

# A Mechanistically Guided Workflow for Relating Complex Reactive Site Topologies to Catalyst Performance in C–H Functionalization Reactions

Ryan C. Cammarota<sup>‡,1</sup>, Wenbin Liu<sup>‡,2</sup>, John Bacsa<sup>2</sup>, Huw M. L. Davies<sup>\*,2</sup> and Matthew S. Sigman<sup>\*,1</sup>

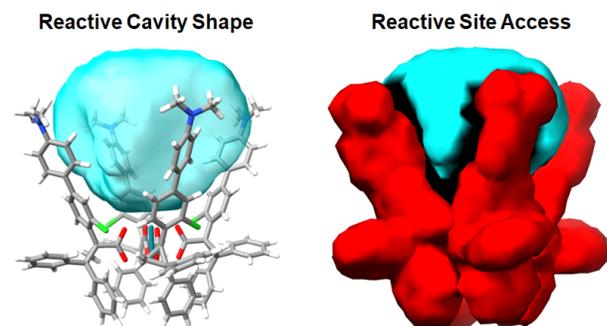
<sup>1</sup>Department of Chemistry, University of Utah, 315 South 1400 East, Salt Lake City, Utah 84112

<sup>2</sup>Department of Chemistry, Emory University, 1515 Dickey Drive, Atlanta, Georgia 30322

**ABSTRACT:** Leveraging congested catalyst scaffolds has emerged as a key strategy for altering innate substrate site-selectivity profiles in C–H functionalization reactions. Similar to enzyme active sites, optimal small molecule catalysts often feature reactive cavities tailored for controlling substrate approach trajectories. However, relating three-dimensional catalyst shape to reaction output remains a formidable challenge, in part due to the lack of molecular features capable of succinctly describing complex reactive site topologies in terms of numerical inputs for machine learning applications. Herein, we present a new set of descriptors, “Spatial Molding for Approachable Rigid Targets” (SMART), which we have applied to quantify reactive site spatial constraints for an expansive library of dirhodium catalysts and to predict site-selectivity for C–H functionalization of 1-bromo-4-pentylbenzene via donor/acceptor carbene intermediates. Optimal site-selectivity for the terminal methylene position was obtained with  $\text{Rh}_2(\text{S-2-Cl-5-MesTPCP})_4$  (30.9:1 rr, 14:1 dr, 87% ee), while C–H functionalization at the electronically activated benzylic site was increasingly favored for  $\text{Rh}_2(\text{TPCP})_4$  catalysts lacking an *ortho*-Cl,  $\text{Rh}_2(\text{S-PTAD})_4$ , and  $\text{Rh}_2(\text{S-TCPTAD})_4$ , respectively. Intuitive global site-selectivity models for 25 disparate dirhodium catalysts were developed via multivariate linear regression to explicitly assess the contributing roles of steric congestion and dirhodium-carbene electrophilicity in controlling the site of C–H functionalization. The workflow utilizes spatial classification to extract descriptors only for reactive catalyst conformers, a nuance that may be widely applicable for establishing close correspondence between ground-state model systems and transition states. Broader still, SMART descriptors are amenable for delineating salient reactive site features to predict reactivity in other chemical and biological contexts.

## Introduction

The advent of catalysts capable of mediating selective C–H functionalization has transformed traditional approaches to organic synthesis.<sup>1</sup> As this field continues to evolve, developing predictive models for catalyst propensity to differentiate between numerous similar substrate C–H bonds would be enabling.<sup>2</sup> Namely, improving predictive modeling capabilities would expedite the design of new catalysts that are increasingly adept at overriding inherent substrate reactivity profiles in high stereoselectivity.<sup>3</sup> This formidable challenge is generally pursued via trial-and-error catalyst modification guided by chemical intuition alone. Nevertheless, substantial progress toward catalyst-controlled intermolecular C–H functionalization<sup>4</sup> has been achieved by leveraging sterically congested catalysts to confine the approach trajectories of substrate toward highly reactive metal-carbene,<sup>5</sup> metal-nitrene,<sup>6</sup> and metal-oxo intermediates.<sup>7</sup> In effect, many of these advances have relied upon peripheral repulsive interactions to bridge the continuum between small molecule catalysts and enzymes.<sup>8</sup> Indeed, many C–H functionalization catalysts that invert innate substrate site-

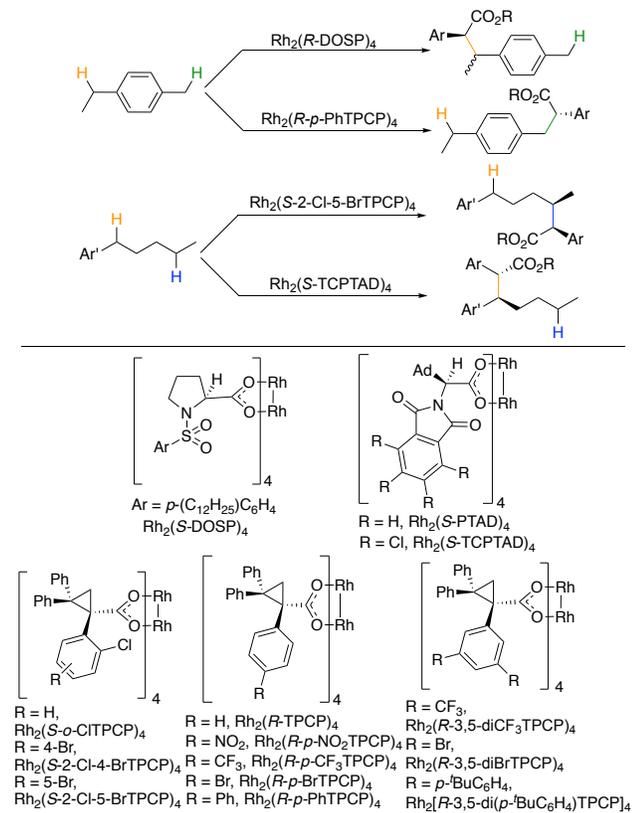


**Figure 1.** Application of new molecular descriptors to enable prediction of site-selectivity in C–H functionalization reactions as a function of complex catalyst cavity shapes.

selectivity profiles have come to resemble miniature enzymes, with both featuring dynamic reactive site cavities with irregular shapes that are critical to function (Figure 1).

The extensive library of dirhodium tetracarboxylate catalysts developed by the Davies group is an illustrative example of specifically tailoring catalyst pocket shapes for mediating site- and stereoselective C–H functionalization.<sup>5h</sup> By tuning catalyst properties, the site-selectivity of benzylic C–H

**Scheme 1.** Dirhodium catalysts developed by Davies and co-workers for controlling C–H functionalization via donor/acceptor carbenes intermediates



insertion via donor/acceptor carbene intermediates was shifted from 2° to 1° sites using a sterically demanding Rh<sub>2</sub>(*R*-*p*-PhTCP)<sub>4</sub> catalyst (TCP = triphenylcyclopropane carboxylate; Scheme 1).<sup>9</sup> Subsequently, further developments showcased stereoselective functionalization of unactivated 1°, 2°, or 3° C–H bonds by identifying the appropriate catalyst.<sup>10</sup> Most recently, a new class of hindered catalysts, Rh<sub>2</sub>(*S*-*o*-CITPCP)<sub>4</sub>, was developed that afforded unprecedented site-selectivity for terminal unactivated methylene C–H bonds (C2) in the presence of electronically activated benzylic (Bn) C–H bonds (Scheme 1).<sup>11</sup> Previous computational and X-ray crystallographic studies have established that these Rh<sub>2</sub>(*S*-*o*-CITPCP)<sub>4</sub> catalysts preferentially adopt “bowl-shaped” C<sub>4</sub> symmetric structures,<sup>11b, 12</sup> which has been hypothesized to play an important role in inverting site-selectivity to functionalize the more sterically accessible C2 position. However, understanding how modest changes in catalyst structure will impact performance is challenging, both in the context of *post hoc* rationalization and *a priori* prediction. This difficulty stems from the intricacy of these catalyst scaffolds, which contain multiple flexible ligands that conspire to induce a unique reactive site cavity shape.

Considering the ubiquity of this type of spatial control element in catalysis, we aimed to build an efficient computational modeling workflow for predicting the performance of large, complex catalysts as a function of reactive site steric constraints and electronic environment. An

invaluable and widely applied computational strategy for understanding catalyst performance is interrogating competing selectivity-determining transition states in the reaction pathway using quantum mechanical calculations.<sup>10a, 11b, 12a, 13</sup> Such analyses, in conjunction with experimental mechanistic studies, are typically performed for a few catalysts of interest after the empirical discovery of an optimal catalyst to rationalize divergent reactivity. However, transition state analysis is prohibitively time-intensive to employ as a virtual screening tool for large, conformationally flexible transition metal catalysts across an extensive library of potential catalyst candidates.<sup>2g, 14</sup> Alternatively, large-scale *in silico* screening can potentially be achieved using multivariate linear regression (MLR)<sup>15</sup> or other machine learning (ML) algorithms,<sup>16</sup> where a predictive model for reaction output can be established as a function of relative differences in more readily computed ground-state molecular descriptors of a simplified catalyst model system (e.g., a free ligand).<sup>17</sup> However, predictive modeling has seldom been applied for catalysts as complex as those developed by the Davies group. This void is primarily due to the lack of molecular descriptors capable of meaningfully quantifying the spatial constraints and accessibility of catalyst reactive sites (*vide infra*), a critical prerequisite for developing intuitive structure-reactivity/selectivity relationships that are sufficiently robust for extrapolation.

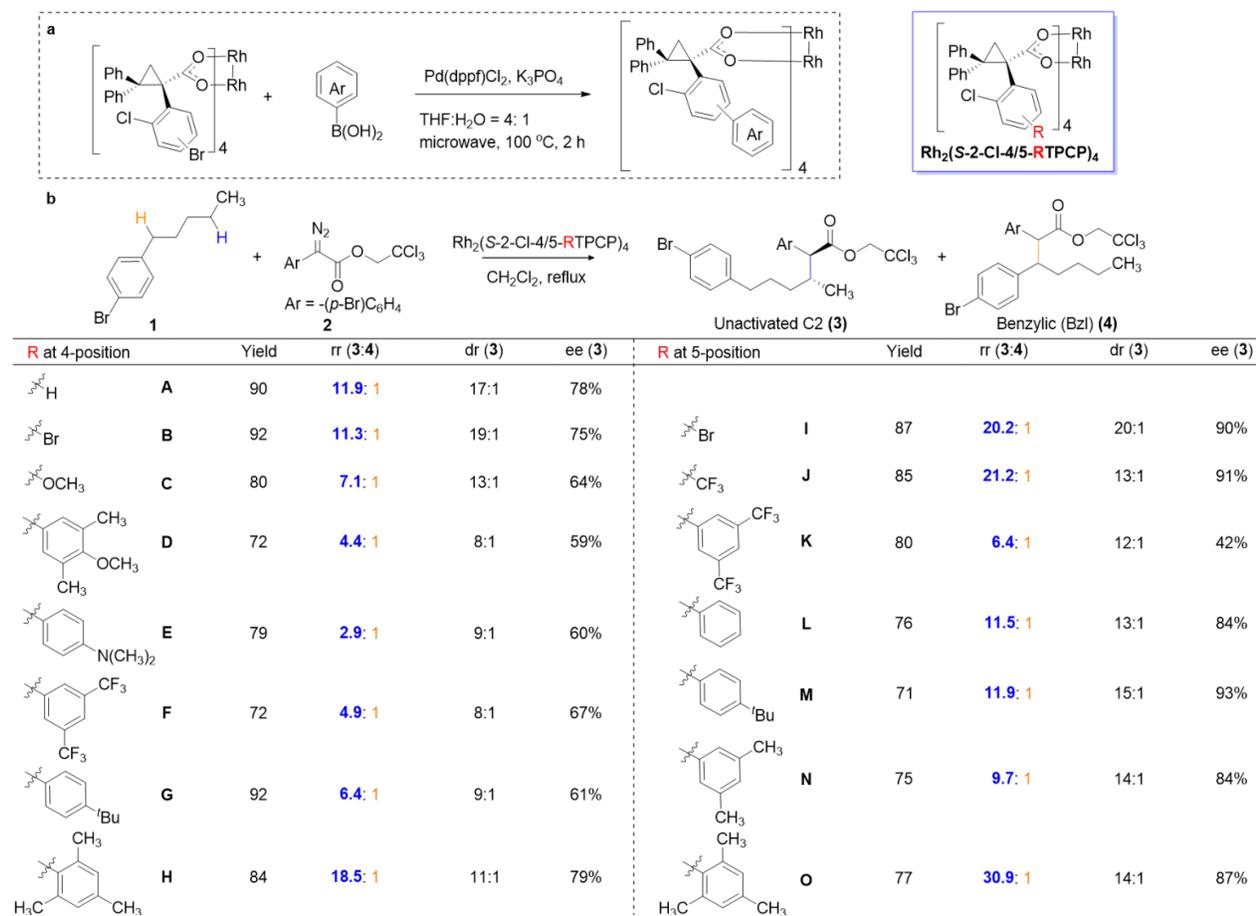
With this as a backdrop, we sought to intimately link catalyst cavity shape with reactivity by developing new molecular descriptors for reactive site spatial constraints. Herein, we introduce “Spatial Molding for Approachable Rigid Targets” (SMART) descriptors (Figure 1), which we demonstrate to be an effective means of quantitatively understanding catalyst steric effects that dictate site-selectivity in the C–H functionalization of 1-bromo-4-pentylbenzene via donor/acceptor dirhodium-carbene intermediates. Global statistical models were developed that leverage SMART cavity constraint and peripheral rigidity descriptors, in conjunction with interpretable catalyst electronic parameters, to predict site-selectivity across a systematically diversified series of 25 dirhodium catalysts. In addition to providing a framework for understanding peripheral repulsive interactions, the model also deconvolutes the interplay of electrostatic and global structural effects on dirhodium-carbene electrophilicity that dictate site-selectivity. More broadly, the tools applied herein are adaptable for capturing the underlying steric effects that govern other reaction types, in both synthetic and biological contexts, where substrate access to a reactive site is restricted by irregular peripheral cavity constraints.

## Results and Discussion

### Synthesis and Evaluation of a Diversified Catalyst Library

Introducing an *ortho*-Cl substituent to the C1 phenyl ring of TCP ligands was previously found to have a dramatic effect on site-selectivity in the dirhodium-catalyzed C–H functionalization of 1-bromo-4-pentylbenzene, with Rh<sub>2</sub>(*S*-2-Cl-5-BrTCP)<sub>4</sub> (**B**) affording optimal C2:Bn site-selectivity (20.2:1 rr) and enantioselectivity (90% ee).<sup>11b</sup> Seeking to further understand catalyst performance and improve upon

**Scheme 2.** Synthesis (a) and evaluation of site-selectivity (b) for an expanded library of  $\text{Rh}_2(\text{S-}o\text{-CITPCP})_4$  catalysts.

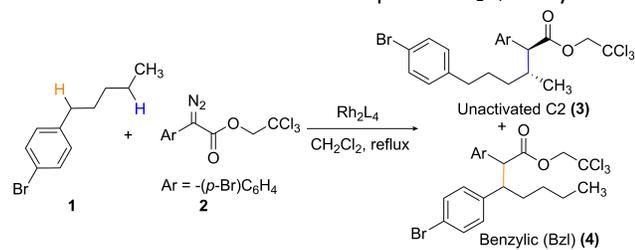


this result, we designed and synthesized a diversified library of  $\text{Rh}_2(\text{S-}o\text{-CITPCP})_4$  catalysts (see SI).<sup>18</sup> Notably, complexes with additional aryl rings were synthesized via four-fold Pd-catalyzed cross-coupling from the parent dirhodium complexes,  $\text{Rh}_2(\text{S-}2\text{-Cl-}5\text{-BrTPCP})_4$  (**I**) and **B** (Scheme 2a).<sup>10c</sup>

This expanded series of  $\text{Rh}_2(\text{S-}o\text{-CITPCP})_4$  catalysts was evaluated in the C–H functionalization of 1-bromo-4-pentylbenzene with 2,2,2-trichloroethyl-2-(4-bromo-phenyl)-2-diazoacetate (Scheme 2b). Much like **B** and **I**, all new  $\text{Rh}_2(\text{S-}o\text{-CITPCP})_4$  catalysts evaluated also exhibited a strong preference for C2 functionalization, with site-selectivity ranging from 2.9:1 to 30.9:1 rr. While more sterically hindered catalysts should prefer to functionalize the more accessible substrate C2 site,<sup>5h</sup> quantitative trends that intuitively relate catalyst properties with performance are not discernable (Scheme 2b). Although C2 site-selectivity was generally higher for  $\text{Rh}_2(\text{S-}o\text{-CITPCP})_4$  catalysts with aryl substitution at the 5-position rather than the 4-position (c.f., **M** to **G**, **K** to **F**, **O** to **H**, **I** to **B**), catalysts within each subclass bearing both electron-rich and electron-poor aryl groups afforded similar site-selectivity (c.f., **D** to **F**, **K** to **N**). Excitingly, catalysts featuring bulky mesityl groups at either the 4- (**H**) or 5-position (**O**) gave markedly improved C2:Bn site-selectivity. Optimal site-selectivity (30.9:1 rr) was obtained using **O**, which also afforded favorable yield (77%) and stereoselectivity (14:1 dr, 87% ee).

To better understand the unique performance of  $\text{Rh}_2(\text{S-}o\text{-CITPCP})_4$  catalysts, it is necessary to elucidate why other classes of dirhodium catalysts do not perform as well. In this vein, the data set for the C–H functionalization of 1-bromo-4-pentylbenzene under identical conditions was expanded to investigate the site-selectivity afforded by 10 more dissimilar dirhodium tetracarboxylate catalysts (Scheme 3), building upon the originally reported catalyst screening.<sup>11b</sup> The parent  $\text{Rh}_2(\text{R-TPCP})_4$  catalyst (**P**) showed a modest predilection for functionalization at the electronically activated benzylic site (entry 1). Similarly, *para*-substituted TCPCP derivatives,  $\text{Rh}_2(\text{R-}p\text{-XTPCP})_4$  (**Q-T**), also gave minimal differentiation between functionalization at the C2 and benzylic sites (1:1.8 to 2.1:1 rr, entries 2-5). In addition, three di-*meta*-substituted derivatives,  $\text{Rh}_2(\text{3,5-diXTPCP})_4$  (**U-W**), were found to exhibit a slight preference for the C2 site (1.5:1 to 3.2:1 rr, entries 6-8). Notably, within both the  $\text{Rh}_2(\text{3,5-diXTPCP})_4$  and  $\text{Rh}_2(\text{p-XTPCP})_4$  subclasses, the catalysts with the bulkiest substituents afforded the highest C2 site-selectivity ( $X = p\text{-}^t\text{BuC}_6\text{H}_4$  and Ph, respectively). Lastly, two phthalimido catalysts,  $\text{Rh}_2(\text{S-PTAD})_4$  (**X**) and  $\text{Rh}_2(\text{S-TCPTAD})_4$  (**Y**), were found to favor benzylic C–H functionalization. This preference was especially strong for **Y** (1:8.7 rr, entry 10). Intriguingly, X-ray crystallographic and computational analyses indicate that **Y** adopts a similar  $\text{C}_4$  symmetric bowl-shaped structure to those of  $\text{Rh}_2(\text{S-}o\text{-CITPCP})_4$  catalysts that facilitate selective C2 functionalization.<sup>11b, 19</sup>

### Scheme 3. Performance of more disparate Rh<sub>2</sub>L<sub>4</sub> catalysts.



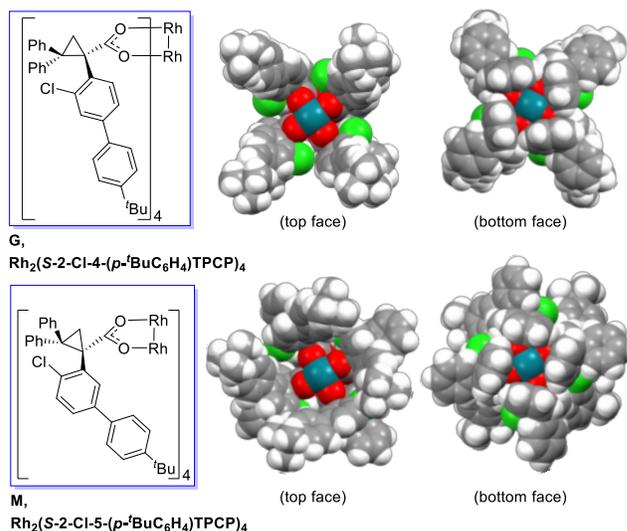
Entry	Rh <sub>2</sub> L <sub>4</sub>	rr (3:4) <sup>b</sup>
1	Rh <sub>2</sub> (R-TPCP) <sub>4</sub> , <b>P</b>	1: 2.5 (29, 71)
2	Rh <sub>2</sub> (R- <i>p</i> -NO <sub>2</sub> TPCP) <sub>4</sub> , <b>Q</b>	1: 1.8 (36, 64)
3	Rh <sub>2</sub> (R- <i>p</i> -CF <sub>3</sub> TPCP) <sub>4</sub> , <b>R</b>	1: 1.2 (45, 55)
4	Rh <sub>2</sub> (S- <i>p</i> -BrTPCP) <sub>4</sub> , <b>S</b>	1.8: 1 (64, 36)
5	Rh <sub>2</sub> (S- <i>p</i> -PhTPCP) <sub>4</sub> , <b>T</b>	2.1: 1 (68, 32)
6	Rh <sub>2</sub> (S-3,5-diCF <sub>3</sub> TPCP) <sub>4</sub> , <b>U</b>	1.8: 1 (64, 36)
7	Rh <sub>2</sub> (S-3,5-diBrTPCP) <sub>4</sub> , <b>V</b>	1.5: 1 (60, 40)
8	Rh <sub>2</sub> (R-3,5-di( <i>p</i> - <sup>t</sup> BuC <sub>6</sub> H <sub>4</sub> )TPCP) <sub>4</sub> , <b>W</b>	3.2: 1 (76, 23)
9	Rh <sub>2</sub> (S-PTAD) <sub>4</sub> , <b>X</b>	1: 2.0 (33, 67)
10	Rh <sub>2</sub> (S-TCPTAD) <sub>4</sub> , <b>Y</b>	1: 8.7 (10, 90)

<sup>a</sup>Conditions: **2** (0.3 mmol) in 6 mL CH<sub>2</sub>Cl<sub>2</sub> solution was added over 3 h to a solution of Rh<sub>2</sub>L<sub>4</sub> (1.0 mol%) and **1** (0.6 mmol) in 3 mL CH<sub>2</sub>Cl<sub>2</sub> under reflux. The reaction was stirred for an additional 1 h. <sup>b</sup>Determined by crude <sup>1</sup>H NMR in lieu of isolation.

#### Workflow for Catalyst Performance Prediction

With experimental data for 25 dirhodium tetracarboxylate catalysts in hand, catalyst structures and properties must next be accurately obtained to establish robust quantitative structure-reactivity relationships. Single-crystal X-ray diffraction studies are an invaluable experimental tool for obtaining structural insights, and X-ray structures were determined for four new catalysts (**G** and **M** in Figure 2; **H** and **N** in Figure S44). Despite the larger substituents on the *ortho*-Cl-aryl rings, all four catalysts were found to adopt C<sub>4</sub> symmetric configurations in the solid-state similar to those previously observed for **A**, **B**, and **I**.<sup>11b, 12b</sup> As shown in Figure 2, access to the bottom Rh face is significantly blocked by four closely packed phenyl rings. Thus, dirhodium-carbene formation can only occur on the top Rh face within the axially chiral, “bowl-shaped” pocket induced by hindered rotation of the *ortho*-Cl-aryl groups, which are maximally separated to minimize *ortho*-Cl steric clashes.

To enable virtual screening of potential catalysts that have not yet been synthesized, a prediction platform must rely on calculated structures. Given the large number of atoms and possible conformations<sup>20</sup> for each of the 25 catalysts, a QM/MM approach<sup>21</sup> was utilized to efficiently identify preferred catalyst conformations (M06-D3/def2tzvp-SDD(Rh)-SMD(CH<sub>2</sub>Cl<sub>2</sub>)/[B3LYP-D3(BJ)/6-31G\*-lanl2dz(Rh)]:UFF).<sup>22</sup> Adapting the methods of Houk and Musaev<sup>10a</sup> to include the C1 aryl and cyclopropyl rings in the QM partition was found to be critical for matching the “bowl” shapes adopted by Rh<sub>2</sub>(S-*o*-



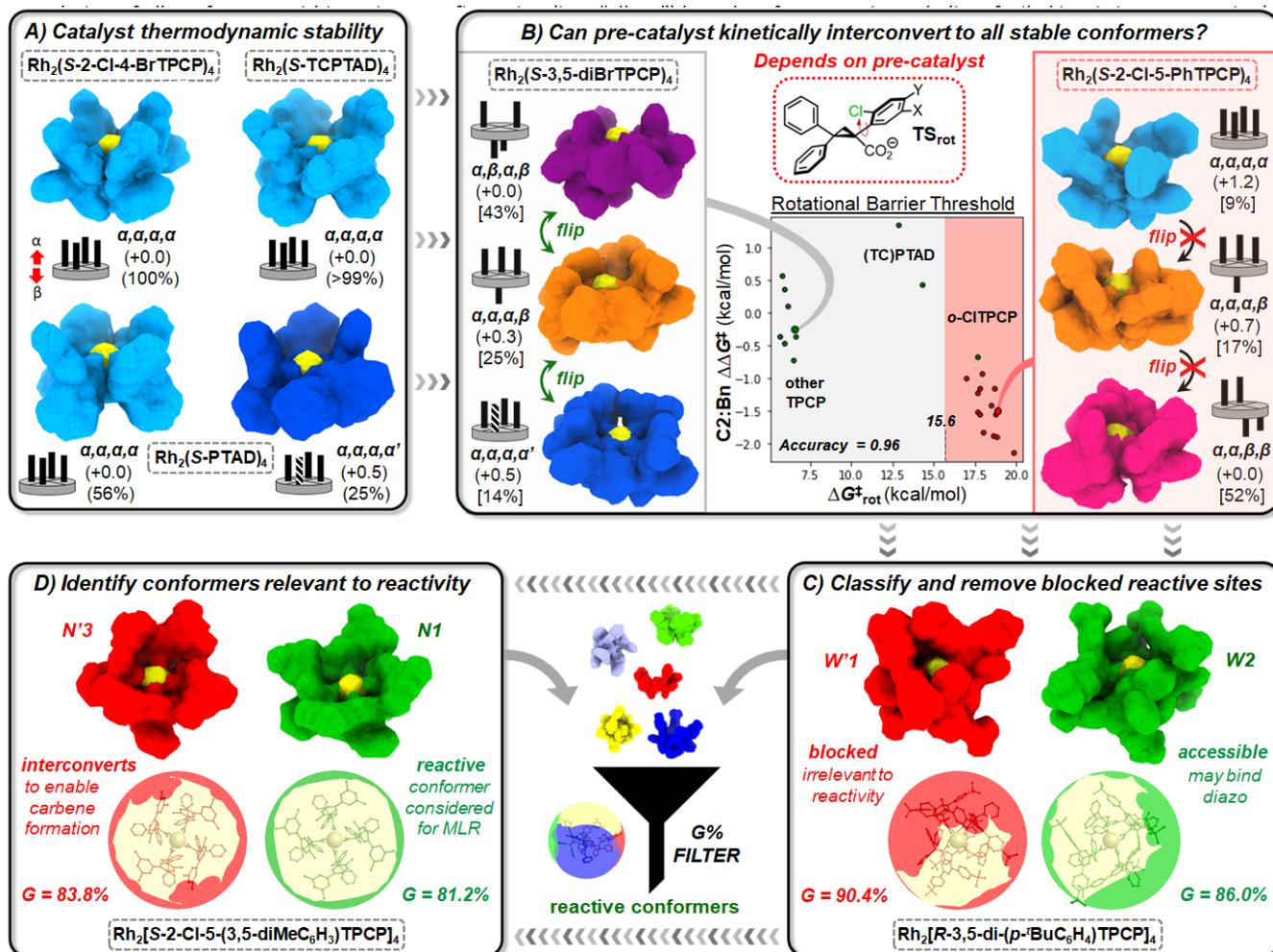
**Figure 2.** Space-filling diagrams of representative Rh<sub>2</sub>(S-2-Cl-4-Ar)TPCP<sub>4</sub> (**G**) and Rh<sub>2</sub>(S-2-Cl-5-Ar)TPCP<sub>4</sub> (**M**) catalysts based on X-ray crystal structures (Ar = *p*-<sup>t</sup>BuC<sub>6</sub>H<sub>4</sub>).

CITPCP)<sub>4</sub> catalysts in the solid-state (Figures S1-S4). Recent studies of Rh<sub>2</sub>(TPCP)<sub>4</sub> catalysts have highlighted the potential for crystal packing effects to give rise to drastic differences between solid-state structures and those relevant for reactivity in solution.<sup>12a</sup> Thus, the accuracy of our QM/MM approach for describing catalyst conformational ensembles in CH<sub>2</sub>Cl<sub>2</sub> was also verified via comparison with structures and relative energetics from previous QM studies (Figure S6).<sup>11b, 12, 23</sup>

The results of our detailed analysis of the ground state catalyst conformations and the interconversion between them are summarized in Figure 3. The “chiral crown” α,α,α,α geometry is unambiguously the most stable for Rh<sub>2</sub>(S-2-Cl-4-XTPCP)<sub>4</sub> and phthalimido catalysts (e.g., **B**, **X**, and **Y** in Figure 3A).<sup>10b, 11b, 12b, 19</sup> Our analysis indicates that several “all-up” conformers are possible for some catalysts (Figures S8-S20). For example, we find **X** to be less structurally rigid than **Y** (Figure 3A), which is consistent with experimental studies for Rh<sub>2</sub>(S-[TC]PTTL)<sub>4</sub> complexes.<sup>19, 24</sup> It is also necessary to consider whether interconversion from each pre-catalyst geometry to the most stable conformers is kinetically feasible under the reaction conditions (Figure 3B).<sup>5h, 11b, 12a, 25</sup> Catalyst ligand flipping transition states are costly to identify, so we instead calculated the related rotational barriers for ligand *M/P* isomer interconversion ( $\Delta G^{\ddagger}_{\text{rot}}$ ) to assess relative catalyst rigidity (Figure S21).<sup>11b</sup> Interestingly, a well-defined classification is observed based on  $\Delta G^{\ddagger}_{\text{rot}}$ ,<sup>26</sup> where Rh<sub>2</sub>(S-*o*-CITPCP)<sub>4</sub> catalysts with rotationally hindered ligands afford higher C2:Bn site-selectivity (Figure 3B, Tables S10-S11).

Reasoning that relative  $\Delta G^{\ddagger}_{\text{rot}}$  values would track with catalyst ligand flipping barriers, our prediction platform uses  $\Delta G^{\ddagger}_{\text{rot}}$  to classify whether each pre-catalyst can interconvert to adopt its most stable configurations (Figure 3B). Notably, on the right side of the single-node  $\Delta G^{\ddagger}_{\text{rot}}$  classification, Rh<sub>2</sub>(S-2-Cl-5-Ar)TPCP<sub>4</sub> pre-catalysts, which have been isolated in α,α,α,α configurations analogous to those of other Rh<sub>2</sub>(S-*o*-

**Figure 3.** Mechanistically guided workflow that accounts for thermodynamic and kinetic considerations to identify relevant catalyst conformers from which to extract molecular descriptors for statistical modeling. The axial Rh site of interest is shown in yellow for each catalyst conformer, along with relative free energies (kcal/mol) and equilibrium populations (% at 313 K in CH<sub>2</sub>Cl<sub>2</sub>). [%] indicates the net



CITPCP)<sub>4</sub> catalysts, are predicted to be unable to interconvert under the reaction conditions to generate more thermodynamically stable  $\alpha, \beta, \alpha, \beta$  and  $\alpha, \alpha, \beta, \beta$  configurations. Interestingly, the isolation of  $Rh_2(S-2-Cl-5-ArTPCP)_4$  catalysts in the  $\alpha, \alpha, \alpha, \alpha$  configuration is likely the dual consequence of their synthesis, via cross-coupling with **I** in its pre-formed  $\alpha, \alpha, \alpha, \alpha$  geometry, and their prohibitive ligand flipping barriers.

#### Classification of Catalyst Conformers Relevant to Reactivity

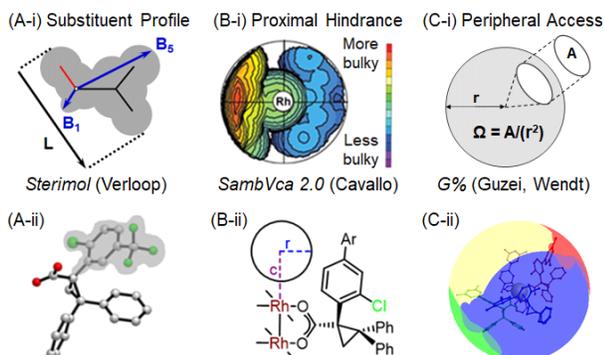
Next, a robust method is required for classifying which  $Rh_2L_4$  catalyst conformers present in solution are relevant to consider for extracting properties that best predict dirhodium-carbene C–H insertion site-selectivity. Ideally, mechanistic insights could be leveraged to globally identify relevant ground state catalyst conformers that closely correspond with C–H insertion transition states. Recently, a thorough computational analysis of stereoselective donor/acceptor carbene C–H insertion catalyzed by  $Rh_2(S-p-BrTPCP)_4$  was reported, building on prior studies by Musaev and Houk.<sup>27</sup> The initial reaction sequence exhibits non-Curtin-Hammett behavior, where dirhodium-carbene formation via N<sub>2</sub> loss is rapid relative to diazo dissociation.<sup>27a</sup> Thus, the catalyst shape to which diazo initially binds will likely closely resemble that of the downstream dirhodium-carbene species that dictates C–H insertion site-selectivity.<sup>30</sup>

To exploit this mechanistic link, a spatial filtering protocol was devised to automatically identify the most reactive catalyst “shapes” toward diazo binding (Figures 3C-D). The kinetic rates of diazo binding to a catalyst of a given electrophilicity should depend primarily on the spatial accessibility of the reactive site to approach, which can be quantified in terms of Guzei and Wendt’s G-parameter (G%).<sup>29, 31</sup> G% can be interpreted as the percentage of vector paths, originating on the surface of the surrounding coordination sphere, by which the diazo precursor cannot approach Rh due to steric blocking. Therefore, only catalyst conformers with lower G% values will be kinetically competent to bind diazo, and consequently the dirhodium-carbene species generated will feature reactive sites that spatially resemble those of these accessible  $Rh_2L_4$  conformers. Based on this premise, we implemented a G% spatial classification filter (see SI for full details). Two illustrative examples of this strategy are depicted (Figures 3C-D). Although stable conformation **W'1** of  $Rh_2[R-3,5-di-(p-BuC_6H_4)TPCP]_4$  is present in the solution-state ensemble, its Rh binding sites are clearly significantly blocked (Figure 3C). In contrast, conformer **W2** is significantly more accessible for diazo to approach. Thus, **W'1** and all other conformers where the reactive site is largely inaccessible were not considered for extracting molecular descriptors for MLR modeling (Figure S20). In a more subtle example, although

conformer **N'3** is relatively accessible (Figure 3D), a methyl group protrudes into the reactive site and restricts the available space for diazo binding. Therefore, it stands to reason that **N'3** would readily interconvert to a more accessible conformer like **N1** to spatially accommodate diazo binding and subsequent dirhodium-carbene formation and reactivity (Figure S14, Tables S3-S5).

#### Surveying Available Steric Descriptors for Catalyst Cavities

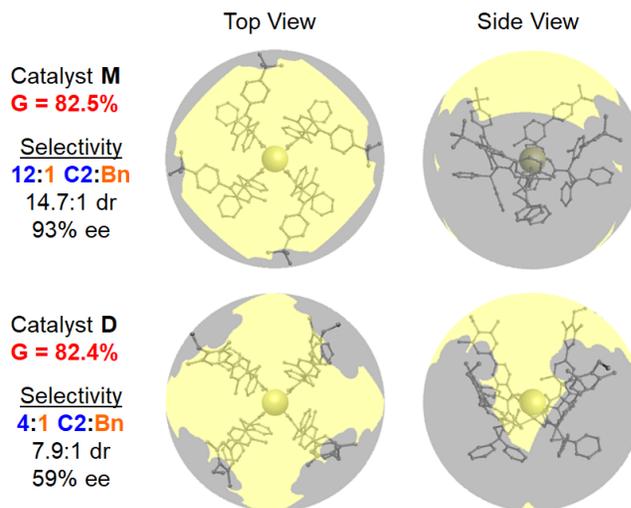
With  $Rh_2L_4$  conformers relevant to reactivity in hand, numerical features describing these catalyst reactive site shapes and electronic environments must next be extracted for MLR modeling. It is particularly challenging to succinctly quantify the spatial constraints imposed by the intricate reactive pockets of these catalysts. Indeed, our initial survey of modern steric descriptors was unsatisfactory for establishing robust correlations with site-selectivity (*vide infra*). Sterimol values<sup>32</sup> are best deployed for describing the spatial profile of a single substituent (Figure 4A) rather than for capturing cavity constraints that arise based on the size and orientation of many substituents. Descriptors based on percent buried volume ( $\%V_{bur}$ ),<sup>33</sup> such as those introduced by Cavallo (*SambVca*, Figure 4B),<sup>33b,33c</sup> are more appropriate for capturing how reactive site topology impacts organometallic catalysis.<sup>26e,34</sup> However, catalyst  $\%V_{bur}$  defined for varying spheres exhibited poor univariate correlations with observed site-selectivity ( $R^2 \leq 0.12$  with  $\Delta\Delta G^\ddagger_{exp}$ ; Figures S23-S24), presumably due to an inability to account for peripheral repulsive interactions (Figure S22).<sup>35</sup> Even  $G\%$  (Figure 4C)



**Figure 4.** Surveying available descriptors for quantifying steric profile and spatial accessibility. Note that  $G\% = 100 \cdot \Omega / (4\pi)$ . Figure B-i was adapted from ref. 29 (<https://pubs.acs.org/doi/10.1021/acs.organomet.6b00371>), copyright 2016 ACS. Further permissions related to the material excerpted should be directed to the ACS.

shows essentially no correlation with site-selectivity ( $R^2 = 0.03$ ; Figure S25).<sup>29</sup>

We hypothesized that  $G\%$  does not correlate with site-selectivity because it is unable to distinguish between different catalyst reactive site shapes. This hypothesis was validated by considering an instructive comparison between the structures of  $Rh_2[S-2-Cl-5-(p\text{-}^tBuC_6H_4)TPCP]_4$  (**M**) and  $Rh_2[S-2-Cl-4-(3,5\text{-}diMe-4\text{-}OMeC_6H_2)TPCP]_4$  (**D**; Figure 5). **M** features a compact bowl shape where substrate is primarily allowed to approach via the top aperture. In contrast,  $Rh_2[S-2-Cl-4\text{-}ArTPCP]_4$  catalysts like **D** have larger, more permeable bowl structures



**Figure 5.** Illustrative comparison of catalysts  $Rh_2[S-2-Cl-5-(p\text{-}^tBuC_6H_4)TPCP]_4$  (**M**) and  $Rh_2[S-2-Cl-4-(3,5\text{-}diMe-4\text{-}OMeC_6H_2)TPCP]_4$  (**D**), which highlights the inability of  $G\%$  to capture shape differences that control reaction output.

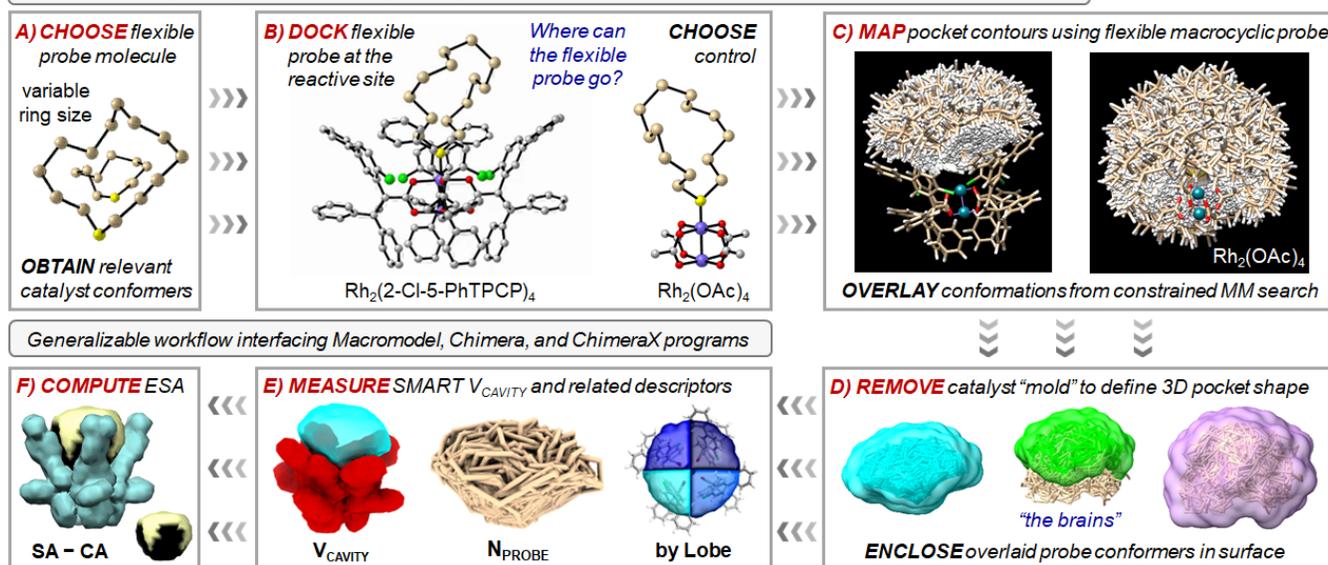
that additionally permit the substrate to approach the dirhodium-carbene via spatial gaps between adjacent ligands. Despite their clear differences in cavity shape and site-selectivity outputs (Figure 5), catalysts **M** and **D** have nearly equivalent net spatial accessibility in terms of  $G\%$  (82.5% and 82.4%, respectively). In addition, electronic descriptors like Rh partial charges and d-orbital energies are nearly identical for the two catalysts (see SI). Therefore, we qualitatively reasoned that the inferior C2:Bn site-selectivity afforded by **D** likely results from its less spatially confined reactive site, which permits the hindered but electronically activated benzylic C-H bond to more readily access the dirhodium-carbene. Moreover, given the inability of  $G\%$ ,  $\%V_{bur}$ , and Sterimol descriptors to capture differences in catalyst cavity shape that impact performance, it became clear that more sophisticated spatial descriptors must be developed.

#### SMART Descriptors for Quantifying Spatial Constraints

Most of the steric parameters surveyed are designed to address the interrelated questions of how bulky a substituent or ligand is and where its bulk resides in space relative to a particular site of interest. The implication of these parameter types in organometallic catalysis is that if the ligand atoms occupy a particular region, then substrate(s) cannot occupy that space to engage in a reaction at a nearby metal center. However, an artifact of this approach is that not all steric bulk in a particular region (e.g.,  $\%V_{bur}$  in an arbitrarily defined sphere) contributes equally, or at all, to confining the reactive site.  $G\%$  elegantly defines accessibility by eliminating these artifacts, but in doing so it obscures whether steric bulk is peripheral or proximal, such that reactive sites with distinctly different shapes may have equivalent net accessibility (*vide supra*).

To complement these existing steric parameters, we sought to develop a descriptor set that quantifies the *empty void space* at a reactive site in a manner that better captures all relevant proximal and peripheral spatial constraints. We have

**Spatial Molding for Approachable Rigid Targets (SMART) Descriptors for Reactive Site Steric Constraints**



**Figure 6.** Workflow for obtaining SMART descriptors for spatial constraints using MacroModel and UCSF Chimera programs.

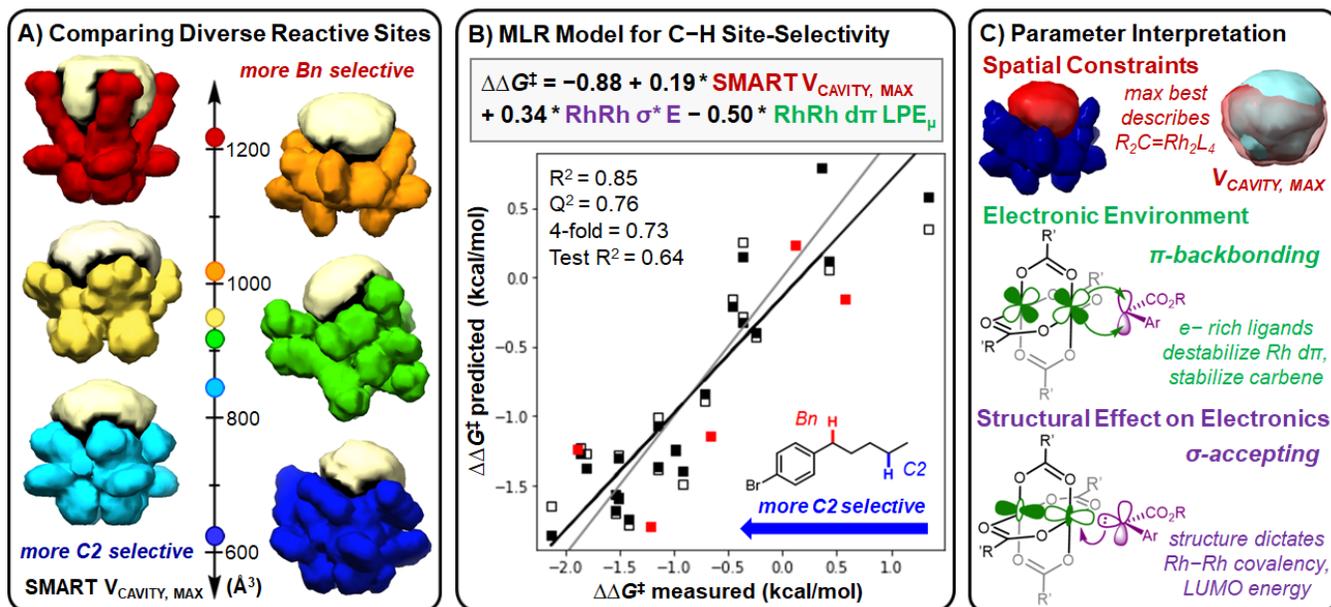
dubbed these descriptors "Spatial Molding for Approachable Rigid Targets", or SMART descriptors (Figure 6). This approach treats the reactive site of each relevant catalyst conformation as an "approachable rigid target" to which a probe molecule can be docked (Figures 6A-B). The spatial constraints of the reactive site are then directly captured by assessing what space the probe molecule can occupy in a constrained conformational search where the catalyst atoms are frozen (Figure 6C; MacroModel<sup>20a</sup>/OPLS3e<sup>20b</sup>). This workflow is illustrated for  $\text{Rh}_2(\text{S-2-Cl-5-PhTPCP})_4$  (**L**) in comparison to a  $\text{Rh}_2(\text{OAc})_4$  control catalyst with minimal axial steric hindrance. A macrocyclic thioether probe was employed (Figure 6A), which is flexible enough to map the contours of the catalyst pocket and exhibits maximal sensitivity to changes in catalyst shape (see SI). The overlaid conformations of the macrocyclic probe can then be enclosed within a 3D surface using UCSF ChimeraX 1.1,<sup>36</sup> which allows for visualization and quantification of all possible space that the probe can occupy given the spatial constraints of the catalyst pocket (Figure 6D).

From there, various SMART features can be tabulated and utilized as inputs for relating feature space to catalyst performance via statistical modeling. For example,  $V_{\text{CAVITY}}$  is the volume in  $\text{\AA}^3$  of the surface enclosing all probe conformers, and it provides a direct measure of reactive site spatial constraints (Figure 6E). In comparison to  $\text{Rh}_2(\text{OAc})_4$ , where the probe is free to move with minimal constraints and the resulting surface enclosing all probe conformers has a relatively large volume, bulky dirhodium catalysts like **L** have significantly smaller  $V_{\text{CAVITY}}$  values. Hence,  $V_{\text{CAVITY}}$  can be interpreted as a quantitative scale for the degree of reactive site spatial confinement. Conceptually, tabulating  $V_{\text{CAVITY}}$  descriptors across catalyst conformers is somewhat related to the inverse of average steric occupancy (ASO) descriptors recently pioneered by Denmark.<sup>16e</sup> However, an important difference is that not all void space is relevant to consider as space in which a reaction may occur. By leveraging the space that a probe

molecule can occupy as the basis for defining  $V_{\text{CAVITY}}$ , irrelevant regions within a given radius of the reactive site that are inaccessible for accommodating incoming substrates are rightly excluded from consideration (i.e., channels, crevices). Notably, this approach also provides a related  $N_{\text{PROBE}}$  descriptor (Figure 6E), where a smaller number ( $N$ ) of probe conformations are possible for more confined reactive sites.

Building upon the cavity surface defined to compute  $V_{\text{CAVITY}}$ , a new feature, entry surface area (ESA), was developed to quantify the absolute accessibility of the catalyst reactive site. While  $G\%$  also captures accessibility, it is a relative metric. Thus, among two catalysts with equivalent  $G\%$ , one may have a larger absolute surface area through which substrate can enter the reactive site (Figure S42). We hypothesized that this absolute "entry surface area" (ESA in  $\text{\AA}^2$ ) could capture how accessible the catalyst pocket is for an approaching substrate of a given size, which is a somewhat similar idea to a previous effort to quantify an "accessible molecular surface."<sup>37</sup> ESA was tabulated by first generating a catalyst surface, followed by measuring the contact area between the catalyst surface and the cavity surface (see SI for details).<sup>36, 38</sup> This contact area (CA; shaded black in Figure 6F) represents the surface area on the catalyst pocket through which substrate cannot enter due to repulsive interactions with the ligands. ESA is computed by subtracting this contact area (CA) from the total surface area (SA) of the cavity (Figure 6E). It should be noted that both ESA and  $V_{\text{CAVITY}}$  can be computed for a quadrant, or "lobe", of the full cavities ("the brains"), or in terms of their deviation across lobes to capture pocket asymmetry. Additionally, SMART descriptors can be computed as Boltzmann-average, maximum, minimum, delta ( $\Delta$ ), or weighted standard deviation values by considering parameters and their variance across all relevant catalyst conformations.

In essence, SMART descriptors represent the adaptation for small molecule catalysts of a host of previous methods



**Figure 7.** (A) Cavity confinement visual scale for catalysts **E**, **S**, **Y**, **W**, **J**, and **O** in descending  $V_{\text{CAVITY, MAX}}$  order (**E** =  $\text{Rh}_2[\text{S-2-Cl-4-(p-NMe}_2\text{C}_6\text{H}_4)\text{TPCP}]_4$ , **S** =  $\text{Rh}_2[\text{S-}i>p\text{-BrTPCP}]_4$ , **Y** =  $\text{Rh}_2[\text{S-TCPTAD}]_4$ , **W** =  $\text{Rh}_2[\text{R-3,5-di}(p\text{-}^t\text{BuC}_6\text{H}_4)\text{TPCP}]_4$ , **J** =  $\text{Rh}_2[\text{S-2-Cl-5-CF}_3\text{TPCP}]_4$ , **O** =  $\text{Rh}_2[\text{S-2-Cl-5-MesTPCP}]_4$ ). (B) MLR model for catalyst-controlled C-H site-selectivity, where greater C2 site-selectivity is defined as a more negative  $\Delta\Delta G^\ddagger$  value. An 80:20 training/validation partitioning was performed to yield equidistant site-selectivity outputs. Internal cross-validation was performed with  $K=4$  ( $n=20!/\{5! \cdot 15!\}$ ), while hollow squares indicate leave-one-out (LOO) predictions, the precision of which are reflected in  $Q^2$ . (C) Chemical interpretation of the parameters in the model.

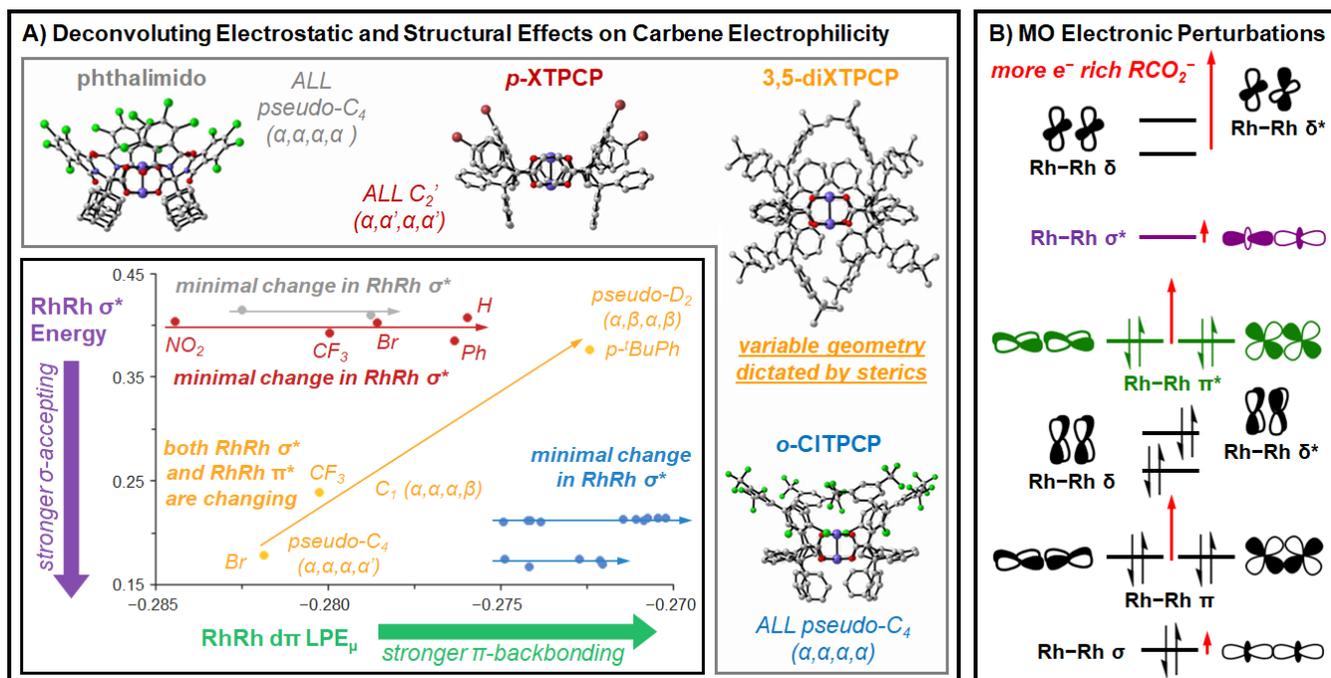
developed for analyzing protein cavities and channels.<sup>39</sup> An important difference is that the reactive sites of small molecule catalysts are generally much less spatially enclosed than binding sites in proteins and enzymes, the latter of which are often defined based on cavity filling with small spheres that can access spatial regions that slightly larger spheres cannot.<sup>39g, 40</sup> This definition is not as applicable to small molecule catalysts, as even highly sterically hindered reactive sites like those of the dirhodium catalysts described here have apertures through which very large spheres could fit. Alternatively, many less hindered catalysts do not have true cavities (e.g.,  $\text{Rh}_2\text{OAc}_4$ ), such that spheres of nearly all sizes can access their reactive sites.<sup>41</sup> Furthermore, grid-based methods for defining cavities in biology can suffer from the challenges presented by the so-called “mouth opening ambiguity” (MOA), which stems from the inherent difficulty involved in defining where a shallow cavity ends and where the exterior void space begins.<sup>39g, 41-42</sup> This can be especially problematic for describing reactive sites with irregular shapes. By defining the reactive pocket to be anywhere that a given docked probe molecule can reside, SMART descriptors provide a minimally arbitrary MOA solution that is generalizable for directly comparing the spatial constraints imposed across a diverse set of reactive sites. Thus, SMART can allow disparate catalysts to be unified in a single model (Figures 7A-B), which should enable more reliable out-of-sample extrapolations for *in silico* screening applications.

#### Statistical Model for Catalyst-Controlled Site-Selectivity

A MLR model was subsequently pursued to understand the relative influence of cavity spatial constraints on C-H functionalization site-selectivity in comparison to the effects of

other catalyst properties. A statistically robust model was found that features one spatial descriptor (SMART  $V_{\text{CAVITY, MAX}}$ ) and two electronic descriptors capturing carbene stabilization by the dirhodium catalyst via both  $\pi$ -backbonding and  $\sigma$ -accepting interactions (Figures 7B-C). The model was obtained using a forward-stepwise search algorithm<sup>15, 43</sup> with 117 normalized descriptors for each catalyst as the input.<sup>44</sup> Catalyst descriptors included 18 spatial descriptors (SMART, G%, Sterimol, % $V_{\text{bur}}$ ) and 13 electronic descriptors derived from Natural Bond Orbital (NBO) analysis<sup>45</sup> (Rh charges, *d*-orbital energies), with additional descriptors included based on the maximum, minimum,  $\Delta$ , and weighted standard deviation values of these properties among all reactive conformers for each catalyst (Table S12). High internal cross-validation  $Q^2$  and  $K$ -fold metrics that approach  $R^2$ ,<sup>46</sup> along with an adequate test  $R^2$  despite a high data-to-parameter ratio,<sup>47</sup> collectively indicate that the model is not the result of overfitting the training set data.<sup>48</sup> Additionally, the model retains similar parameter coefficients and statistical metrics for alternative Kennard-Stone<sup>49</sup> training/validation splits (Figure S30), indicating that the favorable performance of the model is not due to fortuitous training set partitioning.

Beyond its favorable performance, the model is also chemically interpretable and enables the relative contributions of intuitive catalyst properties to be explicitly assessed (Figure 7C). Catalyst spatial constraints are accounted for in terms of SMART  $V_{\text{CAVITY, MAX}}$  values, which are shown for representative catalysts with varying cavity shapes in Figure 7A. The positive coefficient for  $V_{\text{CAVITY, MAX}}$  in the model signifies that catalysts with more spatially restricted reactive sites (i.e., smaller absolute cavity volumes, more negative normalized z-scores)



**Figure 8.** Deconvoluting catalyst electronic effects in the MLR model via qualitative electronic structure perturbation analysis.

will afford increased C2:Bn site-selectivity (i.e., more negative  $\Delta\Delta G^\ddagger$  value from the product of the positive coefficient and more negative z-score). This is consistent with our qualitative hypothesis that more sterically hindered catalysts will be less readily able to facilitate C-H functionalization at the more hindered benzylic site. Furthermore, the *maximum*  $V_{CAVITY}$  values across the reactive conformers of each catalyst were found to best capture the relative spatial constraints. Presumably,  $Rh_2L_4$  catalysts will adopt conformations most similar to those that maximize void space near their reactive sites to accommodate carbene formation and substrate approach in the selectivity-determining transition states. This principle can be better appreciated by visually overlaying the largest (red) and smallest (blue) cavities among the reactive conformations for a given catalyst, as displayed for **T** in Figure 7C. Additionally, a similar MLR model can also be obtained wherein  $V_{CAVITY, MAX}$  is replaced with a SMART  $N_{PROBE, MAX}$  descriptor (Figure S31), indicating that increased C2 site-selectivity is afforded by catalysts for which a smaller maximum number of probe orientations are possible within their reactive sites (*ceteris paribus*).

Based on the model coefficient magnitudes, cavity spatial constraints are not as influential as carbene electronic stabilization in controlling site-selectivity across this disparate catalyst set. The most important catalyst property is RhRh  $d\pi$  LPE $_{\mu}$ , which is the average energy of Rh  $d_{xz}/d_{yz}$  orbitals obtained from NBO analysis.<sup>45</sup> NBO describes dirhodium electronic structure to consist of  $\sigma$ -bonding (filled) and  $\sigma$ -antibonding (empty) orbitals derived from  $d_z^2-d_z^2$  overlap, along with three localized lone-pairs at each Rh ( $d_{xy}$ ,  $d_{xz}$ ,  $d_{yz}$ ). This treatment does not consider mixing of localized Rh  $d\pi$  orbitals to engender Rh-Rh  $\pi$ -bonding (filled) and  $\pi$ -antibonding (filled) orbitals, the latter of which are the relevant frontier molecular

orbitals that engage in  $\pi$ -backbonding interactions to stabilize the singlet carbene fragment (Figure 7C). Nevertheless, higher energy localized Rh  $d\pi$  orbitals will give rise to higher energy RhRh  $\pi^*$  orbitals upon mixing that have more favorable energetic overlap for  $\pi$ -backbonding interactions with the conserved donor/acceptor carbene moiety. The negative model coefficient for RhRh  $d\pi$  LPE $_{\mu}$  signifies that catalysts that better stabilize the carbene via  $\pi$ -backbonding (i.e., higher energy Rh  $d\pi$ , more positive z-score) will afford greater C2 site-selectivity (i.e., more negative  $\Delta\Delta G^\ddagger$ ). In other words, catalysts that better quench latent carbene electrophilicity will afford more stable dirhodium-carbenes with larger C-H insertion transition state barriers, which in this case results in greater discrimination between the C2 and Bn C-H bonds in accordance with the oft-maligned reactivity-selectivity principle.<sup>3k, 5h, 50</sup> Similarly, catalysts with a lower energy RhRh  $\sigma^*$  orbital (i.e., more negative z-score multiplied by a positive coefficient) will better stabilize the carbene via  $\sigma$ -accepting interactions (Figure 7C), and will therefore afford greater C2 site-selectivity by the analogous rationale.

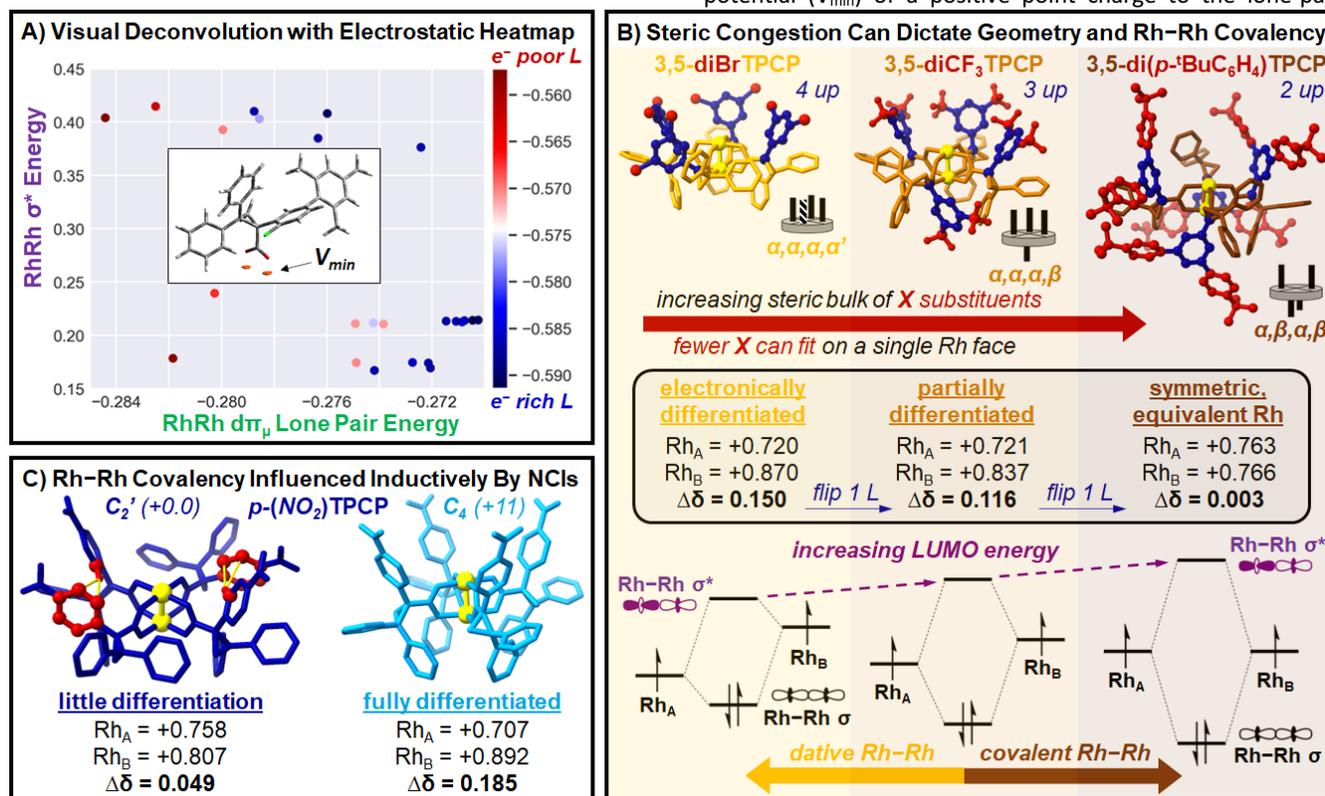
#### Deconvoluting Dirhodium-Carbene Electrophilicity Effects

The relative importance of catalyst electronic properties in dictating site-selectivity was somewhat surprising given our initial qualitative hypotheses. Additionally, we were intrigued by the feature independence of RhRh  $\sigma^*$  E and RhRh  $d\pi$  LPE $_{\mu}$  ( $R^2 = 0.35$ )<sup>51</sup> and wondered how each of these electronic descriptors relates to our intuitive conception of electron-richness in physical organic chemistry. Thus, we proceeded to further deconvolute their precise chemical meaning by visualizing where catalysts in each of the four subclasses lie on a plot of RhRh  $\sigma^*$  E vs. RhRh  $d\pi$  LPE $_{\mu}$  (Figure 8A). Strikingly, the RhRh  $\sigma^*$  energy remains largely invariant for catalysts within the same subclass for phthalimido, *p*-XTPCP, and *o*-CITPCP

catalysts. This can be most clearly observed for the *p*-XTPCP subclass (red), where serial variation of the *para* substituent impacts the relative propensity for  $\pi$ -backbonding across the catalyst set without significantly perturbing the RhRh  $\sigma^*$  energy and the associated catalyst  $\sigma$ -accepting capability (Figure 8A). The same trend holds true within the phthalimido (gray) and *o*-CITPCP catalyst (blue) subclasses, albeit with two slightly offset horizontal lines observed for the latter depending on whether the *o*-CITPCP ligand has 5-Ar substitution.

(Figure 8B). Thus, the observation that varying carboxylate electronics (e.g., across the *p*-XTPCP series) impacts Rh–Rh  $\pi^*$  energy to a much greater extent than Rh–Rh  $\sigma^*$  energy is borne out by electronic structure analysis.

Further visual clarity for the deconvolution of dirhodium catalyst electronic properties as a function of ligand electron-richness is provided by the heatmap in Figure 9A. The colormap quantifies carboxylate ligand electron-richness in terms of the minimum molecular electrostatic attraction potential ( $V_{\min}$ ) of a positive point charge to the lone-pair



**Figure 9.** (A) Visual deconvolution of catalyst electronics using a carboxylate ligand electrostatic heatmap. (B–C) Effects of steric crowding and NCIs on Rh–Rh covalency, which dictates catalyst propensity for carbene stabilization via  $\sigma$ -accepting interactions.

These observations can be reconciled by considering the electronic structure of dirhodium catalysts (Figure 8B).<sup>52</sup> Each Rh fragment possesses square-planar geometry with respect to the carboxylate ligand field and engages in Rh–Rh  $\sigma$ -,  $\pi$ -, and  $\delta$ -interactions via the overlap of  $d_z^2$ ,  $d_{xz}/d_{yz}$ , and  $d_{xy}/d_{x^2-y^2}$  orbitals, respectively. A net Rh–Rh single bond ( $\sigma$ ) results from in-phase  $d_z^2$  overlap, with otherwise equal population of metal-metal bonding and antibonding orbitals of  $\pi$  and  $\delta$  symmetry. The electronic perturbations of more electron-rich carboxylates, such as moving from *p*-NO<sub>2</sub>TPCP to the parent TPCP ligand, are shown with red arrows in Figure 8B. More electron-rich carboxylates will act as stronger  $\sigma$ -donors, which will significantly destabilize the unfilled Rh–Rh  $\delta$  and  $\delta^*$  orbitals derived from weak  $d_{x^2-y^2}$  orbital overlap. Stronger carboxylate  $\sigma$ -donation will also slightly destabilize the Rh–Rh  $\sigma$  and  $\sigma^*$  orbitals via the modest interaction of the ligand field with the torus of each  $d_z^2$  orbital. Carboxylates are also excellent  $\pi$ -donors, and so more electron-rich carboxylates will significantly destabilize the Rh–Rh  $\pi$  and  $\pi^*$  orbitals as well

region of each carboxylate oxygen atom.<sup>53</sup>  $V_{\min}$  has been shown to quantitatively describe the electron-donor propensity of phosphine ligands.<sup>54</sup> A clear visual stratification can be observed in Figure 9A, where horizontal movement along the RhRh  $dT_{\mu}$  LPE <sub>$\mu$</sub>  axis separates more electron-poor carboxylate ligands (red points) from more electron-rich ones (blue points). Importantly, free ligand electron-richness does not appear to have any bearing on the  $\sigma$ -accepting propensity of the corresponding catalyst (i.e., the vertical RhRh  $\sigma^*$  energy coordinate in Figure 9A).

Having established that RhRh  $\sigma^*$  energy is largely independent of carboxylate ligand electronic properties, we hypothesized that catalyst  $\sigma$ -accepting propensity may be primarily dictated by global catalyst structural effects. In support, catalyst structure is invariant within each of the subclasses where RhRh  $\sigma^*$  energy remains constant (Figure 8A), as *p*-XTPCP catalysts universally adopt *C*<sub>2'</sub> symmetric structures ( $\alpha, \alpha', \alpha, \alpha'$ ) and both phthalimido and *o*-CITPCP

catalysts adopt pseudo- $C_4$  symmetric structures ( $\alpha, \alpha, \alpha, \alpha$ ).<sup>5h, 12a</sup> The only catalyst subclass within which RhRh  $\sigma^*$  and  $\pi^*$  energies *both change* is the 3,5-diXTPCP series, and, intriguingly, the major structural configuration also changes as X is varied from Br to CF<sub>3</sub> to *p*-<sup>t</sup>BuC<sub>6</sub>H<sub>4</sub> (Figure 8A). Within this subclass, we find that as the steric bulk of X increases, less X substituents prefer to be oriented above a single Rh face. This is illustrated in Figure 9B, where the major reactive catalyst conformer shifts from “all-up” for X = Br ( $\alpha, \alpha, \alpha, \alpha'$ ) to “three-up” for X = CF<sub>3</sub> ( $\alpha, \alpha, \alpha, \beta$ ) to an alternating “two-up”  $D_2$  symmetric structure for X = *p*-<sup>t</sup>BuC<sub>6</sub>H<sub>4</sub> ( $\alpha, \beta, \alpha, \beta$ ). These structural preferences seem to be dictated by the size of the X substituents (Figures S18-S20). Indeed,  $\alpha, \alpha, \alpha, \beta$  and  $\alpha, \alpha, \alpha, \alpha$  structures for Rh<sub>2</sub>[*R*-3,5-di(*p*-<sup>t</sup>BuC<sub>6</sub>H<sub>4</sub>)TPCP]<sub>4</sub> (**W**) are highly unstable because *p*-<sup>t</sup>BuC<sub>6</sub>H<sub>4</sub> groups are simply too large for three or four ligands to fit about one Rh face without requiring significant structural distortions to minimize steric clashes.<sup>10c</sup>

A subtle consequence of these global structural preferences is that the two Rh atoms can be locally electronically differentiated to varying extents depending on the degree of asymmetry in the preferred catalyst structure. For example, the  $D_2$  symmetric ( $\alpha, \beta, \alpha, \beta$ ) structure of **W** features electronically equivalent Rh atoms, as quantified in terms of the nearly identical computed Rh partial charges from NBO analysis (Figure 9B).<sup>45</sup> Upon flipping one ligand to adopt a  $C_1$  ( $\alpha, \alpha, \alpha, \beta$ ) structure, Rh<sub>2</sub>(*S*-3,5-diCF<sub>3</sub>TPCP)<sub>4</sub> (**U**) features more electronically differentiated Rh atoms ( $\Delta\delta = 0.116$ ). Upon flipping another ligand, maximum electronic differentiation across the subclass is observed for the “all-up” structure of Rh<sub>2</sub>(*S*-3,5-diBrTPCP)<sub>4</sub> (**V**;  $\Delta\delta = 0.150$ ). We propose that this electronic differentiation of the two Rh atoms plays a critical role in dictating the propensity of the dirhodium catalyst to stabilize the donor/acceptor carbene moiety via  $\sigma$ -accepting interactions (RhRh  $\sigma^*$  E in the MLR model). Specifically, electronically identical Rh fragments will have favorable energetic  $d_z^2$  orbital overlap for  $\sigma$ -bonding, which will result in highly covalent Rh–Rh bonding with maximal bonding stabilization (RhRh  $\sigma$ ) and antibonding destabilization (RhRh  $\sigma^*$ ; Figure 9B). As the two Rh become more electronically differentiated, Rh–Rh bonding will become more dative (i.e., partial donor/acceptor character between an electron-rich and an electron-poor Rh). Rh electronic differentiation manifests in less favorable energetic overlap for Rh–Rh  $\sigma$ -bonding, which will give rise to a lower energy catalyst RhRh  $\sigma^*$  orbital that can better stabilize the lone pair of the singlet carbene moiety.

While all Rh–Rh bonds are highly covalent in comparison to heterometallic metal-metal bonds,<sup>55</sup> clearly this subtle continuum toward more dative Rh–Rh bonding leads to dramatic electronic effects on reactivity that have largely been underappreciated when considering how to rationally design a selective dirhodium catalyst. That we were able to elucidate this subtle relationship between global structure and dirhodium-carbene electrophilicity<sup>56</sup> speaks to the power of the multivariate modeling approach when molecular descriptors that closely quantify the underlying catalyst properties that control reactivity are utilized as numerical

inputs. Interestingly, the RhRh  $\sigma^*$  energy for a given Rh<sub>2</sub>(*S*-*o*-CITPCP)<sub>4</sub> catalyst increases markedly upon distorting its  $C_4$  symmetric ( $\alpha, \alpha, \alpha, \alpha$ ) structure to a  $D_2$  symmetric ( $\alpha, \beta, \alpha, \beta$ ) structure with equivalent Rh atoms (Figure S29). Unlike TPCP ligands, the properties of phthalimido ligands are such that even for the maximally electronically differentiated  $\alpha, \alpha, \alpha, \alpha$  structures that these catalysts adopt, the phthalimido and adamantyl Rh faces are only slightly electronically differentiated (Rh  $\Delta\delta < 0.02$ ). Thus, despite their  $\alpha, \alpha, \alpha, \alpha$  structures, Rh<sub>2</sub>(*S*-TCPTAD)<sub>4</sub> and Rh<sub>2</sub>(*S*-PTAD)<sub>4</sub> feature highly covalent Rh–Rh bonding that gives rise to relatively high energy RhRh  $\sigma^*$  orbitals (Figure 8A).

In addition to arising from catalyst structural symmetry effects, it is worth noting that Rh electronic differentiation can also result from inductive effects stemming from noncovalent interactions (NCIs) between neighboring ligands. This scenario is depicted in Figure 9C, where the Rh atoms in Rh<sub>2</sub>(*p*-NO<sub>2</sub>TPCP)<sub>4</sub> become substantially more electronically similar upon structural distortion from an unstable  $C_4$  symmetric ( $\alpha, \alpha, \alpha, \alpha$ ) structure to the preferred  $C_2'$  configuration ( $\alpha, \alpha', \alpha, \alpha'$ ) that features (C–H)– $\pi$  interactions between neighboring ligands with close computed contacts of 2.90 Å. This structural configuration is preserved across the *p*-XTPCP catalyst series,<sup>12a</sup> with the electron-density sequestered in these NCIs presumably inductively diverted from the Rh atom on the proximal face (Figure S29). Collectively, this in-depth analysis has clarified the chemical meaning of the RhRh  $\sigma^*$  E term in the MLR model by uncovering an intimate relationship between global catalyst structure and Rh–Rh covalency, which clearly plays a larger role in promoting selective catalysis via carbene stabilization than has previously been understood.

#### *Influence of Structural Rigidity on Site- and Stereoselectivity*

Catalyst structural dynamics are also undoubtedly important to performance, and these considerations have not yet been emphasized. Sufficient conformational flexibility is needed to allow the dirhodium-carbene species to form and react with substrate amid the highly congested catalyst pocket, but conformational rigidity is also critical to ensure a consistently defined cavity for facilitating selective product formation. The only term that captures catalyst rigidity to any extent in the MLR model is SMART  $V_{\text{CAVITY, MAX}}$ , as catalysts that are more rigid cannot adopt more open conformations with less spatial restrictions on reactivity. We reasoned that among catalysts that are electronically similar, such as the set of 15 Rh<sub>2</sub>(*S*-*o*-CITPCP)<sub>4</sub> catalysts, increased rigidity would be a defining feature that differentiates catalysts that afford more selective C–H functionalization. This hypothesis is supported by the univariate correlation among Rh<sub>2</sub>(*S*-*o*-CITPCP)<sub>4</sub> catalysts between site-selectivity  $\Delta\Delta G^\ddagger$  and  $\Delta G^\ddagger_{\text{rot}}$ , the *M/P* ligand rotational barrier in the previously discussed threshold classification ( $R^2 = 0.46$ ; Figure 3B).

To further probe this hypothesis for the entire catalyst set, we consolidated the two electronic terms in the model into a single RhRh  $\Delta E_{\pi^*/\sigma^*}$  parameter, where a smaller energetic gap between the frontier RhRh  $\pi^*$  and  $\sigma^*$  orbitals is ideal for carbene stabilization via both  $\pi$ -backbonding and  $\sigma$ -accepting.<sup>57</sup> Upon re-modeling with this additional descriptor (Figure 10), a model with improved internal cross-validation ( $Q^2 = 0.79$ , 4-fold = 0.77) and external validation metrics (test  $R^2 = 0.87$ ) was identified that captures dirhodium-carbene electrophilicity (RhRh  $\Delta E_{\pi^*/\sigma^*}$ ), reactive site spatial constraints (SMART  $N_{\text{PROBE, LE}}$ ), and catalyst rigidity (SMART ESA  $\Delta$ ). Catalyst rigidity is quantified by SMART ESA  $\Delta$ , or the change in entry surface area across reactive catalyst conformers. Rigid catalysts with relatively constant ESA across possible reactive conformers were found to afford higher C2 site-selectivity, as were catalysts with more confined reactive sites (i.e., fewer SMART  $N_{\text{PROBE, LE}}$ ) and greater propensity for carbene stabilization (i.e., smaller RhRh  $\Delta E_{\pi^*/\sigma^*}$ ). Although exceedingly high rigidity could hinder catalysis in principle, this limit has evidently not been approached within this catalyst set.

It is also important to note that these same catalyst properties also largely dictate diastereoselectivity as well, since  $\Delta\Delta G^\ddagger$  for C2 product dr is highly correlated with that for C2:Bn site-selectivity ( $R^2 = 0.83$ ; Figure S27). Although C2 enantioselectivity and C2:Bn site-selectivity are not correlated ( $R^2 = 0.04$  for respective  $\Delta\Delta G^\ddagger$ ; Figure S28), threshold analysis indicates that increased catalyst rigidity to provide a more well-defined reactive site is also beneficial for stereoselective catalysis (Figure S38). We envision that SMART descriptors capturing catalyst pocket dynamics will be useful alternatives to traditional steric and RMSD-based rigidity descriptors,<sup>58</sup> as

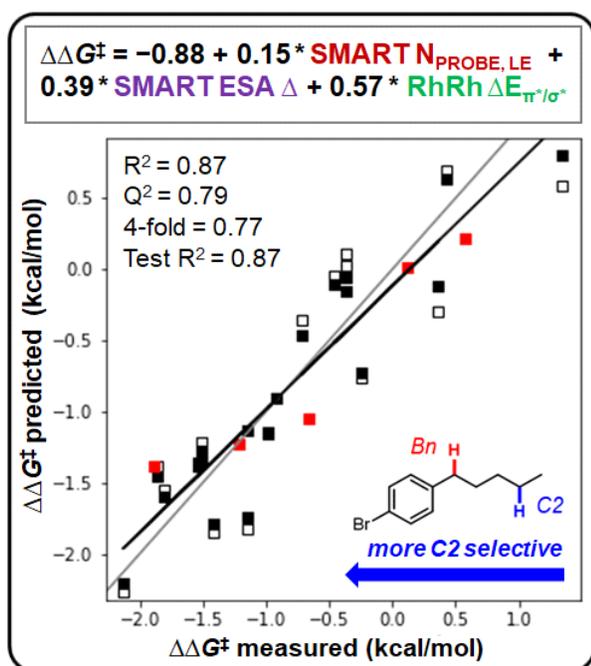
not all changes in atomic positions across conformers represent pertinent changes in reactive site shape.

#### Comparing Models: SMART vs. Other Steric Descriptors

When introducing a new set of molecular descriptors, it is prudent to explicitly compare their strengths and predictive utility to established parameters so that others may appropriately apply them (Table S13).<sup>59</sup> Unsurprisingly, models based on free ligand steric and electronic parameters (**mD-mE**) were of decreased quality and interpretability compared to the catalyst descriptor model in Figure 10 (**mA**), especially in terms of test set predictions ( $\Delta\text{test } R^2 \leq -0.21$ ; Figure S32). Likewise, replacing SMART steric and rigidity descriptors in model **mA** with the best Sterimol combination ( $B1_{R-COO}$  and  $B5_{R-COO} \Delta$ ) results in poor fitting of the training set data (**mF**:  $\Delta R^2 = -0.14$ ,  $\Delta Q^2 = -0.18$ ,  $\Delta 4\text{-fold} = -0.19$ ; Figure S33). The best  $\%V_{\text{bur}}$  combination for catalyst sterics ( $\%V_{\text{bur}, 4.0c, 4.5r}$ ) and rigidity ( $\%V_{\text{bur}, 4.0c, 4.5r} \Delta$ ) performs better than Sterimol (c.f., **mG** to **mF**), but still offers decreased predictive power relative to SMART model **mA** ( $\Delta R^2 = -0.05$ ,  $\Delta Q^2 = -0.08$ ,  $\Delta 4\text{-fold} = -0.11$ ,  $\Delta\text{test } R^2 = -0.14$ ; Figure S33). Combining a descriptor for peripheral rigidity (e.g.,  $G\%$   $\Delta$ , Sterimol  $\Delta$ , or  $\Delta G_{\text{rot}}^\ddagger$ ) with  $\%V_{\text{bur}}$  (proximal sterics) and RhRh  $\Delta E_{\pi^*/\sigma^*}$  (catalyst electronics) also does not lead to improved models (**mI-mK**; Figure S34-S35).

Examining the performance of  $\%V_{\text{bur}}$  further, 14-20 of the 25 catalysts have nearly identical values (i.e.,  $\pm 1\%$ ; depending on sphere position and radius) due to the presence of essentially no proximal ligand atoms directly within the axial cavity (Figures S22 and S39-S40). This inability to differentiate catalysts with little proximal steric hindrance likely explains the larger losses in predictive utility in terms of test  $R^2$  when  $\%V_{\text{bur}}$  descriptors are used. Despite this limitation,  $\%V_{\text{bur}}$  can be successfully deployed in combination with SMART peripheral rigidity descriptors like SMART ESA  $\Delta$  (**mM**; Figure S36). This illustrates that, although SMART  $N_{\text{PROBE}}$  and  $V_{\text{CAVITY}}$  parameters are more descriptive than  $\%V_{\text{bur}}$  for unhindered catalysts (Figures S39-S40), the most unique effects captured by SMART descriptors are described by ESA-type parameters (Figures S42-S43). Indeed, SMART ESA  $\Delta$  cannot be replaced in model **mA** by any other traditional descriptor (e.g.,  $\%V_{\text{bur}} \Delta$  or  $G\%$   $\Delta$ ) without precipitous drops in utility (**mN-mO**:  $\Delta Q^2 \leq -0.11$ ,  $\Delta 4\text{-fold} \leq -0.12$ ,  $\Delta\text{test } R^2 \leq -0.22$ ; Figure S37). This highlights that, relative to traditional steric parameters, SMART analysis enables access to a larger suite of more generalizable, *continuous-scale* descriptors, any one of which may enable reliable predictions for a given reaction.

Lastly, it is notable that model **mA** provides favorable performance by focusing on repulsive interactions in lieu of explicitly accounting for the potential influence of attractive NCIs. We tentatively hypothesize that NCIs between the substrate benzyl moiety and the aryl group of the dirhodium-carbene (e.g.,  $\pi$ - $\pi$  or C-H- $\pi$ ) could act as a minor perturbation to divert C-H insertion away from the benzylic site. Future studies may probe these subtle effects via transition state analysis, as well as by designing an expanded experimental data matrix for MLR modeling that pairs the catalyst diversity



**Figure 10.** Improved MLR model accounting for catalyst rigidity effects. Note that SMART  $N_{\text{PROBE, LE}}$  denotes the number of low energy probe conformers ( $\leq 5$  kcal/mol).

explored here with an array of substrates and diazo precursors with varying propensity to engage in attractive NCl<sub>s</sub>.

## Conclusions

A series of Rh<sub>2</sub>(*S*-*o*-CITPCP)<sub>4</sub> catalysts with systematic perturbations was designed and found to adopt C<sub>4</sub> symmetric bowl-shaped structures with varying depth and permeability. Together with a diverse set of Rh<sub>2</sub>(TPCP)<sub>4</sub> and Rh<sub>2</sub>([TC]PTAD)<sub>4</sub> catalysts, their performance in terms of C–H functionalization site-selectivity between the terminal methylene (C2) and benzylic (Bn) sites in 1-bromo-4-pentylbenzene was evaluated, with Rh<sub>2</sub>(*S*-2-Cl-5-MesTPCP)<sub>4</sub> discovered to be optimal for stereoselective C2 functionalization (30.9:1 rr, 14:1 dr, 87% ee). Based on evaluating the performance of this expanded catalyst library, intuitive quantitative structure-selectivity relationships were uncovered by performing MLR modeling using newly developed SMART descriptors for reactive site spatial constraints. Rigid catalysts with more confined reactive sites were found to afford greater site-selectivity for functionalizing the more sterically accessible C2 position. Additionally, catalyst electronic stabilization of the singlet donor/acceptor carbene fragment via increased  $\pi$ -backbonding and  $\sigma$ -accepting interactions was found to play a comparatively greater role in affording high C2 site-selectivity and dr. Although broad generalization of the reactivity-selectivity principle is an enduring myth,<sup>50</sup> clearly it still may often hold true for catalytic transformations traversing highly reactive intermediates that engage in competing selectivity-determining pathways.

Using the MLR model as inspiration for further mechanistic inquiry, subtle global catalyst structural effects were elucidated that influence Rh–Rh covalency and catalyst propensity for carbene stabilization via  $\sigma$ -accepting interactions. The uncovering of this structural effect on dirhodium-carbene electrophilicity highlights the power of holistically analyzing data for catalysts that are both effective and ineffective for a given reaction class to pinpoint the properties that render an optimal catalyst. We anticipate that the understanding presented here of the many factors that conspire to dictate catalyst-controlled site-selectivity in C–H functionalization will prove valuable for guiding the design of next-generation catalysts using data-driven tools. While typical statistical models are derived from highly simplified model systems, the workflow presented preserves mechanistic fidelity by utilizing a spatial classification filter to efficiently extract descriptors only for catalyst conformers relevant for carbene formation and reactivity. In doing so, this study should inform future efforts to efficiently extend mechanistic insights from a single representative transition state study to predict reactivity across a combinatorial library of potential catalysts using data science techniques. Furthermore, both the computational workflow and SMART descriptors developed in this study are broadly transferrable for quantifying reactive site topology and predicting reactivity in other chemical and biological systems.

## ASSOCIATED CONTENT

## Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Complete experimental procedures, compound characterization, and computational modeling details are described in the Supporting Information. (PDF)

Catalyst property inputs for statistical modeling are available as a Microsoft Excel spreadsheet. (XLSX)

CIF files and reports are available for Rh<sub>2</sub>[*S*-2-Cl-4-(*p*-<sup>t</sup>BuC<sub>6</sub>H<sub>4</sub>)TPCP]<sub>4</sub> (CCDC 2095750), Rh<sub>2</sub>[*S*-2-Cl-4-(2,4,6-triMeC<sub>6</sub>H<sub>4</sub>)TPCP]<sub>4</sub> (CCDC 2089324), Rh<sub>2</sub>[*S*-2-Cl-5-(*p*-<sup>t</sup>BuC<sub>6</sub>H<sub>4</sub>)TPCP]<sub>4</sub> (CCDC 2090926), and Rh<sub>2</sub>[*S*-3,5-diMeC<sub>6</sub>H<sub>4</sub>)TPCP]<sub>4</sub> (CCDC 2089273). (CIF)

## AUTHOR INFORMATION

### Corresponding Author

\*hmdavie@emory.edu

\*matt.sigman@utah.edu

### Author Contributions

‡These authors contributed equally.

### Notes

HMLD is a named inventor on a patent entitled, Dirhodium Catalyst Compositions and Synthetic Processes Related Thereto (US 8,974,428, issued March 10, 2015). The other authors have no competing financial interests.

## ACKNOWLEDGMENT

Financial support was provided by NSF under the CCI Center for Selective C–H Functionalization (CHE-1700982). M.S.S. & R.C.C. would like to thank Ángel L. Mudarra for initial exploration, as well as Tobias Gensch, Beck Miller, Zack Ren, Djameladdin Musaev, and Richmond Sarpong for insightful discussions and suggestions. R.C.C. would also like to acknowledge XSEDE allocation CHE200027 for supplemental computing resources.

## REFERENCES

- (a) Gutekunst, W. R.; Baran, P. S., C–H functionalization logic in total synthesis. *Chem. Soc. Rev.* **2011**, *40* (4), 1976-1991; (b) Hartwig, J. F., Evolution of C–H Bond Functionalization from Methane to Methodology. *J. Am. Chem. Soc.* **2016**, *138* (1), 2-24; (c) Abrams, D. J.; Provencher, P. A.; Sorensen, E. J., Recent applications of C–H functionalization in complex natural product synthesis. *Chem. Soc. Rev.* **2018**, *47* (23), 8925-8967; (d) Cernak, T.; Dykstra, K. D.; Tyagarajan, S.; Vachal, P.; Kraska, S. W., The medicinal chemist's toolbox for late stage functionalization of drug-like molecules. *Chem. Soc. Rev.* **2016**, *45* (3), 546-576; (e) Newhouse, T.; Baran, P. S., If C–H Bonds Could Talk: Selective C–H Bond Oxidation. *Angew. Chem. Int. Ed.* **2011**, *50* (15), 3362-3374; (f) Dalton, T.; Faber, T.; Glorius, F., C–H Activation: Toward Sustainability and Applications. *ACS Cent. Sci.* **2021**, *7* (2), 245-261; (g) Yamaguchi, J.; Yamaguchi, A. D.; Itami, K., C–H Bond Functionalization: Emerging Synthetic Tools for Natural Products and Pharmaceuticals. *Angew. Chem. Int. Ed.* **2012**, *51* (36), 8960-9009.
- (a) Struble, T. J.; Coley, C. W.; Jensen, K. F., Multitask prediction of site selectivity in aromatic C–H functionalization reactions. *React. Chem. Eng.* **2020**, *5* (5), 896-902; (b) Li, X.; Zhang,

- S.-Q.; Xu, L.-C.; Hong, X., Predicting regioselectivity in radical C–H functionalization of heterocycles through machine learning. *Angew. Chem. Int. Ed.* **2020**, *59* (32), 13253-13259; (c) Lei, H.; Rovis, T., A site-selective amination catalyst discriminates between nearly identical C–H bonds of unsymmetrical disubstituted alkenes. *Nat. Chem.* **2020**, *12* (8), 725-731; (d) Hwang, Y.; Jung, H.; Lee, E.; Kim, D.; Chang, S., Quantitative analysis on two-point ligand modulation of iridium catalysts for chemodivergent C–H amidation. *J. Am. Chem. Soc.* **2020**, *142* (19), 8880-8889; (e) Besora, M.; Olmos, A.; Gava, R.; Noverges, B.; Asensio, G.; Caballero, A.; Maseras, F.; Pérez, P. J., A quantitative model for alkane nucleophilicity based on C–H bond structural/topological descriptors. *Angew. Chem. Int. Ed.* **2020**, *59* (8), 3112-3116; (f) Park, Y.; Niemeyer, Z. L.; Yu, J.-Q.; Sigman, M. S., Quantifying structural effects of amino acid ligands in Pd(II)-catalyzed enantioselective C–H functionalization reactions. *Organometallics* **2018**, *37* (2), 203-210; (g) Piou, T.; Romanov-Michailidis, F.; Romanova-Michaelides, M.; Jackson, K. E.; Semakul, N.; Taggart, T. D.; Newell, B. S.; Rithner, C. D.; Paton, R. S.; Rovis, T., Correlating reactivity and selectivity to cyclopentadienyl ligand properties in Rh(III)-catalyzed C–H activation reactions: An experimental and computational study. *J. Am. Chem. Soc.* **2017**, *139* (3), 1296-1310; (h) Gormisky, P. E.; White, M. C., Catalyst-controlled aliphatic C–H oxidations with a predictive model for site-selectivity. *J. Am. Chem. Soc.* **2013**, *135* (38), 14052-14055; (i) Bischoff, A. J.; Nelson, B. M.; Niemeyer, Z. L.; Sigman, M. S.; Movassaghi, M., Quantitative modeling of bis(pyridine)silver(I) permanganate oxidation of hydantoin derivatives: Guidelines for predicting the site of oxidation in complex substrates. *J. Am. Chem. Soc.* **2017**, *139* (43), 15539-15547; (j) Bess, E. N.; DeLuca, R. J.; Tindall, D. J.; Oderinde, M. S.; Roizen, J. L.; Du Bois, J.; Sigman, M. S., Analyzing site selectivity in Rh<sub>2</sub>(esp)<sub>2</sub>-catalyzed intermolecular C–H amination reactions. *J. Am. Chem. Soc.* **2014**, *136* (15), 5783-5789.
3. (a) Shao, Q.; Wu, K.; Zhuang, Z.; Qian, S.; Yu, J.-Q., From Pd(OAc)<sub>2</sub> to chiral catalysts: The discovery and development of bifunctional mono-N-protected amino acid ligands for diverse C–H functionalization reactions. *Acc. Chem. Res.* **2020**, *53* (4), 833-851; (b) Romero, E. A.; Chen, G.; Gembicky, M.; Jazsar, R.; Yu, J.-Q.; Bertrand, G., Understanding the Activity and Enantioselectivity of Acetyl-Protected Aminoethyl Quinoline Ligands in Palladium-Catalyzed β-C(sp<sup>3</sup>)-H Bond Arylation Reactions. *J. Am. Chem. Soc.* **2019**, *141* (42), 16726-16733; (c) Yu, J.-Q.; Shi, Z., C–H activation. In *Topics in Current Chemistry*. Springer: 2010; Vol. 292, p. 57-194; (d) Farr, C. M. B.; Kazerouni, A. M.; Park, B.; Poff, C. D.; Won, J.; Sharp, K. R.; Baik, M.-H.; Blakey, S. B., Designing a Planar Chiral Rhodium Indenyl Catalyst for Regio- and Enantioselective Allylic C–H Amidation. *J. Am. Chem. Soc.* **2020**, *142* (32), 13996-14004; (e) Burg, F.; Breitenlechner, S.; Jandl, C.; Bach, T., Enantioselective oxygenation of exocyclic methylene groups by a manganese porphyrin catalyst with a chiral recognition site. *Chem. Sci.* **2020**, *11* (8), 2121-2129; (f) Annappureddy, R. R.; Jandl, C.; Bach, T., A Chiral Phenanthroline Ligand with a Hydrogen-Bonding Site: Application to the Enantioselective Amination of Methylene Groups. *J. Am. Chem. Soc.* **2020**, *142* (16), 7374-7378; (g) Diesel, J.; Grosheva, D.; Kodama, S.; Cramer, N., A Bulky Chiral N-Heterocyclic Carbene Nickel Catalyst Enables Enantioselective C–H Functionalizations of Indoles and Pyrroles. *Angew. Chem. Int. Ed.* **2019**, *58* (32), 11044-11048; (h) Milan, M.; Salamone, M.; Costas, M.; Bietti, M., The quest for selectivity in hydrogen atom transfer based aliphatic C–H bond oxygenation. *Acc. Chem. Res.* **2018**, *51* (9), 1984-1995; (i) Ye, B.; Cramer, N., Chiral cyclopentadienyls: Enabling ligands for asymmetric Rh(III)-catalyzed C–H functionalizations. *Acc. Chem. Res.* **2015**, *48* (5), 1308-1318; (j) Lu, H.; Zhang, X. P., Catalytic C–H functionalization by metalloporphyrins: recent developments and future directions. *Chem. Soc. Rev.* **2011**, *40* (4), 1899-1909; (k) Davies, H. M. L.; Morton, D., Guiding principles for site selective and stereoselective intermolecular C–H functionalization by donor/acceptor rhodium carbenes. *Chem. Soc. Rev.* **2011**, *40* (4), 1857-1869; (l) Che, C.-M.; Lo, V. K.-Y.; Zhou, C.-Y.; Huang, J.-S., Selective functionalisation of saturated C–H bonds with metalloporphyrin catalysts. *Chem. Soc. Rev.* **2011**, *40* (4), 1950-1975; (m) Hayashi, H.; Uchida, T., Nitrene transfer reactions for asymmetric C–H amination: Recent development. *Eur. J. Org. Chem.* **2020**, *2020* (8), 909-916.
4. For additional examples of site-selective C–H functionalization proceeding via alternative mechanisms, see: (a) Jiao, K.-J.; Xing, Y.-K.; Yang, Q.-L.; Qiu, H.; Mei, T.-S., Site-Selective C–H Functionalization via Synergistic Use of Electrochemistry and Transition Metal Catalysis. *Acc. Chem. Res.* **2020**, *53* (2), 300-310; (b) Karmel, C.; Chen, Z.; Hartwig, J. F., Iridium-Catalyzed Silylation of C–H Bonds in Unactivated Arenes: A Sterically Encumbered Phenanthroline Ligand Accelerates Catalysis. *J. Am. Chem. Soc.* **2019**, *141* (17), 7063-7072; (c) Xia, Y.; Wang, L.; Studer, A., Site-Selective Remote Radical C–H Functionalization of Unactivated C–H Bonds in Amides Using Sulfone Reagents. *Angew. Chem. Int. Ed.* **2018**, *57* (39), 12940-12944; (d) Carestia, A. M.; Ravelli, D.; Alexanian, E. J., Reagent-dictated site selectivity in intermolecular aliphatic C–H functionalizations using nitrogen-centered radicals. *Chem. Sci.* **2018**, *9* (24), 5360-5365; (e) Hartwig, J. F., Catalyst-controlled site-selective bond activation. *Acc. Chem. Res.* **2017**, *50* (3), 549-555; (f) Saito, Y.; Segawa, Y.; Itami, K., *para*-C–H Borylation of Benzene Derivatives by a Bulky Iridium Catalyst. *J. Am. Chem. Soc.* **2015**, *137* (15), 5193-5198; (g) Neufeldt, S. R.; Sanford, M. S., Controlling site selectivity in palladium-catalyzed C–H bond functionalization. *Acc. Chem. Res.* **2012**, *45* (6), 936-946; (h) Wang, D.; Salazar, C. A.; Stahl, S. S., Catalyst-Controlled Regioselectivity in Pd-Catalyzed Aerobic Oxidative Arylation of Indoles. *Organometallics* **2021**, DOI: 10.1021/acs.organomet.1c00139; (i) Salazar, C. A.; Flesch, K. N.; Haines, B. E.; Zhou, P. S.; Musaev, D. G.; Stahl, S. S., Tailored quinones support high-turnover Pd catalysts for oxidative C–H arylation with O<sub>2</sub>. *Science* **2020**, *370* (6523), 1454-1460.
5. (a) Muñoz-Molina, J. M.; Belderrain, T. R.; Pérez, P. J., Trispyrazolylborate coinage metals complexes: Structural features and catalytic transformations. *Coord. Chem. Rev.* **2019**, *390*, 171-189; (b) Weldy, N. M.; Schafer, A. G.; Owens, C. P.; Herting, C. J.; Varela-Alvarez, A.; Chen, S.; Niemeyer, Z.; Musaev, D. G.; Sigman, M. S.; Davies, H. M. L.; Blakey, S. B., Iridium(III)-bis(imidazolyl)phenyl catalysts for enantioselective C–H functionalization with ethyl diazoacetate. *Chem. Sci.* **2016**, *7* (5), 3142-3146; (c) Caballero, A.; Despagnet-Ayoub, E.; Díaz-Requejo, M. M.; Díaz-Rodríguez, A.; González-Núñez, M. E.; Mello, R.; Muñoz, B. K.; Ojo, W.-S.; Asensio, G.; Etienne, M.; Pérez, P. J., Silver-catalyzed C–C bond formation between methane and ethyl diazoacetate in supercritical CO<sub>2</sub>. *Science* **2011**, *332* (6031), 835-838; (d) Suematsu, H.; Katsuki, T., Iridium(III) catalyzed diastereo- and enantioselective C–H bond functionalization. *J. Am. Chem. Soc.* **2009**, *131* (40), 14218-14219; (e) Thu, H.-Y.; Tong, G. S.-M.; Huang, J.-S.; Chan, S. L.-F.; Deng, Q.-H.; Che, C.-M., Highly Selective

- Metal Catalysts for Intermolecular Carbenoid Insertion into Primary C–H Bonds and Enantioselective C–C Bond Formation. *Angew. Chem. Int. Ed.* **2008**, *47* (50), 9747-9751; (f) Yamawaki, M.; Tsutsui, H.; Kitagaki, S.; Anada, M.; Hashimoto, S., Dirhodium(II) tetrakis[N-tetrachlorophthaloyl-(S)-tert-leucinate]: a new chiral Rh(II) catalyst for enantioselective amidation of C–H bonds. *Tetrahedron Lett.* **2002**, *43* (52), 9561-9564; (g) Callot, H. J.; Metz, F., Homologation of n-alkanes using diazoesters and rhodium(III)porphyrins. Enhanced attack on primary C–H bonds. *Tetrahedron Lett.* **1982**, *23* (42), 4321-4324; (h) Davies, H. M. L.; Liao, K., Dirhodium tetracarboxylates as catalysts for selective intermolecular C–H functionalization. *Nat. Rev. Chem.* **2019**, *3* (6), 347-360.
6. (a) Ju, M.; Schomaker, J. M., Nitrene transfer catalysts for enantioselective C–N bond formation. *Nat. Rev. Chem.* **2021**, 1-15 (DOI: 10.1038/s41570-021-00291-4); (b) Fanourakis, A.; Williams, B. D.; Paterson, K. J.; Phipps, R. J., Enantioselective Intermolecular C–H Amination Directed by a Chiral Cation. *J. Am. Chem. Soc.* **2021**, *143* (27), 10070-10076; (c) Baek, Y.; Das, A.; Zheng, S.-L.; Reibenspies, J. H.; Powers, D. C.; Betley, T. A., C–H amination mediated by cobalt organoazide adducts and the corresponding cobalt nitrenoid intermediates. *J. Am. Chem. Soc.* **2020**, *142* (25), 11232-11243; (d) Nasrallah, A.; Lazib, Y.; Boquet, V.; Darses, B.; Dauban, P., Catalytic Intermolecular C(sp<sup>3</sup>)–H Amination with Sulfamates for the Asymmetric Synthesis of Amines. *Org. Process Res. Dev.* **2019**, *24* (5), 724-728; (e) Nasrallah, A.; Boquet, V.; Hecker, A.; Retailleau, P.; Darses, B.; Dauban, P., Catalytic Enantioselective Intermolecular Benzylic C(sp<sup>3</sup>)–H Amination. *Angew. Chem. Int. Ed.* **2019**, *58* (24), 8192-8196; (f) Chiappini, N. D.; Mack, J. B.; Du Bois, J., Intermolecular C(sp<sup>3</sup>)–H amination of complex molecules. *Angew. Chem. Int. Ed.* **2018**, *57* (18), 4956-4959; (g) Nishioka, Y.; Uchida, T.; Katsuki, T., Enantio- and Regioselective Intermolecular Benzylic and Allylic C–H Bond Amination. *Angew. Chem. Int. Ed.* **2013**, *52* (6), 1739-1742; (h) Bakhoda, A.; Jiang, Q.; Badiei, Y. M.; Bertke, J. A.; Cundari, T. R.; Warren, T. H., Copper-Catalyzed C(sp<sup>3</sup>)–H Amidation: Sterically Driven Primary and Secondary C–H Site-Selectivity. *Angew. Chem. Int. Ed.* **2019**, *58* (11), 3421-3425.
7. (a) Vicens, L.; Olivo, G.; Costas, M., Rational Design of Bioinspired Catalysts for Selective Oxidations. *ACS Catal.* **2020**, *10* (15), 8611-8631; (b) Cianfanelli, M.; Olivo, G.; Milan, M.; Klein Gebbink, R. J. M. K.; Ribas, X.; Bietti, M.; Costas, M., Enantioselective C–H lactonization of unactivated methylenes directed by carboxylic acids. *J. Am. Chem. Soc.* **2020**, *142* (3), 1584-1593; (c) White, M. C.; Zhao, J., Aliphatic C–H oxidations for late-stage functionalization. *J. Am. Chem. Soc.* **2018**, *140* (43), 13988-14009; (d) Ottenbacher, R. V.; Talsi, E. P.; Rybalova, T. V.; Bryliakov, K. P., Enantioselective Benzylic Hydroxylation of Arylalkanes with H<sub>2</sub>O<sub>2</sub> in Fluorinated Alcohols in the Presence of Chiral Mn Aminopyridine Complexes. *ChemCatChem* **2018**, *10* (22), 5323-5330; (e) Burg, F.; Gicquel, M.; Breitenlechner, S.; Pöthig, A.; Bach, T., Site- and Enantioselective C–H Oxygenation Catalyzed by a Chiral Manganese Porphyrin Complex with a Remote Binding Site. *Angew. Chem. Int. Ed.* **2018**, *57* (11), 2953-2957; (f) Talsi, E. P.; Samsonenko, D. G.; Ottenbacher, R. V.; Bryliakov, K. P., Highly Enantioselective C–H Oxidation of Arylalkanes with H<sub>2</sub>O<sub>2</sub> in the Presence of Chiral Mn-Aminopyridine Complexes. *ChemCatChem* **2017**, *9* (24), 4580-4586; (g) Milan, M.; Bietti, M.; Costas, M., Highly enantioselective oxidation of nonactivated aliphatic C–H bonds with hydrogen peroxide catalyzed by manganese complexes. *ACS Cent. Sci.* **2017**, *3* (3), 196-204; (h) Bigi, M. A.; Reed, S. A.; White, M. C., Directed metal (oxo) aliphatic C–H hydroxylations: Overriding substrate bias. *J. Am. Chem. Soc.* **2012**, *134* (23), 9721-9726; (i) Chen, M. S.; White, M. C., Combined effects on selectivity in Fe-catalyzed methylene oxidation. *Science* **2010**, *327* (5965), 566-571; (j) Groves, J. T.; Viski, P., Asymmetric hydroxylation by a chiral iron porphyrin. *J. Am. Chem. Soc.* **1989**, *111* (22), 8537-8538.
8. (a) Yang, Y.; Arnold, F. H., Navigating the unnatural reaction space: Directed evolution of heme proteins for selective carbene and nitrene transfer. *Acc. Chem. Res.* **2021**, *54* (5), 1209-1225; (b) Trouvé, J.; Zardi, P.; Al-Shehimi, S.; Roisnel, T.; Gramage-Doria, R., Enzyme-like Supramolecular Iridium Catalysis Enabling C–H Bond Borylation of Pyridines with *meta*-Selectivity. *Angew. Chem. Int. Ed.* **2021**, DOI: 10.1002/anie.202107624; (c) Reyes, R. L.; Sato, M.; Iwai, T.; Suzuki, K.; Maeda, S.; Sawamura, M., Asymmetric remote C–H borylation of aliphatic amides and esters with a modular iridium catalyst. *Science* **2020**, *369* (6506), 970-974; (d) Jia, Z.-J.; Gao, S.; Arnold, F. H., Enzymatic primary amination of benzylic and allylic C(sp<sup>3</sup>)–H bonds. *J. Am. Chem. Soc.* **2020**, *142* (23), 10279-10283; (e) Perez-Rizquez, C.; Rodriguez-Otero, A.; Palomo, J. M., Combining enzymes and organometallic complexes: novel artificial metalloenzymes and hybrid systems for C–H activation chemistry. *Org. Biomol. Chem.* **2019**, *17* (30), 7114-7123; (f) Olivo, G.; Capocasa, G.; Lanzalunga, O.; Di Stefano, S.; Costas, M., Enzyme-like substrate-selectivity in C–H oxidation enabled by recognition. *Chem. Commun.* **2019**, *55* (7), 917-920; (g) Dydio, P.; Key, H. M.; Nazarenko, A.; Rha, J. Y.-E.; Seyedkazemi, V.; Clark, D. S.; Hartwig, J. F., An artificial metalloenzyme with the kinetics of native enzymes. *Science* **2016**, *354* (6308), 102-106; (h) Hyster, T. K.; Knörr, L.; Ward, T. R.; Rovis, T., Biotinylated Rh(III) complexes in engineered streptavidin for accelerated asymmetric C–H activation. *Science* **2012**, *338* (6106), 500-503; (i) Lewis, J. C.; Coelho, P. S.; Arnold, F. H., Enzymatic functionalization of carbon–hydrogen bonds. *Chem. Soc. Rev.* **2011**, *40* (4), 2003-2021; (j) Espinoza, R. V.; Haatveit, K. C.; Grossman, S. W.; Tan, J. Y.; McGlade, C. A.; Khatri, Y.; Newmister, S. A.; Schmidt, J. J.; Garcia-Borràs, M.; Montgomery, J.; Houk, K. N.; Sherman, D. H., Engineering P450 TamI as an Iterative Biocatalyst for Selective Late-Stage C–H Functionalization and Epoxidation of Tirandamycin Antibiotics. *ACS Catal.* **2021**, *11* (13), 8304-8316.
9. Qin, C.; Davies, H. M. L., Role of Sterically Demanding Chiral Dirhodium Catalysts in Site-Selective C–H Functionalization of Activated Primary C–H Bonds. *J. Am. Chem. Soc.* **2014**, *136* (27), 9792-9796.
10. (a) Liao, K.; Yang, Y.-F.; Li, Y.; Sanders, J. N.; Houk, K. N.; Musaev, D. G.; Davies, H. M. L., Design of catalysts for site-selective and enantioselective functionalization of non-activated primary C–H bonds. *Nat. Chem.* **2018**, *10* (10), 1048-1055; (b) Liao, K.; Pickel, T. C.; Boyarskikh, V.; Bacsa, J.; Musaev, D. G.; Davies, H. M. L., Site-selective and stereoselective functionalization of non-activated tertiary C–H bonds. *Nature* **2017**, *551* (7682), 609-613; (c) Liao, K.; Negretti, S.; Musaev, D. G.; Bacsa, J.; Davies, H. M. L., Site-selective and stereoselective functionalization of unactivated C–H bonds. *Nature* **2016**, *533* (7602), 230-234.
11. (a) Liu, W.; Babl, T.; Röther, A.; Reiser, O.; Davies, H. M. L., Functionalization of Piperidine Derivatives for the Site-Selective and Stereoselective Synthesis of Positional Analogues of Methylphenidate. *Chem. Eur. J.* **2020**, *26* (19), 4236-4241; (b) Liu, W.; Ren, Z.; Bosse, A. T.; Liao, K.; Goldstein, E. L.; Bacsa, J.; Musaev,

- D. G.; Stoltz, B. M.; Davies, H. M. L., Catalyst-Controlled Selective Functionalization of Unactivated C–H Bonds in the Presence of Electronically Activated C–H Bonds. *J. Am. Chem. Soc.* **2018**, *140* (38), 12247-12255.
12. (a) Ren, Z.; Musaev, D. G.; Davies, H. M. L., Influence of Aryl Substituents on the Alignment of Ligands in the Dirhodium Tetrakis(1,2,2-Triarylcyclopropane-carboxylate) Catalysts. *ChemCatChem* **2021**, *13* (1), 174-179; (b) Liao, K.; Liu, W.; Niemeyer, Z. L.; Ren, Z.; Bacsa, J.; Musaev, D. G.; Sigman, M. S.; Davies, H. M. L., Site-Selective Carbene-Induced C–H Functionalization Catalyzed by Dirhodium Tetrakis(triarylcyclopropanecarboxylate) Complexes. *ACS Catal.* **2018**, *8* (1), 678-682.
13. For additional examples of mechanistic insights from quantum mechanical calculations for dirhodium-catalyzed C–H functionalization with donor/acceptor carbenes, see: (a) Wertz, B.; Ren, Z.; Bacsa, J.; Musaev, D. G.; Davies, H. M. L., Comparison of 1,2-Diarylcyclopropanecarboxylates with 1,2,2-Triarylcyclopropanecarboxylates as Chiral Ligands for Dirhodium-Catalyzed Cyclopropanation and C–H Functionalization. *J. Org. Chem.* **2020**, *85* (19), 12199-12211; (b) Lee, M.; Ren, Z.; Musaev, D. G.; Davies, H. M. L., Rhodium-Stabilized Diarylcarbenes Behaving as Donor/Acceptor Carbenes. *ACS Catal.* **2020**, *10* (11), 6240-6247; (c) Fu, J.; Ren, Z.; Bacsa, J.; Musaev, D. G.; Davies, H. M. L., Desymmetrization of cyclohexanes by site- and stereoselective C–H functionalization. *Nature* **2018**, *564* (7736), 395-399; (d) Qin, C.; Boyarskikh, V.; Hansen, J. H.; Hardcastle, K. I.; Musaev, D. G.; Davies, H. M. L.,  $D_2$ -Symmetric Dirhodium Catalyst Derived from a 1,2,2-Triarylcyclopropanecarboxylate Ligand: Design, Synthesis and Application. *J. Am. Chem. Soc.* **2011**, *133* (47), 19198-19204.
14. For a discussion of the limitations in interrogating full reaction pathways for large, flexible metal catalysts using quantum mechanical transition state calculations, as well as selected examples of recent methods to automate this process and/or utilize a lower level of theory, see: (a) Vitek, A. K.; Jugovic, T. M. E.; Zimmerman, P. M., Revealing the strong relationships between ligand conformers and activation barriers: A case study of bisphosphine reductive elimination. *ACS Catal.* **2020**, *10* (13), 7136-7145; (b) Minenkov, Y.; Sharapa, D. I.; Cavallo, L., Application of semiempirical methods to transition metal complexes: Fast results but hard-to-predict accuracy. *J. Chem. Theory Comput.* **2018**, *14* (7), 3428-3439; (c) Guan, Y.; Ingman, V. M.; Rooks, B. J.; Wheeler, S. E., AARON: An automated reaction optimizer for new catalysts. *J. Chem. Theory Comput.* **2018**, *14* (10), 5249-5261; (d) Poree, C.; Schoenebeck, F., A holy grail in chemistry: Computational catalyst design: Feasible or fiction? *Acc. Chem. Res.* **2017**, *50* (3), 605-608; (e) Hammes-Schiffer, S., Catalysts by design: The power of theory. *Acc. Chem. Res.* **2017**, *50* (3), 561-566; (f) Patrascu, M. B.; Pottel, J.; Pinus, S.; Bezanson, M.; Norrby, P.-O.; Moitessier, N., From desktop to benchtop with automated computational workflows for computer-aided design in asymmetric catalysis. *Nat. Catal.* **2020**, *3* (7), 574-584; (g) Rosales, A. R.; Wahlers, J.; Limé, E.; Meadows, R. E.; Leslie, K. W.; Savin, R.; Bell, F.; Hansen, E.; Helquist, P.; Munday, R. H.; Wiest, O.; Norrby, P.-O., Rapid virtual screening of enantioselective catalysts using CatVS. *Nat. Catal.* **2019**, *2* (1), 41-45; (h) Hansen, E.; Rosales, A. R.; Tutkowski, B.; Norrby, P.-O.; Wiest, O., Prediction of Stereochemistry using Q2MM. *Acc. Chem. Res.* **2016**, *49* (5), 996-1005; (i) Grimme, S. Exploration of Chemical Compound Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *J. Chem. Theory Comput.* **2019**, *15* (5), 2847-2862.
15. (a) Santiago, C. B.; Guo, J.-Y.; Sigman, M. S., Predictive and mechanistic multivariate linear regression models for reaction development. *Chem. Sci.* **2018**, *9* (9), 2398-2412; (b) 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* (6), 1292-1301.
16. (a) Żurański, A. M.; Martínez Alvarado, J. I.; Shields, B. J.; Doyle, A. G., Predicting Reaction Yields via Supervised Learning. *Acc. Chem. Res.* **2021**, *54* (8), 1856-1865; (b) Rinehart, N. I.; Zahrt, A. F.; Henle, J. J.; Denmark, S. E., Dreams, False Starts, Dead Ends, and Redemption: A Chronicle of the Evolution of a Chemoinformatic Workflow for the Optimization of Enantioselective Catalysts. *Acc. Chem. Res.* **2021**, *54* (9), 2041-2054; (c) Janet, J. P.; Duan, C.; Nandy, A.; Liu, F.; Kulik, H. J., Navigating transition-metal chemical space: Artificial intelligence for first-principles design. *Acc. Chem. Res.* **2021**, *54* (3), 532-545; (d) dos Passos Gomes, G.; Pollice, R.; Aspuru-Guzik, A., Navigating through the maze of homogeneous catalyst design with machine learning. *Trends Chem.* **2021**, *3* (2), 96-110; (e) 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* (6424), DOI: 10.1126/science.aau5631; (f) 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* (6385), 186-190.
17. Durand, D. J.; Fey, N., Building a toolbox for the analysis and prediction of ligand and catalyst effects in organometallic catalysis. *Acc. Chem. Res.* **2021**, *54* (4), 837-848.
18. Hatridge, T. A.; Liu, W.; Yoo, C.-J.; Davies, H. M. L.; Jones, C. W., Optimized Immobilization Strategy for Dirhodium(II) Carboxylate Catalysts for C–H Functionalization and Their Implementation in a Packed Bed Flow Reactor. *Angew. Chem. Int. Ed.* **2020**, *59* (44), 19525-19531.
19. (a) Lindsay, V. N. G.; Lin, W.; Charette, A. B., Experimental Evidence for the All-Up Reactive Conformation of Chiral Rhodium(II) Carboxylate Catalysts: Enantioselective Synthesis of *cis*-Cyclopropane  $\alpha$ -Amino Acids. *J. Am. Chem. Soc.* **2009**, *131* (45), 16383-16385; (b) DeAngelis, A.; Dmitrenko, O.; Yap, G. P. A.; Fox, J. M., Chiral Crown Conformation of  $Rh_2(S-PTTL)_4$ : Enantioselective Cyclopropanation with  $\alpha$ -Alkyl- $\alpha$ -diazoesters. *J. Am. Chem. Soc.* **2009**, *131* (21), 7230-7231.
20. (a) *MacroModel, version 11.8*, Schrödinger, LLC: New York, NY 2017; (b) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgenson, W. L.; Abel, R.; Friesner, R. A., OPLS3: A force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theory Comput.* **2016**, *12* (1), 281-296; (c) Kelley, L. A.; Gardner, S. P.; Sutcliffe, M. J., An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. *Protein Eng. Des. Sel.* **1996**, *9* (11), 1063-1065.
21. (a) Senn, H. M.; Thiel, W., QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed.* **2009**, *48* (7), 1198-1229; (b) Lin, H.; Truhlar, D. G., QM/MM: what have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.* **2007**, *117* (2), 185-199.

22. (a) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, *Gaussian 16, Revision C.01*, Gaussian, Inc.: Wallingford, CT, 2016; (b) Grimme, S.; Ehrlich, S.; Goerigk, L., Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32* (7), 1456-1465; (c) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132* (15), 154104; (d) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G., Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **2009**, *113* (18), 6378-6396; (e) Zhao, Y.; Truhlar, D. G., The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120* (1), 215-241; (f) Johnson, E. R.; Becke, A. D., A post-Hartree-Fock model of intermolecular interactions: Inclusion of higher-order corrections. *J. Chem. Phys.* **2006**, *124* (17), 174104; (g) Weigend, F.; Ahlrichs, R., Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7* (18), 3297-3305; (h) Johnson, E. R.; Becke, A. D., A post-Hartree-Fock model of intermolecular interactions. *J. Chem. Phys.* **2005**, *123* (2), 024101; (i) Becke, A. D.; Johnson, E. R., A density-functional model of the dispersion interaction. *J. Chem. Phys.* **2005**, *123* (15), 154101; (j) Rassolov, V. A.; Pople, J. A.; Ratner, M. A.; Windus, T. L., 6-31G\* basis set for atoms K through Zn. *J. Chem. Phys.* **1998**, *109* (4), 1223-1229; (k) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J., *Ab initio* calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.* **1994**, *98* (45), 11623-11627; (l) Beck, A. D., Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98* (7), 5648; (m) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard III, W. A.; Skiff, W. M., UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114* (25), 10024-10035; (n) Andrae, D.; Häußermann, U.; Dolg, M.; Stoll, H.; Preuß, H., Energy-adjusted *ab initio* pseudopotentials for the second and third row transition elements. *Theor. Chim. Acta* **1990**, *77* (2), 123-141; (o) Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37* (2), 785; (p) Hay, P. J.; Wadt, W. R., *Ab initio* effective core potentials for molecular calculations. Potentials for K to Au including the outermost core orbitals. *J. Chem. Phys.* **1985**, *82* (1), 299-310; (q) Hay, P. J.; Wadt, W. R., *Ab initio* Effective Core Potentials for Molecular Calculations. Potentials for the Transition Metal Atoms Sc to Hg. *J. Chem. Phys.* **1985**, *82*, 270-283; (r) Vosko, S. H.; Wilk, L.; Nusair, M., Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* **1980**, *58* (8), 1200-1211.
23. Full QM optimizations were performed for 8 of the 25 catalysts, including Rh<sub>2</sub>(2-Cl-5-ArTPCP)<sub>4</sub> and Rh<sub>2</sub>(3,5-diXTPCP)<sub>4</sub> catalysts, in order to more reliably assess the relative energetics of competing configurations by appropriately accounting for dispersive interactions between closely approaching neighboring ligands in these catalyst subclasses.
24. (a) Adly, F. G.; Gardiner, M. G.; Ghanem, A., Design and synthesis of novel chiral dirhodium(II) carboxylate complexes for asymmetric cyclopropanation reactions. *Chem. Eur. J.* **2016**, *22* (10), 3447-3461; (b) Lindsay, V. N. G.; Charette, A. B., Design and synthesis of chiral heteroleptic rhodium(II) carboxylate catalysts: Experimental investigation of halogen bond rigidification effects in asymmetric cyclopropanation. *ACS Catal.* **2012**, *2* (6), 1221-1225; (c) DeAngelis, A.; Boruta, D. T.; Lubin, J.-B.; Plampin III, J. N.; Yap, G. P. A.; Fox, J. M., The chiral crown conformation in paddlewheel complexes. *Chem. Commun.* **2010**, *46* (25), 4541-4543.
25. Hansen, J.; Davies, H. M. L., High symmetry dirhodium(II) paddlewheel complexes as chiral catalysts. *Coord. Chem. Rev.* **2008**, *252* (5-7), 545-555.
26. (a) Kluyver, T.; Ragan-Kelley, B.; Pérez, F.; Granger, B. E.; Bussonnier, M.; Frederic, J.; Kelley, K.; Hamrick, J. B.; Grout, J.; Corlay, S.; Ivanov, P.; Avila, D.; Abdalla, S.; Willing, C., *Jupyter Notebooks-a publishing format for reproducible computational workflows*. 2016; Vol. 2016; (b) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É., Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825-2830; (c) *Anaconda Software Distribution*, Anaconda Inc.: **2020**; (d) E. Peters, SigmanGroup/threshold: Single parameter threshold (1.0.0). Zenodo (2021): <http://doi.org/10.5281/zenodo.5227162>; (e) Newman-Stonebraker, S. H.; Smith, S. R.; Borowski, J. E.; Peters, E.; Gensch, T.; B.; Johnson, H. C.; Sigman, M. S.; Doyle, A. G., Univariate classification of phosphine ligation state and reactivity in cross-coupling catalysis. *Science*, **2021**, *374* (6565), 301-308.
27. (a) Ren, Z.; Musaev, D. G.; Davies, H. M. L., Key Selectivity Controlling Elements in Rhodium-Catalyzed C-H Functionalization with Donor/Acceptor Carbenes. *ChemRxiv, Pre-print submitted*. **2021**; (b) Yang, L.-L.; Evans, D.; Xu, B.; Li, W.-T.; Li, M.-L.; Zhu, S.-F.; Houk, K. N.; Zhou, Q.-L., Enantioselective diarylcarbene insertion into Si-H bonds induced by electronic properties of the carbenes. *J. Am. Chem. Soc.* **2020**, *142* (28), 12394-12399; (c) For prior computational studies of dirhodium-catalyzed C-H functionalization, the reader is referred to refs. 13.
28. (a) Pritchard, B. P.; Altarawy, D.; Didier, B.; Gibson, T. D.; Windus, T. L., New basis set exchange: An open, up-to-date resource for the molecular sciences community. *J. Chem. Inf. Model.* **2019**, *59* (11), 4814-4820; (b) Spitznagel, G. W.; Clark, T.; von Ragué Schleyer, P.; Hehre, W. J., An evaluation of the performance of diffuse function-augmented basis sets for second row elements, Na-Cl. *J. Comput. Chem.* **1987**, *8* (8), 1109-1116; (c) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; von Ragué Schleyer, P., Efficient diffuse function-augmented basis sets for anion

calculations. III. The 3-21+ G basis set for first-row elements, Li–F. *J. Comput. Chem.* **1983**, *4* (3), 294-301; (d) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A., Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **1982**, *77* (7), 3654-3665; (e) McLean, A. D.; Chandler, G. S., Contracted Gaussian basis sets for molecular calculations. I. Second row atoms,  $Z=11-18$ . *J. Chem. Phys.* **1980**, *72* (10), 5639-5648; (f) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A., Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.* **1980**, *72* (1), 650-654.

29. (a) Guzei, I. A.; Wendt, M., An improved method for the computation of ligand steric effects based on solid angles. *Dalton Trans.* **2006**, *128* (33), 3991-3999. (b) Kinetic binding rates are also likely exponentially dependent on G% rather than linearly, which further accentuates the reactivity implications of incremental differences in G% (see SI).

30. (a) It is reasonable to speculate as to whether non-Curtin-Hammett behavior would be expected to hold true across a diverse catalyst set. Nevertheless, moving to more hindered  $Rh_2(S\text{-}o\text{-CITPCP})_4$  catalysts would be anticipated to increase the kinetic barriers for diazo dissociation to an equal or greater extent than those for  $N_2$  extrusion. While more stringent spatial constraints will increase  $2^\circ$  C–H insertion barriers such that dirhodium-carbene diastereomer interconversion that has critical implications on stereoselectivity may be permitted (see ref. 27a), the concomitant increase in catalyst ligand flipping barriers due to reactive site congestion should limit large-scale changes in dirhodium-carbene reactive site topology prior to C–H insertion. (b) For a related discussion of relative reaction rates as a function of dirhodium catalyst in asymmetric cyclopropanation reactions, see: Wei, B.; Sharland, J. C.; Lin, P.; Wilkerson-Hill, S. M.; Fullilove, F. A.; McKinnon, S.; Blackmond, D. G.; Davies, H. M. L., In Situ Kinetic Studies of Rh(II)-Catalyzed Asymmetric Cyclopropanation with Low Catalyst Loadings. *ACS Catal.* **2019**, *10* (2), 1161-1170.

31. (a) Bilbrey, J. A.; Kazez, A. H.; Locklin, J.; Allen, W. D., Exact Ligand Solid Angles. *J. Chem. Theory Comput.* **2013**, *9* (12), 5734-5744; (b) Taverner, B. C., Improved algorithm for accurate computation of molecular solid angles. *J. Comput. Chem.* **1996**, *17* (14), 1612-1623; (c) White, D.; Taverner, B. C.; Leach, P. G. L.; Coville, N. J., Solid angles I. The radial profile. *J. Organomet. Chem.* **1994**, *478* (1), 205-211; (d) Komatsuzaki, T.; Akai, I.; Sakakibara, K.; Hirota, M., A new steric substituent constant  $\Omega_s$ . Characteristic feature and comparison with some other steric constants. *Tetrahedron* **1992**, *48* (9), 1539-1556; (e) Hirota, M.; Sakakibara, K.; Komatsuzaki, T.; Akai, I., A new steric substituent constant  $\Omega_s$  based on molecular mechanics calculations. *Comput. Chem.* **1991**, *15* (3), 241-248; (f) Taverner, B. C.; Smith, J. M.; White, D. P.; Coville, N. J., Quantification of ligand-ligand interactions using solid angles. *S. Afr. J. Chem.* **1997**, *50* (2), 59-66.

32. (a) Verloop, A., In *Drug Design*, Ariens, E. J., Ed. Academic Press: 1976; Vol. III, p 133; (b) Hansch, C. and Leo, A., *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*. American Chemical Society: Washington, D.C., 1995, p. 1-580; (c) Verloop, A. and Tipker, J., In *Biological Activity and Chemical Structure*, Buisman, J. A., Ed. Elsevier: 1977; p 63; (d) Verloop, A. and Tipker, J., In *QSAR in Drug Dosing and Toxicology*, Hadzi, B. and Jerman-Blazic, B., Ed. Elsevier: 1987; p 97; (e) Brethomé, A. V.; Fletcher, S. P.; Paton, R. S., Conformational Effects on Physical-Organic Descriptors: The Case of Sterimol Steric Parameters. *ACS Catal.* **2019**, *9* (3), 2313-2323; (f) Harper, K. C.; Bess, E. N.; Sigman, M. S., Multidimensional steric parameters in the analysis of asymmetric catalytic reactions. *Nat. Chem.* **2012**, *4* (5), 366-374; (g) For recent improvements to Sterimol parameters, see: Luchini, G.; Paton, R. S. DBSTEP: DFT Based Steric Parameters. 2021, DOI: 10.5281/zenodo.4702097; (h) The authors note that "buried Sterimol" descriptors have also been explored to a lesser extent.

33. (a) Clavier, H.; Nolan, S. P., Percent buried volume for phosphine and N-heterocyclic carbene ligands: steric properties in organometallic chemistry. *Chem. Commun.* **2010**, *46* (6), 841-861; (b) Falivene, L.; Credendino, R.; Poater, A.; Petta, A.; Serra, L.; Oliva, R.; Scarano, V.; Cavallo, L., SambVca 2. A Web Tool for Analyzing Catalytic Pockets with Topographic Steric Maps. *Organometallics* **2016**, *35* (13), 2286-2293; (c) Falivene, L.; Cao, Z.; Petta, A.; Serra, L.; Poater, A.; Oliva, R.; Scarano, V.; Cavallo, L., Towards the online computer-aided design of catalytic pockets. *Nat. Chem.* **2019**, *11*, 872-879.

34. Durand, D. J.; Fey, N., Computational ligand descriptors for catalyst design. *Chem. Rev.* **2019**, *119* (11), 6561-6594.

35. (a) Wu, K.; Doyle, A. G., Parameterization of phosphine ligands demonstrates enhancement of nickel catalysis via remote steric effects. *Nat. Chem.* **2017**, *9* (8), 779-784; (b) Tolman, C. A., Steric effects of phosphorus ligands in organometallic chemistry and homogeneous catalysis. *Chem. Rev.* **1977**, *77* (3), 313-348.

36. (a) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E., UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **2021**, *30* (1), 70-82; (b) Goddard, T. D.; Huang, C. C.; Meng, E. C.; Pettersen, E. F.; Couch, G. S.; Morris, J. H.; Ferrin, T. E., UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **2018**, *27* (1), 14-25.

37. Angermund, K.; Baumann, W.; Dinjus, E.; Fornika, R.; Görls, H.; Kessler, M.; Krüger, C.; Leitner, W.; Lutz, F., Complexes  $[(P_2)Rh(hfacac)]$  as model compounds for the fragment  $[(P_2)Rh]$  and as highly active catalysts for  $CO_2$  hydrogenation: The accessible molecular surface (AMS) model as an approach to quantifying the intrinsic steric properties of chelating ligands in homogeneous catalysis. *Chem. Eur. J.* **1997**, *3* (5), 755-764.

38. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605-1612.

39. (a) Dundas, J.; Adamian, L.; Liang, J., Structural signatures of enzyme binding pockets from order-independent surface alignment: a study of metalloendopeptidase and NAD binding proteins. *J. Mol. Biol.* **2011**, *406* (5), 713-729; (b) Huang, B., MetaPocket: a meta approach to improve protein ligand binding site prediction. *OMICS* **2009**, *13* (4), 325-330; (c) Krone, M.; Kozlíková, B.; Lindow, N.; Baaden, M.; Baum, D.; Parulek, J.; Hege, H.-C.; Viola, I. In *Visual analysis of biomolecular cavities: State of the art*, Computer Graphics Forum, Wiley Online Library: 2016; pp 527-551; (d) Brezovsky, J.; Chovancova, E.; Gora, A.; Pavelka, A.; Biedermannova, L.; Damborsky, J., Software tools for identification, visualization and analysis of protein tunnels and channels. *Biotechnol. Adv.* **2013**, *31* (1), 38-49; (e) Zhao, J.; Cao, Y.; Zhang, L., Exploring the computational methods for protein-ligand binding site prediction. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 417-426; (f) Henrich, S.; Salo-Ahen, O. M. H.; Huang, B.; Rippmann, F. F.; Cruciani, G.; Wade, R. C., Computational approaches to

- identifying and characterizing protein binding sites for ligand design. *J. Mol. Recognit.* **2010**, *23* (2), 209-219; (g) Simões, T.; Lopes, D.; Dias, S.; Fernandes, F.; Pereira, J.; Jorge, J.; Bajaj, C.; Gomes, A. In *Geometric detection algorithms for cavities on protein surfaces in molecular graphics: A survey*, Computer graphics forum, Wiley Online Library: 2017; pp 643-683.
40. (a) Connolly, M. L., Analytical molecular surface calculation. *J. Appl. Crystallogr.* **1983**, *16* (5), 548-558; (b) Connolly, M. L., Solvent-accessible surfaces of proteins and nucleic acids. *Science* **1983**, *221* (4612), 709-713; (c) Schneider, S.; Zacharias, M., Combining geometric pocket detection and desolvation properties to detect putative ligand binding sites on proteins. *J. Struct. Biol.* **2012**, *180* (3), 546-550; (d) Kawabata, T.; Go, N., Detection of pockets on protein surfaces using small and large probe spheres to find putative ligand binding sites. *Proteins* **2007**, *68* (2), 516-529; (e) Brady, Jr., G. P.; Stouten, P. F., Fast prediction and visualization of protein binding pockets with PASS. *J. Comput. Aided Mol. Des.* **2000**, *14* (4), 383-401.
41. Benkaidali, L.; André, F.; Maouche, B.; Siregar, P.; Benyettou, M.; Maurel, F.; Petitjean, M., Computing cavities, channels, pores and pockets in proteins from non-spherical ligands models. *Bioinformatics* **2014**, *30* (6), 792-800.
42. (a) Kleywegt, G. J.; Jones, T. A., Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **1994**, *50* (2), 178-185; (b) Oliveira, S. H.; Ferraz, F. A.; Honorato, R. V.; Xavier-Neto, J.; Sobreira, T. J.; de Oliveira, P. S., KVFinder: steered identification of protein cavities as a PyMOL plugin. *BMC Bioinform.* **2014**, *15* (1), 1-8; (c) Coleman, R. G.; Sharp, K. A., Protein pockets: Inventory, shape, and comparison. *J. Chem. Inf. Model.* **2010**, *50* (4), 589-603.
43. (a) Bendel, R. B.; Afifi, A. A., Comparison of stopping rules in forward "stepwise" regression. *J. Am. Stat. Assoc.* **1977**, *72* (357), 46-53; (b) Bursac, Z.; Gauss, C. H.; Williams, D. K.; Hosmer, D. W., Purposeful selection of variables in logistic regression. *Source Code Biol. Med.* **2008**, *3* (1), 1-8.
44. A normalized z-score indicates the number of standard deviations ( $\sigma$ ) above or below the population mean ( $\mu$ ) for a given observation ( $x$ ), where  $z = (x - \mu) / \sigma$ , as discussed in: Marquandt, D., You should standardize the predictor variables in your regression models. Discussion of: A critique of some ridge regression methods. *J. Am. Stat. Assoc.* **1980**, *75* (369), 87-91.
45. E. D. Glendening, J., K. Badenhop, A. E. Reed, J. E. Carpenter, J. A. Bohmann, C. M. Morales, P. Karafiloglou, C. R. Landis, and F. Weinhold *NBO 7.0*, Theoretical Chemistry Institute, University of Wisconsin, Madison, 2018.
46. (a) Wold, S., Validation of QSAR's. *Quant. Struct.-Act. Relat.*, in *Wiley Online Library*. **1991**, *10*, 191-193; (b) Golbraikh, A.; Tropsha, A., Beware of q<sup>2</sup>! *J. Mol. Graph. Model.* **2002**, *20* (4), 269-276; (c) Gramatica, P., Principles of QSAR models validation: internal and external. *QSAR Comb. Sci.* **2007**, *26* (5), 694-701.
47. Wold, S.; Dunn III, W. J., Multivariate quantitative structure-activity relationships (QSAR): conditions for their applicability. *J. Chem. Inform. Comput. Sci.* **1983**, *23* (1), 6-13.
48. Hawkins, D. M., The problem of overfitting. *J. Chem. Inform. Comput. Sci.* **2004**, *44* (1), 1-12.
49. Kennard, R. W.; Stone, L. A., Computer aided design of experiments. *Technometrics* **1969**, *11* (1), 137-148.
50. Mayr, H.; Ofial, A. R., The Reactivity-Selectivity Principle: An Imperishable Myth in Organic Chemistry. *Angew. Chem. Int. Ed.* **2006**, *45* (12), 1844-1854.
51. Note that descriptors in a robust MLR model should not be intercorrelated. A threshold of  $R^2=0.4$  was utilized to eliminate models that contain colinear descriptors. See: (a) Farrar, D. E.; Glauber, R. R., Multicollinearity in regression analysis: The problem revisited. *Rev. Econ. Stat.* **1967**, 92-107; (b) Slinker, B. K.; Glantz, S. A., Multiple regression for physiological data analysis: the problem of multicollinearity. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **1985**, *249* (1), R1-R12.
52. (a) Note that the molecular orbital diagram in Figure 8B is qualitative and is not precisely drawn to scale. For further discussions of dirhodium electronic structure highlighting the impact of ligand field and charge state, please see the references that follow; (b) Cotton, F. A.; Dikarev, E. V.; Petrukhina, M. A., Using structures formed by dirhodium tetra(trifluoroacetate) with polycyclic aromatic hydrocarbons to prospect for maximum  $\pi$ -electron density: Hückel calculations get it right. *J. Am. Chem. Soc.* **2001**, *123* (47), 11655-11663; (b) Varela-Álvarez, A.; Yang, T.; Jennings, H.; Kornecki, K. P.; Macmillan, S. N.; Lancaster, K. M.; Mack, J. B. C.; Du Bois, J.; Berry, J. F.; Musaev, D. G., Rh<sub>2</sub>(II, III) catalysts with chelating carboxylate and carboxamidate supports: Electronic structure and nitrene transfer reactivity. *J. Am. Chem. Soc.* **2016**, *138* (7), 2327-2341; (c) Warzecha, E.; Berto, T. C.; Wilkinson, C. C.; Berry, J. F., Rhodium Rainbow: A Colorful Laboratory Experiment Highlighting Ligand Field Effects of Dirhodium Tetraacetate. *J. Chem. Educ.* **2019**, *96* (3), 571-576; (d) Warzecha, E.; Berto, T. C.; Berry, J. F., Axial ligand coordination to the C-H amination catalyst Rh<sub>2</sub>(esp)<sub>2</sub>: A structural and spectroscopic study. *Inorg. Chem.* **2015**, *54* (17), 8817-8824.
53. Scrocco, E.; Tomasi, J., The electrostatic molecular potential as a tool for the interpretation of molecular properties. In *New concepts II*, Springer: 1973; pp 95-170.
54. Suresh, C. H.; Koga, N., Quantifying the electronic effect of substituted phosphine ligands via molecular electrostatic potential. *Inorg. Chem.* **2002**, *41* (6), 1573-1578.
55. (a) Eisenhart, R. J.; Clouston, L. J.; Lu, C. C., Configuring bonds between first-row transition metals. *Acc. Chem. Res.* **2015**, *48* (11), 2885-2894; (b) Berry, J. F., Metal-Metal Bonded Compounds of the Group IX Elements. In *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*, Elsevier: 2021, p. 4-42; (c) Cooper, B. G.; Napoline, J. W.; Thomas, C. M., Catalytic applications of early/late heterobimetallic complexes. *Catal. Rev.* **2012**, *54* (1), 1-40; (d) Cotton, F. A.; Murillo, C. A.; Walton, R. A., *Multiple bonds between metal atoms*. Springer Science & Business Media: 2005, p. 465-632; (e) Collman, J. P.; Barnes, C. E.; Woo, L. K., Systematic variation of metal-metal bond order in metalloporphyrin dimers. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80* (24), 7684-7688.
56. We use this language to differentiate the effects discussed here from conventional stereoelectronic effects, where molecular structure and reactivity is impacted by spatial orbital overlap considerations. Here the intertwinement is between global structure and Rh energetic orbital overlap considerations, which results in perturbations to catalyst electronic properties (i.e., RhRh  $\sigma^* E$ ) that impact reactivity.

57. Although  $RhRh \Delta E_{\pi^*/\sigma^*}$  is colinear with  $RhRh \sigma^* E$ , the former fits the data slightly better in the improved model and, in principle, accounts for both  $\sigma$ -bonding and  $\pi$ -backbonding effects.

58. (a) Bell, E. W.; Zhang, Y., DockRMSD: an open-source tool for atom mapping and RMSD calculation of symmetric molecules through graph isomorphism. *J. Cheminformatics* **2019**, *11* (1), 1-9; (b) Baber, J. C.; Thompson, D. C.; Cross, J. B.; Humblet, C., GARD: a generally applicable replacement for RMSD. *J. Chem. Inf. Model.* **2009**, *49* (8), 1889-1900; (c) Coutsiaris, E. A.; Seok, C.; Dill, K. A., Using quaternions to calculate RMSD. *J. Comput. Chem.* **2004**, *25* (15), 1849-1857; (d) Kabsch, W., A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. Sec. A* **1976**, *32* (5), 922-923; (e) Schulz-Gasch, T.; Schärfer, C.; Guba, W.; Rarey, M., TFD: Torsion fingerprints as a new measure to compare small molecule conformations. *J. Chem. Inf. Model.* **2012**, *52* (6), 1499-

1512; (f) Bero, S. A.; Muda, A. K.; Choo, Y. H.; Muda, N. A.; Pratama, S. F., Similarity Measure for Molecular Structure: A Brief Review. *J. Phys. Conf. Ser.* **2017**, *892*, 012015.

59. We thank the reviewers for insightful comments that led to the inclusion of more explicit comparisons of models with SMART descriptors to those with traditional steric descriptor types. To minimize fortuitous training/validation partitioning bias in these statistical metric comparisons between models with different descriptor types, difference values (i.e.,  $\Delta R^2$ ,  $\Delta Q^2$ ,  $\Delta 4$ -fold, and  $\Delta \text{test } R^2$ ) were tabulated based on the average metrics across three different 20:5 TS/Vs splits (see SI section 8 and Table S13).

For Table of Contents Only:

