Title: Linking mechanistic analysis of catalytic reactivity cliffs to ligand classification

Authors: Samuel H. Newman-Stonebraker[1†], Sleight R. Smith[2†], Julia E. Borowski[1], Ellyn Peters[2], Tobias Gensch[2], Heather C. Johnson[3], Matthew S. Sigman[2]*, Abigail G. Doyle[1]*

Affiliations:

[1]Department of Chemistry, Princeton University, Princeton, NJ 08544.

[2]Department of Chemistry, University of Utah, Salt Lake City, UT 84112.

[3]Process Research and Development, Merck & Co., Inc., Rahway, NJ 07065

*Corresponding author. Email: sigman@chem.utah.edu (M.S.S.), agdoyle@princeton.edu (A.G.D.)

[†]*These authors contributed equally*

Abstract: Statistical analysis of reaction data with molecular descriptors can enable chemists to identify reactivity cliffs that result from a mechanistic dependence on a specific structural feature. In this study, we develop a broadly applicable and quantitative classification workflow that identifies reactivity cliffs in eleven Ni- and Pd-catalyzed cross-coupling datasets employing monodentate phosphine ligands. A unique ligand steric descriptor, $\%V_{\text{bur}}$ (*min*), is found to divide these datasets into active and inactive regions at a similar threshold value. Organometallic studies demonstrate that this threshold corresponds to the binary outcome of bisligated versus monoligated metal and that $\%V_{\text{bur}}$ (*min*) is a physically meaningful and predictive representation of ligand structure in catalysis. Taken together, we expect that this strategy will be of broad value in mechanistic investigation of structure-reactivity relationships, while providing a means to rationally partition datasets for data-driven modeling.

Main text:

For almost 100 years, chemists have used data-driven modeling to understand the many complex relationships connecting chemical structure to reactivity. The paradigmatic example is the

29   Hammett equation (*1*), a linear free energy relationship (LFER) between an experimental

30   observable and a molecular descriptor (*2*). LFERs have been used extensively to understand

31   reaction mechanism (*3, 4*) and, in certain cases, to predict reaction outcomes (*5*). While many

32   reactions are continuously dependent on a single descriptor, as captured by LFERs, some processes

33   may exhibit reactivity cliffs, wherein a criterion or threshold value of a given feature must be met

34   in order for the reaction to occur (i.e., a binary response) (*6*). If a particular molecular feature can

35   define a reactivity cliff and classify molecules by structure, this could be used to predict reaction

36   performance of unseen examples while also resulting in the development of key mechanistic

37   insights (**Fig. 1A**). Thus, the identification and understanding of these reactivity cliffs is of great

38   interest, as the ability to bifurcate datasets—and the molecules within them—on the basis of

39   distinct mechanistic outcomes would augment expert chemists' intuition and simplify subsequent

40   data-driven modeling.

41   Transition metal-catalyzed cross coupling represents an interesting case study for

42   identifying reactivity cliffs. This reaction class is synthetically important due to its widespread use

43   in the construction of pharmaceuticals (*7*) and materials (*8*). In addition, the success of these

44   reactions is highly dependent on ligand identity, for which monodentate phosphines are ubiquitous

45   (*9, 10*). As a result, chemists have developed numerous tools to quantitatively describe the diverse

46   steric and electronic properties of these ligands—for which thousands of unique structures exist—

47   to uncover structure-reactivity relationships (SRRs) in experimental data (*11*) (**Fig. 1B**). These

48   descriptors, including Tolman Cone Angle (*12, 13*), Solid Angle (*14*), Sterimol (*15*), and percent

49   buried volume (%V_{bur}) (*16, 17*), have been used successfully in linear modeling of cross-coupling

50   reactions (*5, 18*) (**Fig. 1C**). However, there are situations in which seemingly similar ligands afford

substantially different responses (*19*), suggesting the presence of discontinuous trends in ligand activity (*20*).

Considering this, we hypothesized that reactivity cliffs were embedded in a phosphine structural feature that had yet to be recognized. Identification of such a feature would not only allow for the development of a statistical tool to classify active and inactive ligands but also provide a quantitative handle for mechanistically rationalizing ligand performance. Herein we provide a workflow and analysis to achieve these goals. Using the organophosphorus(III) descriptor library recently developed by Sigman and Aspuru-Guzik (*Kraken*) that possesses over 190 conformationally representative descriptors for each of several thousand monodentate phosphines (*21*) (**Fig. 1D**), we were able to classify eleven Ni and Pd catalytic case studies into mechanistically bifurcated regions of reactivity based on catalyst ligation state, consistent with spectroscopic/crystallographic organometallic studies. We found that minimum percent buried volume ($\%V_{bur}$ (*min*))—a previously unexplored descriptor that quantifies the proximal steric bulk of the smallest energetically accessible conformation of a given ligand—is the single descriptor able to achieve this classification. Ultimately, we demonstrate that this analysis was able to reveal non-intuitive trends in organometallic chemistry and thereby serves as an important mechanistic tool, complementary to LFERs (*18, 19*), to understand and predict monodentate phosphine SRRs and catalyst speciation in cross-coupling catalysis (**Fig. 1E**).
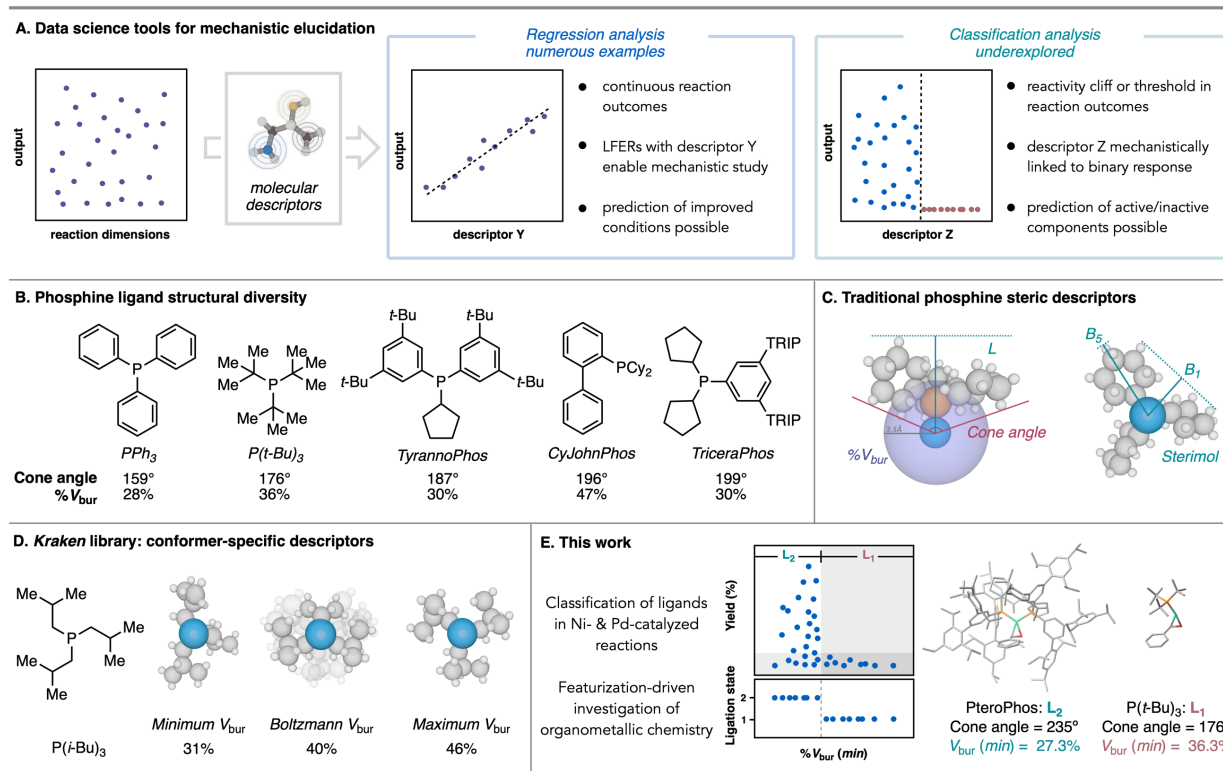
Fig. 1. Introduction. (**A**) Tools to understand trends in chemical reactivity using mechanistically relevant molecular descriptors. (**B**) Examples of monodentate phosphines used in Ni and Pd cross-coupling reactions, including TyrannoPhos and TriceraPhos (DinoPhos ligands) recently reported by the Doyle lab. TRIP = 2,4,6-triisopropylphenyl. (**C**) Commonly employed methods of quantifying phosphine steric properties. (**D**) Phosphine descriptor library (*Kraken*) capturing multiple ligand conformers. (**E**) This work.

Results and Discussion

Exploration of phosphine steric features in Ni-catalysis.

Our initial platform for probing the presence of ligand reactivity cliffs was inspired by a recent study from the Doyle group that identified a new class of phosphine ligands—the DinoPhos ligands (TriceraPhos and TyrannoPhos (**Fig. 1B**))—for Ni-catalyzed cross-coupling of acetals with aryl boroxines (*19*). Having both large cone angles and small %V_{bur} values, these ligands are distinct from other phosphines that often have a 1:1 relationship between these features (*22*). The large cone angles suggested that the DinoPhos ligands may behave like the sterically bulky phosphines developed for Pd catalysis (e.g., P(*t*-Bu)₃ or BrettPhos), where formation of

84   monoligated Pd (L_1 complexes) has been shown to be critical for reaction success (*23-27*).

85   However, the small $\%V_{bur}$ characterized these ligands as similar to smaller structures (e.g., PPh₃)

86   that tend to form bisligating (denoted as L_2) complexes. The observation that this unique steric

87   profile was essential for reaction success in Ni catalysis raised questions about its impact on

88   speciation and prompted us to use Ni-catalyzed cross-coupling reactions as a case study for the

89   identification of ligand reactivity cliffs.

90   The original study by the Doyle lab included 19 phosphines, which we supplemented with

91   15 additional ligands to span the entire range of $\%V_{bur}$ (**I**, Figure 2A). We also collected data on

92   more traditional Csp^2–Csp^2 Suzuki–Miyaura coupling reactions with aryl chlorides using high-

93   throughput experimentation (HTE) (*28*). The ligand set was selected from a combined Doyle and

94   Merck inventory to encompass 90 monodentate ligands that are evenly distributed throughout the

95   feature space from the Sigman/Aspuru-Guzuk virtual library (*21*) as determined by principal

96   component analysis. These ligands were evaluated for four cross-coupling reactions wherein the

97   aryl halide and aryl boronic acid were altered (**II-V**, Figure 2A). The coupling partners included a

98   range of electronic and steric features that could impact different aspects of the elementary steps

99   of the catalytic cycle. Notably, the DinoPhos ligands were top performers in each of these

100   reactions, along with DrewPhos (*29*), a triarylphosphine ligand that possesses large 3,5-

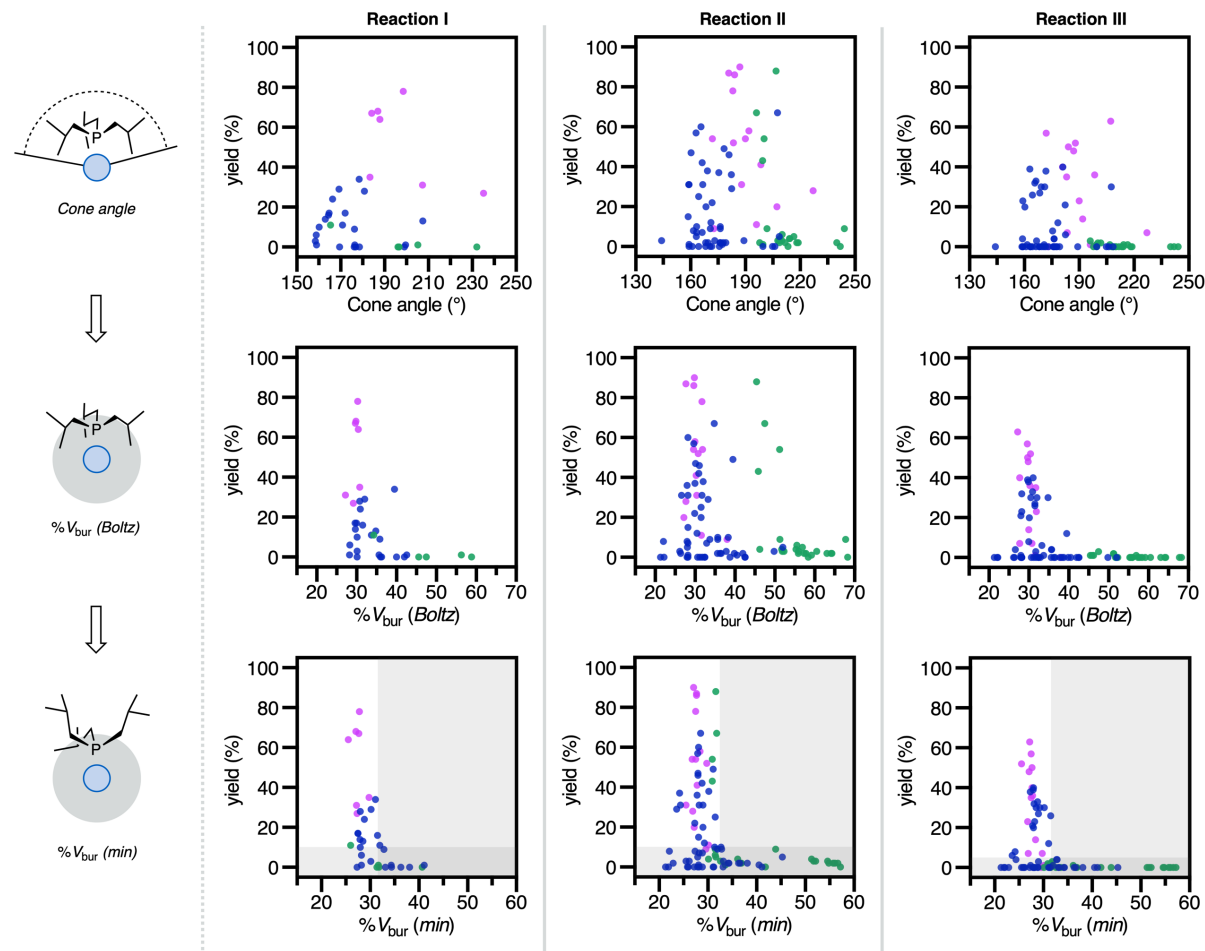101   substituents similar to the DinoPhos ligands.

102   With these data, we sought to determine how the recently computed phosphine steric

103   descriptors relate to reaction performance. For each reaction, the yield was evaluated against

104   representative descriptors in the library; as three examples, the cone angle (*30*), Boltzmann

105   averaged $\%V_{bur}$ of the ligand's conformational ensemble ($\%V_{bur}$ (*Boltz*)), and $\%V_{bur}$ of the library

106   conformer with the smallest buried volume ($\%V_{bur}$ (*min*)) are shown in Figure 2B. While the cone

107   angle descriptors did not provide clear reactivity cutoffs, both $\%V_{bur}$ (*Boltz*) and $\%V_{bur}$ (*min*)

108   allowed for visual classification of ligand performance into "active" and "inactive" bins (details

109   of the classification tool will be described below), though several outliers were present with the

110   former. $\%V_{bur}$ (*min*) revealed sharp cutoffs in reactivity, wherein nearly all ligands above 32%

111   were found to be unreactive. Ligands that were unsuccessful at promoting reactivity below 32%

112   V_{bur} (*min*) were grouped in their own bin, wherein electronic features and catalyst-poisoning

113   functionality (e.g., cyano, carbonyl, and halide groups) appeared to be largely responsible for their

114   inactivity. Notably, the use of $\%V_{bur}$ (*min*) but not $\%V_{bur}$ (*Boltz*) allowed for various highly-flexible

115   scaffolds such as $P(i\text{-Bu})_3$, PBn_3, and small Buchwald-type phosphines (*31, 32*), to be successfully

116   classified. Thus, this steric descriptor was best able to categorize phosphines into active and

117   inactive groups, with a reactivity threshold of ~32% V_{bur} (*min*) appearing to be largely independent

118   of the substrate pairing.

Fig. 2. Investigation of phosphine steric parameters in Ni catalyzed datasets. (**A**) Ni-Catalyzed Suzuki-Miyaura datasets collected with monodentate phosphine ligands, including the DinoPhos ligands. (**B**) Analysis of datasets with steric parameters cone angle, $\%V_{bur}$ Boltz. and $\%V_{bur}$ min.. Pink dots = DinoPhostype ligands, green dots = Buchwald-type ligands, blue dots = all others.

Mechanistic origin of $\%V_{bur}$ (min) reactivity thresholds. The presence of reaction-independent reactivity cliffs relative to $\%V_{bur}$ (*min*) in these data prompted us to investigate the mechanistic basis for this descriptor's significance. Given the importance of speciation in cross-coupling catalysis, we questioned whether this descriptor is predictive of the thermodynamic favorability of

128   L$_2$M vs. L$_1$M speciation. Historically, the development of large ligands—intuitively captured

129   through the cone angle descriptor—is an established principle to encourage L$_1$M and achieve

130   reactivity in Pd catalysis (*23-27*). While TriceraPhos and TyrannoPhos both have cone angle

131   values higher than many phosphines known to form L$_1$ species, %V_{bur} (*min*) categorizes their

132   reactivity with those that form L$_2$ Ni and Pd complexes (*23, 33*). Thus, this interpretation of the

133   observed reactivity cliffs would suggest that steric bulk within the metal's first coordination sphere

134   primarily governs the ligation state outcome, regardless of the overall size of the phosphine.

135   To interrogate this hypothesis, we surveyed a subset (28 phosphines) of the ligands used across

136   the Ni-catalyzed SMC reactions to determine ligation state spectroscopically. L$_n$Ni(4-

137   fluorobenzaldehyde) was chosen as a platform, as its three NMR handles (^{1}H, ^{19}F, and ^{31}P)

138   provided a readout of the *in-situ* ligation state at the metal center upon reaction of Ni(COD)$_2$ with

139   the aldehyde and two equivalents of phosphine (**VI**, Figure 3A) (*34, 35*). Each ligand was assigned

140   as forming L$_2$Ni or L$_1$Ni complexes based on the observed spectra. The results of these experiments

141   were then evaluated with steric features from the library to determine which classified the ligands

142   into regions of similar ligation states (Figure 3B, see SI for additional parameters). As observed

143   with the cross-coupling reaction yields, cone angle was unable to partition the ligands. Indeed, the

144   DinoPhos ligands and DrewPhos all formed L$_2$Ni complexes exclusively, despite having some of

145   the largest cone angle values among ligands evaluated. %V_{bur} (*Boltz*) was able to correctly group

146   the DinoPhos-type ligands with other L$_2$Ni-forming phosphines; however, some flexible ligands

147 (MeJohnPhos, P(*i*-Bu)$_3$, and PBn$_3$) remained

148 misclassified with L$_1$Ni-forming ligands.

149 Similar to the catalytic reactions, %V_{bur} (*min*)

150 resolved these outliers and resulted in a sharp

151 cutoff just below 32% between the L$_2$Ni and

152 L$_1$Ni regimes. Notably, this value closely

153 matched the reactivity thresholds observed in

154 reactions **I-V**.

155   Given this finding, we hypothesized that

156 one could predict the speciation of new ligands,

157 enabling a pre-screening of structures that may

158 be challenging to access synthetically. To test

159 this, we conceptualized a new DinoPhos ligand,

160 PteroPhos, which possesses two aryl groups

161 with 2,4,6-triisopropylphenyl (TRIP)

162 substituents at the 3,5-positions (Figure 4A). Its

163 computed cone angle is 235º, one of the largest

164 among all monodentate phosphines studied.

165 Despite the enormous size of this ligand as

166 ascertained by visual inspection and its cone

167 angle value, its relatively low %V_{bur} (*min*) of

168 only 27.2% suggested that PteroPhos should

169 form L$_2$Ni complexes and be effective in Ni-catalyzed SMCs. Indeed, upon preparation of this
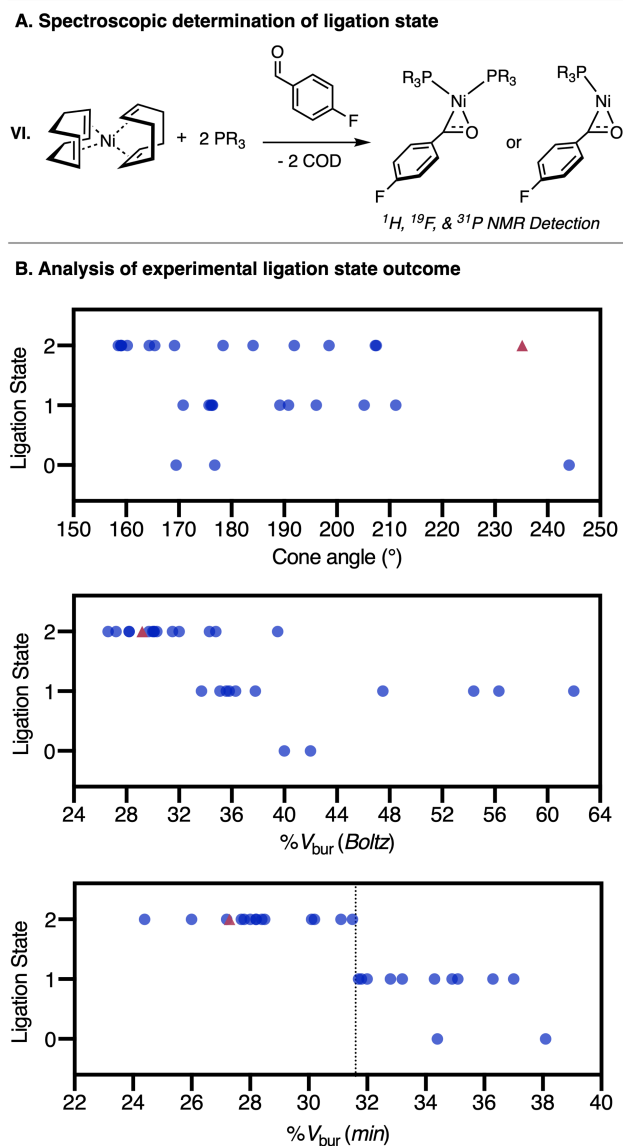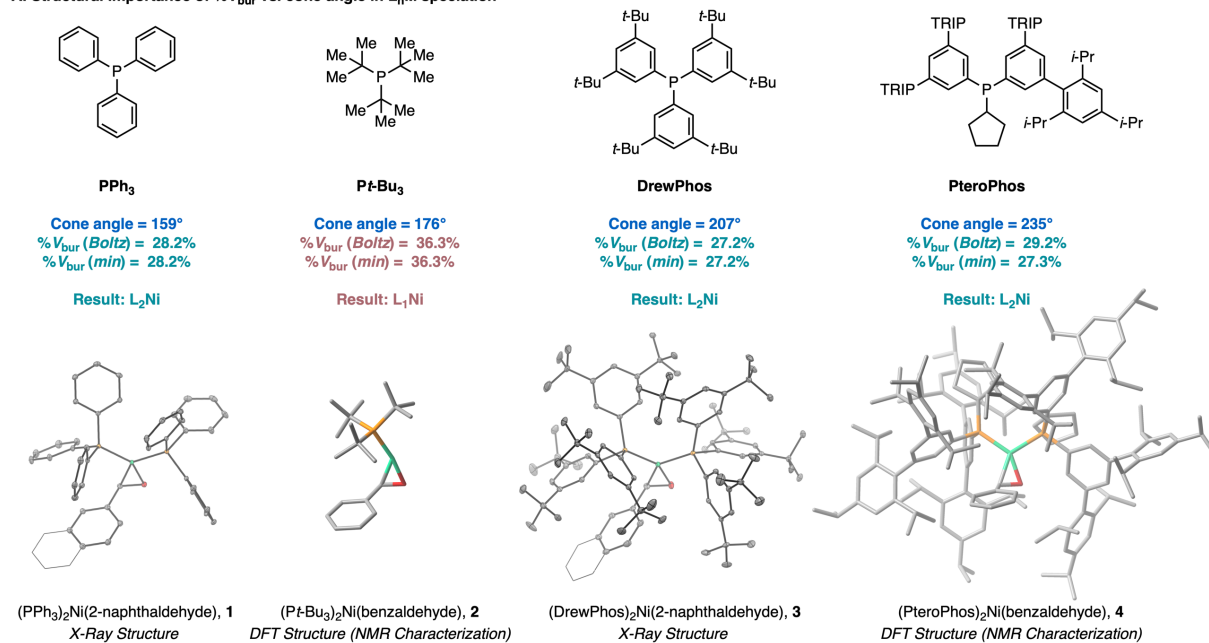

A. Spectroscopic determination of ligation state

B. Analysis of experimental ligation state outcome

Fig. 3. Ligation state studies. (**A**) Reaction of Ni(COD)$_2$, 4-fluorobenzaldehyde, and 2 equivalents of phosphine for spectroscopic determination of ligation state using ^{1}H, ^{19}F, and ^{31}P NMR spectroscopy in C$_6$D$_6$ (25 °C). (**B**) Ligation state experiments plotted against cone angle, %V_{bur} (*Boltz*), and %V_{bur} (*min*). Red triangle represents PteroPhos. (*t*-Bu)BrettPhos (non-ligating) datapoint beyond x-axis range for %V_{bur} (*min*)/(*Boltz*) plots.
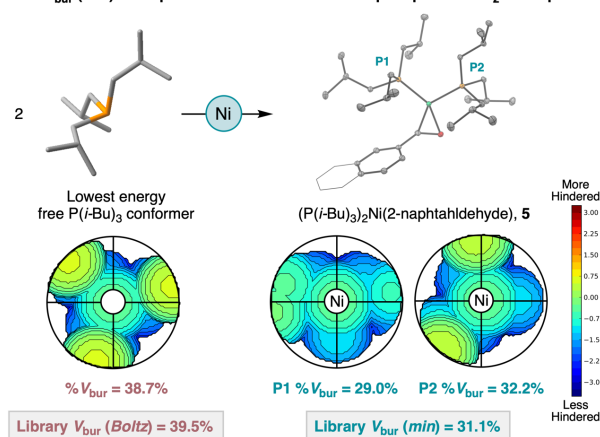
170 ligand, we found that it readily formed a L_2Ni(4-fluorobenzaldehyde) complex (Figure 3B) and

171 was moderately active in Ni-catalyzed reactions **I-V**, consistent with the $\%V_{bur}$ (*min*) classification

172 analysis.

173   We then sought to understand the structural significance of $\%V_{bur}$ (*min*) through

174 crystallographic and computational studies (complexes **1-4**, Figure 4A), first to rationalize why

175 cone angle is not predictive of speciation. While our attempts at crystallizing L_2Ni(benzaldehyde)

176 complexes bearing the DinoPhos ligands were not successful, we were able to obtain X-ray quality

177 crystals of (DrewPhos)$_2$Ni(2-naphthaldehyde) (**3**, Figure 4A). As observed in the solid state, the

178 two phosphines are positioned 116º *cis* to each other on the complex; since DrewPhos possesses a

179 cone angle of 207º, a literal interpretation of cone angle as a description of phosphine size would

180 suggest that there is 90º overlap of two solids, giving rise to considerable steric repulsion or

181 distortion. Yet, when compared with the X-ray structure of (PPh$_3$)$_2$Ni(2-naphthaldehyde) (**1**,

182 Figure 4A), the two complexes exhibit very similar Ni–P bond lengths and P–Ni–P "bite angles",

183 despite DrewPhos having a cone angle 48º larger than PPh$_3$. What cone angle does not capture is

184 the non-uniform topology of the ligand structure distal to the metal. These less dense and more

185 conformationally flexible regions can be arranged in a manner that accommodates a second ligand

186 and/or substrate within the "cone", a recognized limitation that has led to the development of

187 alternative steric parameters (*16, 36*). On this basis, cone angle is unsuccessful at classifying the

188 ligation state and reactivity of the phosphines under study. Indeed, when compared with that of L_1

189 (P(*t*-Bu)$_3$)Ni(benzaldehyde) (**2**) (Figure 4A), the DFT-optimized structure of

190 (PteroPhos)$_2$Ni(benzaldehyde) (**4**) (Figure 4A) provides the most striking example of L_2

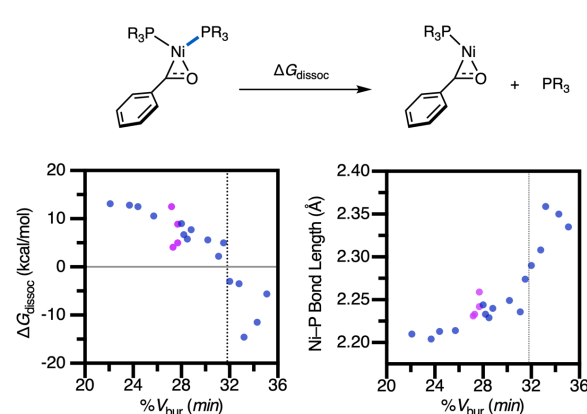191 complexation being agnostic to substantial remote steric bulk.

A. Structural importance of %V_{bur} vs. cone angle in L_nM speciation

PPh₃
Cone angle = 159°
%V_{bur} (*Boltz*) = 28.2%
%V_{bur} (*min*) = 28.2%
Result: L_2Ni

Pt-Bu₃
Cone angle = 176°
%V_{bur} (*Boltz*) = 36.3%
%V_{bur} (*min*) = 36.3%
Result: L_1Ni

DrewPhos
Cone angle = 207°
%V_{bur} (*Boltz*) = 27.2%
%V_{bur} (*min*) = 27.2%
Result: L_2Ni

PteroPhos
Cone angle = 235°
%V_{bur} (*Boltz*) = 29.2%
%V_{bur} (*min*) = 27.3%
Result: L_2Ni

(PPh₃)₂Ni(2-naphthaldehyde), **1**
X-Ray Structure

(Pt-Bu₃)₂Ni(benzaldehyde), **2**
DFT Structure (NMR Characterization)

(DrewPhos)₂Ni(2-naphthaldehyde), **3**
X-Ray Structure

(PteroPhos)₂Ni(benzaldehyde), **4**
DFT Structure (NMR Characterization)

B. %V_{bur} (*min*) best quantifies structure of flexible phosphines in L_2M complexes

Lowest energy
free P(i-Bu)₃ conformer
%V_{bur} = 38.7%
Library V_{bur} (*Boltz*) = 39.5%

(P(i-Bu)₃)₂Ni(2-naphtahldehyde), **5**
P1 %V_{bur} = 29.0%   P2 %V_{bur} = 32.2%
Library V_{bur} (*min*) = 31.1%

C. DFT Studies

ΔG_{dissoc}

Fig. 4. Mechanistic studies of %V_{bur} (*min*) ligation state threshold. (**A**) Ligands with large cone angles but relatively small %V_{bur} values can form L_2Ni, supported structurally by solid-state structures of (PPh₃)₂Ni(2-naphthaldehyde)[a] and (DrewPhos)₂Ni(2-naphthaldehyde)[a], along with DFT structures of (P(t-Bu)₃)Ni(benzaldehyde) and (PteroPhos)₂Ni(benzaldehyde) optimized at the B3LYP-D3/6-31G(d,p)[SDD] level of theory. (**B**) Formation of (P(i-Bu)₃)₂Ni(2-naphthaldehyde)[a], with %V_{bur} values of the ligands bound are in agreement with the library value of %V_{bur} (*min*) (**C**) DFT calculated dissociation energies and bond lengths for L_2Ni(benzaldehyde) complexes. Ni–P bond length calculated from highlighted bond. DinoPhos-type ligands (including DrewPhos) highlighted in purple. Calculations performed at M11-L/def2-TZVP|SMD(benzene)//B3LYP-D3/6-31G(d,p)[SDD]|SMD(benzene) level of theory. [a](ORTEP diagram with 30% thermal ellipsoids shown. Hydrogens omitted for clarity).

While this case study highlights the dependence of speciation on steric bulk within the first coordination sphere, which is captured by the %V_{bur} parameters, the ligands investigated in Fig.

4A had small variations between their %V_{bur} (*min*) and %V_{bur} (*Boltz*) values. To better understand the structural importance of %V_{bur} (*min*) in the classification of the catalytic (Figure 2B) and ligation state experiments (Figure 3B), the L_2Ni-forming ligand with the greatest difference between %V_{bur} (*min*) and %V_{bur} (*Boltz*) values, P(*i*-Bu)$_3$, was investigated. For this phosphine, lowest energy conformer and Boltzmann average %V_{bur} values (38.7% and 39.5%, respectively) are both substantially higher than other L_2Ni-forming ligands, and even higher than that of L_1Ni-forming P(*t*-Bu)$_3$ (36.3%). However, an X-ray crystal structure of (P(*i*-Bu)$_3$)$_2$Ni(2-naphthaldehyde) (**5**) (Figure 4B) confirmed that two phosphines were bound to Ni, with %V_{bur} values determined to be 29.0% and 32.2%, nearly matching that of %V_{bur} (*min*) (31.1%). Thus, only %V_{bur} (*min*), which was designed to capture the degree of conformational flexibility, simulates the redistribution of steric bulk out of the metal's first coordination sphere in order to accommodate both ligands and the substrate for phosphines like P(*i*-Bu)$_3$.

Finally, given the sharp catalytic and ligation state reactivity cliffs, we postulated that there would be a strong dependence on the thermodynamics of phosphine dissociation with %V_{bur} (*min*). Out of all 28 phosphines studied in Figure 3, only one (Cy$_2$P(*t*-Bu)) was found to have any spectroscopically observable equilibrium between L_2Ni and L_1Ni (1:4 L_2:L_1), with a %V_{bur} (*min*) value of 32.0% near the speciation cutoff. Using Cy$_2$P(*t*-Bu) as a reference, the free energy of ligand dissociation (ΔG_{dissoc}) was calculated for 20 L_2Ni(benzaldehyde) complexes using DFT (Figure 4C). A %V_{bur} (*min*) cutoff of 32% cleanly separated the regions of positive and negative ΔG_{dissoc}, with the remote steric bulk of the DinoPhos ligands and DrewPhos having minimal impacts on dissociation energy. Furthermore, the sharp downtick in the ΔG_{dissoc} values as the %V_{bur} (*min*) of the phosphine approached 32% corresponded with a significant uptick in the Ni–P bond length of the computed L_2 complex, a reflection of the increased steric pressure caused by filling

228   Ni's first coordination sphere at this value. Together, these studies demonstrate the strong physical

229   significance of the $\%V_{\text{bur}}$ (*min*) parameter and the 32% threshold value observed experimentally

230   in the datasets.

231   ***Development of threshold analysis algorithm****.* With the identification of sharp reactivity cliffs in

232   Ni SMC datasets, we sought to formalize the analysis to aid in the automated discovery of

233   thresholds and consequent classification of ligands within a diverse array of reaction datasets.

234   While $\%V_{\text{bur}}$ (*min*) was the most successful parameter in classifying ligands in the Ni-catalyzed

235   cross-coupling reactions, the workflow was designed to use all available parameters, and was

236   implemented as follows: the user first defines the % yield or selectivity for a "successful" reaction

237   (either above "ligand-less" control reactivity or a single catalyst turnover), after which the data is

238   parsed by a single-node decision tree algorithm equipped with any descriptor sets provided to it.

239   The algorithm then determines the location of the threshold (if present) for each feature (see SI for

240   details). A consequence of this analysis is the dissection of datasets into four sectors of a confusion

241   matrix: 1) *active* ligands that *meet* the threshold speciation criterion (true positives), 2) *inactive*

242   ligands that correctly *do not meet* the criterion (true negatives), 3) *active* ligands that *do not meet*

243   the criterion (false negatives), and 4) *inactive* ligands that *meet* the criterion but fail for other

244   reasons (false positives) (*37*). These groupings can then in theory be analyzed further using

245   statistical modeling or as targets for mechanistic interrogation (*38*).

246   Applying this workflow to the Ni datasets, we were able to evaluate the algorithm's ability

247   to computationally identify and quantify the thresholds observed empirically at the outset (see SI

248   for details). Most of the false positives in the classifications were ligands that were either electron-

249   poor or contained catalyst-poisoning functionality, and thus failed for reasons other than

250   speciation. As an example, P(4-CF$_3$Ph)$_3$ was found to form L$_2$Ni(4-fluorobenzaldehyde)

251   complexes but was inactive in all of the catalytic reactions.

252   ***Threshold analysis in Pd-catalyzed cross-coupling reactions.*** Considering the success of the

253   classification tool for Ni-catalyzed cross-coupling reactions, we sought to determine if %V_{bur} (*min*)

254   could also be applied in identifying speciation more broadly. In particular, we investigated various

255   Pd-catalyzed cross-coupling reactions wherein the formation of L$_1$Pd versus L$_2$Pd species has been

256   demonstrated to be a key facet of determining reaction outcome (*27, 39*). In one case study, we

257   probed a Pd-catalyzed SMC of aryl triflates, for which L$_2$Pd is proposed to be necessary for C–O

258   bond oxidative addition (**VII**, Figure 5A) (*27, 40*). The reliance on L$_2$Pd in catalysis also provided

259   an opportunity to test the DinoPhos ligands for the first time in Pd-catalysis. We generated a dataset

260   using 56 phosphines and found a %V_{bur} (*min*) reactivity threshold at 32.4% (**VII**, Figure 5B). The

261   reactivity cliff mirrored those of the Ni systems in both the cutoff value and directionality of

262   active/inactive regions. Further, DrewPhos and CyTyrannoPhos were both correctly classified as

263   active ligands based on their %V_{bur} (*min*) values below the threshold and performed similarly to

264   "undecorated" PPh$_3$ and CyPPh$_2$, respectively. Spectroscopic studies corroborated these results,

265   with the detection of L$_2$Pd(dba) by ^{31}P NMR for DrewPhos and the DinoPhos ligands studied,

266   including less active CyTriceraPhos (see SI for details). A necessity of the analysis for reaction

267   **VII** was to partition and classify Buchwald-type phosphines separately, regardless of %V_{bur} (*min*)

268   (see SI). This is likely due to the fact that they can occupy two coordination sites, a design element

269   of these ligands that allows them to mimic an L$_2$Pd species during catalysis through stabilizing Pd-

270   arene interactions (*24, 40, 41*). Consistent with previous investigations by the Sigman lab into the

271   chemoselectivity of aryl triflate SMCs (*40*), P(*o*-Tol)$_3$ was an outlier in the reaction, despite likely

272   favoring L$_1$Pd species under the reaction conditions (*42*).

273   In the next set of case studies, various reactions where L_1Pd species are implicated were

274   evaluated with the classification workflow. We generated datasets for two Pd-catalyzed SMC

275   couplings (**VIII & IX**, Figure 5A) with aryl halides, including one reaction with sterically hindered

276   coupling partners, as well as a dataset for a Buchwald-Hartwig amination (**X**, Figure 5A). For the

277   two SMCs, the classification tool revealed a threshold at approximately 29% V_{bur} (*min*), with the

278   active region occurring above that value, indicating that larger ligands—as defined by %V_{bur}

279   (*min*)—are required for effective catalysis (**VIII & IX**, Figure 5B). Notably, when comparing

280   reactions **VII** and **VIII,** the opposite directionalities of the observed thresholds are consistent with

281   the extensive literature precedent studying the effects of ligation state on the chemoselectivity of

282   aryl triflate vs. aryl chloride bond activation (*27, 39, 40*). The presence of steric bulk on both

283   coupling partners in reaction **IX** did not significantly impact the value at which the threshold

284   occurred, analogous to what was observed for a sterically hindered nucleophile in reaction **III** of

285   the Ni-catalyzed datasets. Perhaps the most striking example is that of the Buchwald-Hartwig

286   amination (**X**, Figure 5B) with a clear reactivity cliff between high yielding reactions and those

287   with 0% yield at and below 31% V_{bur} (*min*). Two datasets mined from the literature were also

288   evaluated: a stereospecific Pd-catalyzed SMC previously studied by Biscoe, Sigman and

289   coworkers (*18*) (**XI**, Figure 5A) and a Heck reaction reported by Hartwig and coworkers (*43*) (**XII**,

290   Figure 5A). In both cases, %V_{bur} (*min*) thresholds were observed and suggest that L_1Pd species are

291   responsible for catalysis (**XI & XII**, Figure 5B).
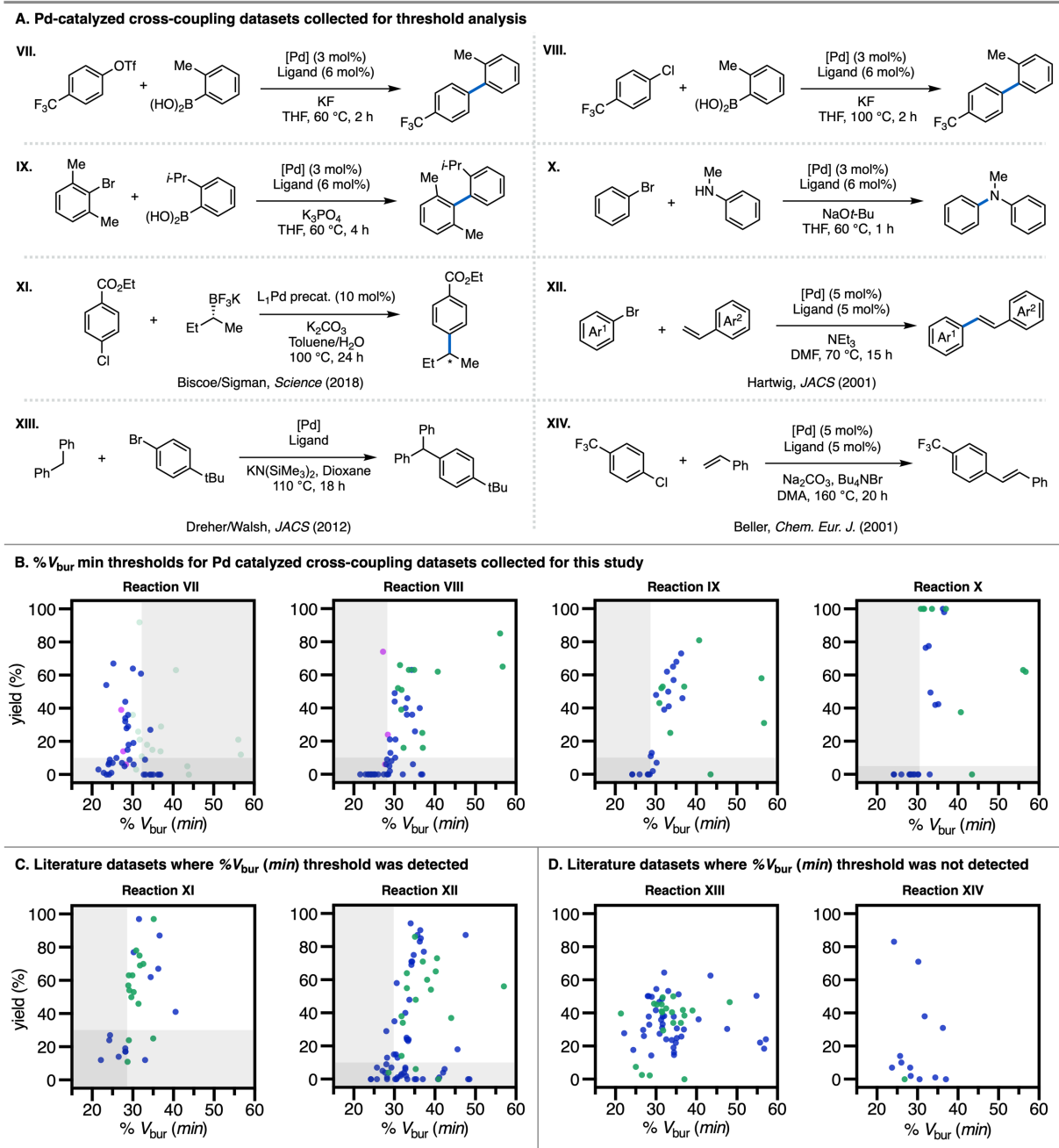
Fig. 5. Threshold analysis for Pd-catalyzed cross-coupling reactions. (**A**) Reaction schemes for reactions analyzed. **VII-X** collected for this study (see SI for exact reaction conditions), **XI**(*18*), **XII** (*43*), **XIII**(*44*), and **XIV**(*45*) mined from literature sources. (**B**) Threshold analysis of Pd datasets **VII-X**. (**C**) Threshold analysis of literature Pd datasets **XI-XII**. (**D**) Threshold analysis of literature Pd datasets **XIII-XIV**, where no %V_{bur} (*min*) threshold is present. For all plots, pink dots = DinoPhos-type ligands, green dots = Buchwald-type ligands, blue dots = all others.

We recognized that there are circumstances where ligand steric properties may not have as significant of an impact on the reaction outcome, and thus a threshold would not be expected. In

our literature mining efforts, we identified two datasets where no phosphine steric threshold was found: a Pd-catalyzed Csp^3-H arylation reported by Dreher, Walsh and coworkers (*44*) (**XIII**, Figure 5) and a Pd-catalyzed Heck cross-coupling studied by Beller and coworkers (*45*) (**XIV**, Figure 5). Possible explanations for these examples include high levels of ligandless background reactivity and/or catalytic cycles that are less sensitive to ligation state, possibly due to the generation of catalytically-active Pd nanoparticles (*46, 47*). The ability to rapidly identify these scenarios in datasets can serve as a valuable mechanistic probe given the physical importance of $\%V_{bur}$ (*min*) (*vide supra*).

Summary and Outlook: The threshold values and directionality found for each of the datasets studied in this work are summarized in Fig. 6. Comparing the Ni- and Pd-catalyzed reactions reveals several mechanistically interesting features about Ni and Pd catalysis, as well as phosphine steric effects more generally. First, the opposite threshold directionalities of the Ni- and Pd-catalyzed Suzuki–Miyaura coupling reactions of aryl halides (II-V, IX,

Fig. 6. Summary of reactivity thresholds. $\%V_{bur}$ (*min*) reactivity threshold locations and directionalities for all Ni (**I**–**V** (green)) and Pd (**VII**–**XII** (blue)) catalyzed reactions under study (excluding **XIII** & **VIV** where no threshold was found). The L_1 and L_2 regions defined by the spectroscopic ligation state studies (**VI** (purple)).

VIII) point to clear differences in mechanistic requirements between the two metals with regard to speciation for this reaction (*48*). Thus, the classification workflow provides a means to compare the two metals head-to-head and sheds light on the orthogonal ligand design principles often necessary for each metal (*19, 49*). Additionally, despite the differences in

324 coordination sphere size between Ni and Pd, the location of the threshold for the Ni datasets and

325 the Pd datasets with the sharpest reactivity cliffs (**VII** and **X**) occur at approximately the same

326 location of 32% V_{bur} (*min*), linking the catalytic outcome for both metals to ligand-controlled

327 thermodynamic L_1 vs. L_2 propensity.

328 It is also worth noting that the observed thresholds occurred with different sensitivities and

329 with some variation in value. Indeed, an overlap area can be observed wherein ligands with %V_{bur}

330 (*min*) values between 29% and 32% work in all of the reactions tested with both metals (with the

331 exception of reaction **X**). Accordingly, this may signify a region of %V_{bur} (*min*) values in which

332 L_2M is thermodynamically favored in the resting state but L_1M is still attainable; within this region,

333 the equilibrium between L_2M and L_1M would be impacted by the temperature and concentration

334 of the reaction. This is known to be the case with phosphines such as PCy_3 (*39*), and the decrease

335 in ΔG_{dissoc} between 29% and 32% (Figure 4C) corroborates this hypothesis. Interestingly, even in

336 the data for Ni-catalyzed cross-couplings, the most active ligands generally fell within these limits

337 (Figure 2). This observation suggests that L_1 intermediates and/or transition states may be relevant

338 in Ni catalysis (*50, 51*), but that the ability to attain an L_2 resting state is necessary (*33, 52, 53*), as

339 evidenced by the sharp reactivity cliffs.

340 In conclusion, we have developed a strategy for the binary classification of monodentate

341 phosphine ligation state and reactivity in cross-coupling catalysis. In searching the feature space

342 of structurally diverse monodentate phosphines, we identified %V_{bur} (*min*) as the parameter

343 capable of bifurcating datasets on the basis of distinct mechanistic outcomes. We envision that this

344 easily adopted tool should facilitate mechanistic studies of related organometallic reactions and

345 enable reaction development through prediction of active/inactive and mono/bis-ligating

346 phosphines prior to synthesis. While we recognize %V_{bur} (*min*) will not capture reactivity trends

347 across all phosphines, the ability to identify outliers (especially false negatives) in the analysis can

348 motivate the development of new parameters and targeted mechanistic study. Taken together, this

349 study highlights how classification analysis can serve as an important mechanistic and predictive

350 tool, complementary to LFERs, to facilitate statistical modeling of reactivity cliffs in chemistry.

References:

352 1.   L. P. Hammett, The Effect of Structure upon the Reactions of Organic Compounds.
353 Benzene Derivatives. *Journal of the American Chemical Society* **59**, 96-103 (1937).

354 2.   P. R. Wells, Linear Free Energy Relationships. *Chemical Reviews* **63**, 171-219 (1963).

355 3.   E. N. Jacobsen, W. Zhang, M. L. Guler, Electronic tuning of asymmetric catalysts. *Journal*
356 *of the American Chemical Society* **113**, 6703-6704 (1991).

357 4.   J. O. Schreck, Nonlinear Hammett relationships. *J. Chem. Ed.* **48**, 103 (1971).

358 5.   C. B. Santiago, J. Y. Guo, M. S. Sigman, Predictive and mechanistic multivariate linear
359 regression models for reaction development. *Chem Sci* **9**, 2398-2412 (2018).

360 6.   N. A. Romero, D. A. Nicewicz, Organic Photoredox Catalysis. *Chemical Reviews* **116**,
361 10075-10166 (2016).

362 7.   D. G. Brown, J. Boström, Analysis of Past and Present Synthetic Methodologies on
363 Medicinal Chemistry: Where Have All the New Reactions Gone? *J Med Chem* **59**, 4443-4458
364 (2016).

365 8.   S. Xu, E. H. Kim, A. Wei, E.-i. Negishi, Pd- and Ni-catalyzed cross-coupling reactions in
366 the synthesis of organic electronic materials. *Science and Technology of Advanced Materials* **15**,
367 044201 (2014).

368 9.   J. F. Hartwig, *Organotransition Metal Chemistry: From Bonding to Catalysis*. (University
369 Science Books, 2010).

370 10.   C. C. C. Johansson Seechurn, M. O. Kitching, T. J. Colacot, V. Snieckus, Palladium-
371 Catalyzed Cross-Coupling: A Historical Contextual Perspective to the 2010 Nobel Prize.
372 *Angewandte Chemie International Edition* **51**, 5062-5085 (2012).

373 11.   D. J. Durand, N. Fey, Computational Ligand Descriptors for Catalyst Design. *Chem Rev*
374 **119**, 6561-6594 (2019).

375 12.   C. A. Tolman, Phosphorus ligand exchange equilibriums on zerovalent nickel. Dominant
376 role for steric effects. *J. Am. Chem. Soc.* **92**, 2956-2965 (1970).

377   13.   C. A. Tolman, Steric effects of phosphorus ligands in organometallic chemistry and
378   homogeneous catalysis. *Chem. Rev.* **77**, 313-348 (1977).

379   14.   D. White, B. C. Tavener, P. G. L. Leach, N. J. Coville, Solid angles I. The radial profile.
380   *Journal of Organometallic Chemistry* **478**, 205-211 (1994).

381   15.   K. C. Harper, E. N. Bess, M. S. Sigman, Multidimensional steric parameters in the analysis
382   of asymmetric catalytic reactions. *Nat Chem* **4**, 366-374 (2012).

383   16.   A. Poater *et al.*, SambVca: A Web Application for the Calculation of the Buried Volume
384   of N-Heterocyclic Carbene Ligands. *European Journal of Inorganic Chemistry* **2009**, 1759-1766
385   (2009).

386   17.   H. Clavier, S. P. Nolan, Percent buried volume for phosphine and N-heterocyclic carbene
387   ligands: steric properties in organometallic chemistry. *Chem Commun (Camb)* **46**, 841-861 (2010).

388   18.   S. Zhao *et al.*, Enantiodivergent Pd-catalyzed C–C bond formation enabled through ligand
389   parameterization. *Science* **362**, 670 (2018).

390   19.   K. Wu, A. G. Doyle, Parameterization of phosphine ligands demonstrates enhancement of
391   nickel catalysis via remote steric effects. *Nat Chem* **9**, 779-784 (2017).

392   20.   H. Y. Liu, K. Eriks, A. Prock, W. P. Giering, Quantitative analysis of ligand effects
393   (QALE). Systematic study of iron-phosphorus bond lengths and their relationship to steric
394   thresholds. *Organometallics* **9**, 1758-1766 (1990).

395   21.   T. Gensch *et al.*, A comprehensive discovery platform for organophosphorus ligands for
396   catalysis. *ChemRxiv*, (2021).

397   22.   H. Clavier, S. P. Nolan, Percent buried volume for phosphine and N-heterocyclic carbene
398   ligands: steric properties in organometallic chemistry. *Chem. Comm.* **46**, 841-861 (2010).

399   23.   F. Barrios-Landeros, B. P. Carrow, J. F. Hartwig, Effect of ligand steric properties and
400   halide identity on the mechanism for oxidative addition of haloarenes to trialkylphosphine Pd(0)
401   complexes. *J Am Chem Soc* **131**, 8141-8154 (2009).

402   24.   R. Martin, S. L. Buchwald, Palladium-Catalyzed Suzuki−Miyaura Cross-Coupling
403   Reactions Employing Dialkylbiaryl Phosphine Ligands. *Accounts of Chemical Research* **41**, 1461-
404   1473 (2008).

405   25.   A. A. Thomas, A. F. Zahrt, C. P. Delaney, S. E. Denmark, Elucidating the Role of the
406   Boronic Esters in the Suzuki–Miyaura Reaction: Structural, Kinetic, and Computational
407   Investigations. *Journal of the American Chemical Society* **140**, 4401-4416 (2018).

408   26.   G. C. Fu, The development of versatile methods for palladium-catalyzed coupling reactions
409   of aryl electrophiles through the use of P(t-Bu)3 and PCy3 as ligands. *Acc Chem Res* **41**, 1555-
410   1564 (2008).

27.   F. Schoenebeck, K. N. Houk, Ligand-controlled regioselectivity in palladium-catalyzed cross coupling reactions. *J Am Chem Soc* **132**, 2496-2497 (2010).

28.   F.-S. Han, Transition-metal-catalyzed Suzuki–Miyaura cross-coupling reactions: a remarkable advance from palladium to nickel catalysts. *Chemical Society Reviews* **42**, 5270-5298 (2013).

29.   A. P. Cinderella, B. Vulovic, D. A. Watson, Palladium-Catalyzed Cross-Coupling of Silyl Electrophiles with Alkylzinc Halides: A Silyl-Negishi Reaction. *J Am Chem Soc* **139**, 7741-7744 (2017).

30.   Cone angle values used in text are those for the Boltzmann weighted average of the conformational ensemble from the descriptor library. Minimum cone angles were also evaluated but for simplicity are shown in the SI.

31.   A. J. Kendall, L. N. Zakharov, D. R. Tyler, Steric and Electronic Influences of Buchwald-Type Alkyl-JohnPhos Ligands. *Inorg Chem* **55**, 3079-3090 (2016).

32.   Buchwald-type phosphines are known to act as structurally bidentate-like ligands with Pd due to stabilizing interactions between the arene and the metal. While little is currently known about the structural behavior of these ligands with Ni, this semi-bidentate binding may enable certain Ni-catalyzed reactions to occur despite only one phosphine bound; their mode of action in these case studies is the subject of ongoing investigation..

33.   E. A. Standley, S. J. Smith, P. Müller, T. F. Jamison, A Broadly Applicable Strategy for Entry into Homogeneous Nickel(0) Catalysts from Air-Stable Nickel(II) Complexes. *Organometallics* **33**, 2012-2018 (2014).

34.   D. Walther, Reaktionen von heteroolefinen an zentralmetallen in niedrigen oxidationsstufen: stabile aldehydkomplexe des nickel(O) und verwandte verbindungen. *Journal of Organometallic Chemistry* **190**, 393-401 (1980).

35.   M. Ohashi, H. Saijo, T. Arai, S. Ogoshi, Nickel(0)-Catalyzed Formation of Oxaaluminacyclopentenes via an Oxanickelacyclopentene Key Intermediate: Me2AlOTf-Assisted Oxidative Cyclization of an Aldehyde and an Alkyne with Nickel(0). *Organometallics* **29**, 6534-6540 (2010).

36.   I. A. Guzei, M. Wendt, An improved method for the computation of ligand steric effects based on solid angles. *Dalton Transactions*, 3991-3999 (2006).

37.   Given that other ligand features can and will lead to poor catalyst performance, the presence of false positives should not be viewed as detrimental to the success/accuracy of the classifier.

38.   Z. L. Niemeyer *et al.*, Parameterization of Acyclic Diaminocarbene Ligands Applied to a Gold(I)-Catalyzed Enantioselective Tandem Rearrangement/Cyclization. *Journal of the American Chemical Society* **139**, 12943-12946 (2017).

447   39.   A. F. Littke, C. Dai, G. C. Fu, Versatile Catalysts for the Suzuki Cross-Coupling of
448   Arylboronic Acids with Aryl and Vinyl Halides and Triflates under Mild Conditions. *Journal of*
449   *the American Chemical Society* **122**, 4020-4028 (2000).

450   40.   Z. L. Niemeyer, A. Milo, D. P. Hickey, M. S. Sigman, Parameterization of phosphine
451   ligands reveals mechanistic pathways and predicts reaction outcomes. *Nat Chem* **8**, 610-617
452   (2016).

453   41.   K. L. Billingsley, K. W. Anderson, S. L. Buchwald, A Highly Active Catalyst for Suzuki–
454   Miyaura Cross-Coupling Reactions of Heteroaryl Compounds. *Angewandte Chemie International*
455   *Edition* **45**, 3484-3488 (2006).

456   42.   J. F. Hartwig, F. Paul, Oxidative Addition of Aryl Bromide after Dissociation of Phosphine
457   from a Two-Coordinate Palladium(0) Complex, Bis(tri-o-tolylphosphine)Palladium(0). *Journal of*
458   *the American Chemical Society* **117**, 5373-5374 (1995).

459   43.   J. P. Stambuli, S. R. Stauffer, K. H. Shaughnessy, J. F. Hartwig, Screening of
460   Homogeneous Catalysts by Fluorescence Resonance Energy Transfer. Identification of Catalysts
461   for Room-Temperature Heck Reactions. *Journal of the American Chemical Society* **123**, 2677-
462   2678 (2001).

463   44.   J. Zhang, A. Bellomo, A. D. Creamer, S. D. Dreher, P. J. Walsh, Palladium-Catalyzed
464   C(sp3)–H Arylation of Diarylmethanes at Room Temperature: Synthesis of Triarylmethanes via
465   Deprotonative-Cross-Coupling Processes. *Journal of the American Chemical Society* **134**, 13765-
466   13772 (2012).

467   45.   A. Zapf, M. Beller, Palladium Catalyst Systems for Cross-Coupling Reactions of Aryl
468   Chlorides and Olefins. *Chemistry – A European Journal* **7**, 2908-2915 (2001).

469   46.   S. S. Zalesskiy, V. P. Ananikov, Pd2(dba)3 as a Precursor of Soluble Metal Complexes
470   and Nanoparticles: Determination of Palladium Active Species for Catalysis and Synthesis.
471   *Organometallics* **31**, 2302-2309 (2012).

472   47.   N. W. J. Scott *et al.*, The ubiquitous cross-coupling catalyst system 'Pd(OAc)2'/2PPh3
473   forms a unique dinuclear PdI complex: an important entry point into catalytically competent cyclic
474   Pd3 clusters. *Chemical Science* **10**, 7898-7906 (2019).

475   48.   S. Z. Tasker, E. A. Standley, T. F. Jamison, Recent advances in homogeneous nickel
476   catalysis. *Nature* **509**, 299-309 (2014).

477   49.   C. M. Lavoie, M. Stradiotto, Bisphosphines: A Prominent Ancillary Ligand Class for
478   Application in Nickel-Catalyzed C–N Cross-Coupling. *ACS Catalysis* **8**, 7228-7250 (2018).

479   50.   H. Shirataki, M. Ohashi, S. Ogoshi, Nickel-Catalyzed Three-Component Coupling
480   Reaction of Tetrafluoroethylene and Aldehydes with Silanes via Oxa-Nickelacycles. *European*
481   *Journal of Organic Chemistry* **2019**, 1883-1887 (2019).

51.   We recognize that there are some instances where Ni will be L_1 throughout the catalytic cycle allowing ligands like P(t-Bu)$_3$ to be employed successfully. See ref (50).

52.   N. Hazari, P. R. Melvin, M. M. Beromi, Well-defined nickel and palladium precatalysts for cross-coupling. *Nat Rev Chem* **1**, (2017).

53.   P.-A. Payard, L. A. Perego, I. Ciofini, L. Grimaud, Taming Nickel-Catalyzed Suzuki-Miyaura Coupling: A Mechanistic Focus on Boron-to-Nickel Transmetalation. *ACS Catalysis* **8**, 4812-4823 (2018).

Author Contributions: T.G., M.S.S., and A.G.D. conceived the project. S.H.N.-S., S.R.S., J.E.B., and H.C.J. performed the experiments and analyzed the data. E.P., T.G., and S.R.S. developed the classification algorithm. M.S.S. and A.G.D. oversaw the project. S.H.N.-S., J.E.B., S.R.S., M.S.S., and A.G.D. wrote the manuscript with contributions from all authors.

Competing Interests: The authors declare no competing interests.

Data and Materials Availability: X-ray crystallographic data are available free of charge from the Cambridge Crystallographic Data Centre, under reference numbers 2069892-2069895. Coordinates of all computed structures are included in a separate zip file. All other experimental, computational, spectroscopic, and crystallographic data are included in the supplementary materials. The code used in this project is available at https://github.com/SigmanGroup/Threshold.

Supplementary Materials
Materials and Methods
Figs. S1 to S116
Tables S1 to S18
References