to identify where extensive reaction development is necessary (e.g. overriding predicted selectivity), or where "simple" systems are likely to be successful. As we expand this oxidative addition reactivity map to incorporate additional sets of reaction conditions and catalysts, we expect it will not only shed new light on the mechanistic aspects of cross-coupling, but also find widespread use in refining and augmenting computer-aided synthesis design and automated reaction discovery/optimization. We are also exploring quantitative reactivity models based on key steps in other catalytic mechanisms to generate accurate and general predictions across the synthesis landscape.

References:

(1)   Campeau, L.-C.; Hazari, N. Cross-Coupling and Related Reactions: Connecting Past Success to the Development of New Reactions for the Future. *Organometallics* **2019**, *38*, 3–35. https://doi.org/10.1021/acs.organomet.8b00720.

(2)   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.

(3)   Vinogradova, E. V.; Zhang, C.; Spokoyny, A. M.; Pentelute, B. L.; Buchwald, S. L. Organometallic Palladium Reagents for Cysteine Bioconjugation. *Nature* **2015**, *526*, 687–691. https://doi.org/10.1038/nature15739.

(4)   Wang, K.; Osuka, A.; Song, J. Pd-Catalyzed Cross Coupling Strategy for Functional Porphyrin Arrays. *ACS Cent. Sci.* **2020**, *6*, 2159–2178. https://doi.org/10.1021/acscentsci.0c01300.

(5)   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*, 49–53. https://doi.org/10.1126/science.1259203.

(6)   Shevlin, M. Practical High-Throughput Experimentation for Chemists. *ACS Med. Chem. Lett.* **2017**, *8*, 601–607. https://doi.org/10.1021/acsmedchemlett.7b00165.

(7)   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.

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

(9)   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.

(10)   Gallegos, L. C.; Luchini, G.; St. John, P. C.; Kim, S.; Paton, R. S. Importance of Engineered and Learned Molecular Representations in Predicting Organic Reactivity, Selectivity, and Chemical Properties. *Acc. Chem. Res.* **2021**, *54*, 827–836. https://doi.org/10.1021/acs.accounts.0c00745.

(11)   Gensch, T.; Gomes, G. dos P.; Friederich, P.; Peters, E.; Gaudin, T.; Pollice, R.; Jorner, K.; Nigam, A.; D'Addario, M. L.; Sigman, M. S.; Aspuru-Guzik, A. A Comprehensive Discovery Platform for Organophosphorus Ligands for Catalysis. *ChemRxiv* **2021**. https://doi.org/10.26434/chemrxiv.12996665.v1.

(12)   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.

(13)   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.

(14) Zahrt, A. F.; Henle, J. J.; Rose, B. T.; Wang, Y.; Darrow, W. T.; Denmark, S. E. Prediction of Higher-Selectivity Catalysts by Computer-Driven Workflow and Machine Learning. *Science* **2019**, *363*, eaau5631. https://doi.org/10.1126/science.aau5631.

(15) Toyao, T.; Maeno, Z.; Takakusagi, S.; Kamachi, T.; Takigawa, I.; Shimizu, K. Machine Learning for Catalysis Informatics: Recent Applications and Prospects. *ACS Catal.* **2020**, *10*, 2260–2297. https://doi.org/10.1021/acscatal.9b04186.

(16) Sandfort, F.; Strieth-Kalthoff, F.; Kühnemund, M.; Beecks, C.; Glorius, F. A Structure-Based Platform for Predicting Chemical Reactivity. *Chem* **2020**, *6*, 1379–1390. https://doi.org/10.1016/j.chempr.2020.02.017.

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

(18) Labinger, J. A. Tutorial on Oxidative Addition. *Organometallics* **2015**, *34*, 4784–4795. https://doi.org/10.1021/acs.organomet.5b00565.

(19) 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.

(20) Sandford, C.; Fries, L. R.; Ball, T. E.; Minteer, S. D.; Sigman, M. S. Mechanistic Studies into the Oxidative Addition of Co(I) Complexes: Combining Electroanalytical Techniques with Parameterization. *J. Am. Chem. Soc.* **2019**, *141*, 18877–18889. https://doi.org/10.1021/jacs.9b10771.

(21) Senn, H. M.; Ziegler, T. Oxidative Addition of Aryl Halides to Palladium(0) Complexes: A Density-Functional Study Including Solvation. *Organometallics* **2004**, *23*, 2980–2988. https://doi.org/10.1021/om049963n.

(22) Ahlquist, M.; Norrby, P.-O. Oxidative Addition of Aryl Chlorides to Monoligated Palladium(0): A DFT-SCRF Study. *Organometallics* **2007**, *26*, 550–553. https://doi.org/10.1021/om0604932.

(23) McMullin, C. L.; Jover, J.; Harvey, J. N.; Fey, N. Accurate Modelling of Pd(0) + PhX Oxidative Addition Kinetics. *Dalton Trans.* **2010**, *39*, 10833–10836. https://doi.org/10.1039/C0DT00778A.

(24) Kashihara, M.; Gordon, C. P.; Copéret, C. Reactivity of Substituted Benzenes toward Oxidative Addition Relates to NMR Chemical Shift of the Ipso-Carbon. *Org. Lett.* **2020**, *22*, 8910–8915. https://doi.org/10.1021/acs.orglett.0c03300.

(25) Amatore, C.; Jutand, A.; Khalil, F.; M'Barki, M. A.; Mottier, L. Rates and Mechanisms of Oxidative Addition to Zerovalent Palladium Complexes Generated in Situ from Mixtures of $Pd^0(dba)_2$ and Triphenylphosphine. *Organometallics* **1993**, *12*, 3168–3178. https://doi.org/10.1021/om00032a045.

(26) Portnoy, M.; Milstein, D. Mechanism of Aryl Chloride Oxidative Addition to Chelated Palladium(0) Complexes. *Organometallics* **1993**, *12*, 1665–1673. https://doi.org/10.1021/om00029a026.

(27) Barrios-Landeros, F.; Carrow, B. P.; Hartwig, J. F. Effect of Ligand Steric Properties and Halide Identity on the Mechanism for Oxidative Addition of Haloarenes to Trialkylphosphine Pd(0) Complexes. *J. Am. Chem. Soc.* **2009**, *131*, 8141–8154. https://doi.org/10.1021/ja900798s.

(28) Mitchell, E. A.; Jessop, P. G.; Baird, M. C. A Kinetics Study of the Oxidative Addition of Bromobenzene to $Pd(PCy_3)_2$ (Cy = Cyclohexyl) in a Nonpolar Medium: The Influence on Rates of Added PCy_3 and Bromide Ion. *Organometallics* **2009**, *28*, 6732–6738. https://doi.org/10.1021/om900679w.

(29) 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 SNAr-Type Mechanism. *Chem. Eur. J.* **2015**, *21*, 7858–7865. https://doi.org/10.1002/chem.201406210.

(30) A. Grimmel, S.; Reiher, M. The Electrostatic Potential as a Descriptor for the Protonation Propensity in Automated Exploration of Reaction Mechanisms. *Faraday Discussions* **2019**, *220*, 443–463. https://doi.org/10.1039/C9FD00061E.

(31) 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.

(32) 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.

(33) 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.

(34) Consonni, V.; Ballabio, D.; Todeschini, R. Comments on the Definition of the Q^2 Parameter for QSAR Validation. *J. Chem. Inf. Model.* **2009**, *49*, 1669–1678. https://doi.org/10.1021/ci900115y.

(35)   Paenurk, E.; Kaupmees, K.; Himmel, D.; Kütt, A.; Kaljurand, I.; Koppel, I. A.; Krossing, I.; Leito, I. A Unified View to Brønsted Acidity Scales: Do We Need Solvated Protons? *Chem. Sci.* **2017**, *8*, 6964–6973. https://doi.org/10.1039/C7SC01424D.

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

(37)   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.

(38)   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.

(39)   Schoenebeck, F.; Houk, K. N. Ligand-Controlled Regioselectivity in Palladium-Catalyzed Cross Coupling Reactions. *J. Am. Chem. Soc.* **2010**, *132*, 2496–2497. https://doi.org/10.1021/ja9077528.

(40)   Niemeyer, Z. L.; Milo, A.; Hickey, D. P.; Sigman, M. S. Parameterization of Phosphine Ligands Reveals Mechanistic Pathways and Predicts Reaction Outcomes. *Nature Chemistry* **2016**, *8*, 610–617. https://doi.org/10.1038/nchem.2501.

(41)   Handy, S. T.; Zhang, Y. A Simple Guide for Predicting Regioselectivity in the Coupling of Polyhaloheteroaromatics. *Chem. Commun.* **2006**, 299–301. https://doi.org/10.1039/B512948F.

(42)   Legault, C. Y.; Garcia, Y.; Merlic, C. A.; Houk, K. N. Origin of Regioselectivity in Palladium-Catalyzed Cross-Coupling Reactions of Polyhalogenated Heterocycles. *J. Am. Chem. Soc.* **2007**, *129*, 12664–12665. https://doi.org/10.1021/ja075785o.

(43)   Almond-Thynne, J.; C. Blakemore, D.; C. Pryde, D.; C. Spivey, A. 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.

(44)   Yang, W.; Wang, Y.; Corte, J. R. Efficient Synthesis of 2-Aryl-6-Chloronicotinamides via PXPd2-Catalyzed Regioselective Suzuki Coupling. *Org. Lett.* **2003**, *5*, 3131–3134. https://doi.org/10.1021/ol035188g.

(45)   Blaise, E.; Kümmerle, A. E.; Hammoud, H.; de Araújo-Júnior, J. X.; Bihel, F.; Bourguignon, J.-J.; Schmitt, M. Access to 4-Alkylaminopyridazine Derivatives via Nitrogen-Assisted Regioselective Pd-Catalyzed Reactions. *J. Org. Chem.* **2014**, *79*, 10311–10322. https://doi.org/10.1021/jo501930s.

(46)   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.

(47)   Dai, X.; Chen, Y.; Garrell, S.; Liu, H.; Zhang, L.-K.; Palani, A.; Hughes, G.; Nargund, R. Ligand-Dependent Site-Selective Suzuki Cross-Coupling of 3,5-Dichloropyridazines. *J. Org. Chem.* **2013**, *78*, 7758–7763. https://doi.org/10.1021/jo401096u.

(48)   Ji, J.; Li, T.; Bunnelle, W. H. Selective Amination of Polyhalopyridines Catalyzed by a Palladium−Xantphos Complex. *Org. Lett.* **2003**, *5*, 4611–4614. https://doi.org/10.1021/ol0357696.

(49)   Keylor, M. H.; Niemeyer, Z. L.; Sigman, M. S.; Tan, K. L. Inverting Conventional Chemoselectivity in Pd-Catalyzed Amine Arylations with Multiply Halogenated Pyridines. *J. Am. Chem. Soc.* **2017**, *139*, 10613–10616. https://doi.org/10.1021/jacs.7b05409.

(50)   Sicre, C.; Alonso-Gómez, J.-L.; Cid, M. M. Regioselectivity in Alkenyl(Aryl)-Heteroaryl Suzuki Cross-Coupling Reactions of 2,4-Dibromopyridine. A Synthetic and Mechanistic Study. *Tetrahedron* **2006**, *62*, 11063–11072. https://doi.org/10.1016/j.tet.2006.09.040.

(51)   Scott, N. W. J.; Ford, M. J.; Jeddi, N.; Eyles, A.; Simon, L.; Whitwood, A. C.; Tanner, T.; Willans, C. E.; Fairlamb, I. J. S. A Dichotomy in Cross-Coupling Site Selectivity in a Dihalogenated Heteroarene: Influence of Mononuclear Pd, Pd Clusters, and Pd Nanoparticles—the Case for Exploiting Pd Catalyst Speciation. *J. Am. Chem. Soc.* **2021**, *143*, 9682–9693. https://doi.org/10.1021/jacs.1c05294.

(52)   Garg, N. K.; Sarpong, R.; Stoltz, B. M. The First Total Synthesis of Dragmacidin D. *J. Am. Chem. Soc.* **2002**, *124*, 13179–13184. https://doi.org/10.1021/ja027822b.

(53)   Yang, C.-G.; Liu, G.; Jiang, B. Preparing Functional Bis(Indole) Pyrazine by Stepwise Cross-Coupling Reactions: An Efficient Method to Construct the Skeleton of Dragmacidin D. *J. Org. Chem.* **2002**, *67*, 9392–9396. https://doi.org/10.1021/jo026450m.

(54)   Garg, N. K.; Stoltz, B. M. A Unified Synthetic Approach to the Pyrazinone Dragmacidins. *Chem. Commun.* **2006**, 3769–3779. https://doi.org/10.1039/B605929E.

Author contributions: *Jingru Lu*: Conceptualization, Methodology, Investigation, Validation, Formal analysis, Writing. *Sofia Donnecke*: Investigation, Formal analysis, Writing. *Irina Paci*: Conceptualization, Methodology, Supervision, Writing. *David C. Leitch*: Conceptualization, Methodology, Formal analysis, Supervision, Writing.

Competing interests: Authors declare no competing interests.

Data and materials availability: All data is available in the main text or the supplementary information files.