Leveraging Regio- and Stereoselective C(sp^3)–H Functionalization of Silyl Ethers to Train a Logistic Regression Classification Model for Predicting Site-Selectivity Bias

Yannick T. Boni[1,‡], Ryan C. Cammarota[2,‡,#], Kuangbiao Liao[1,$], Matthew S. Sigman*[,2] and Huw M. L. Davies*[,1]

[1]Department of Chemistry, Emory University, 1515 Dickey Drive, Atlanta, Georgia 30322, United States

[2]Department of Chemistry, University of Utah, 315 South 1400 East, Salt Lake City, Utah 84112, United States

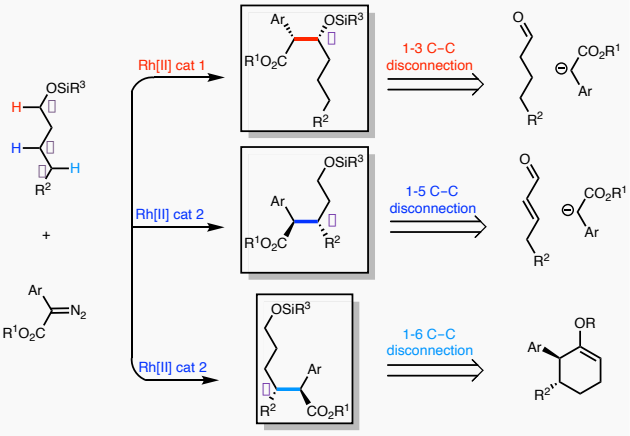
ABSTRACT: The C–H functionalization of silyl ethers via carbene-induced C–H insertion represents an efficient synthetic disconnection strategy. In this work, site- and stereoselective C(sp^3)–H functionalization at α, γ, δ and even more distal positions to the siloxy group has been achieved using donor/acceptor carbene intermediates. By exploiting the predilections of Rh_2(*R*-TCPTAD)$_4$ and Rh_2(*S*-2-Cl-5-BrTPCP)$_4$ catalysts to target either more electronically activated or more spatially accessible C–H sites, respectively, divergent desired products can be formed with good diastereocontrol and enantiocontrol. Notably, the reaction can also be extended to enable desymmetrization of meso silyl ethers. Leveraging the broad substrate scope examined in this study, we have trained a machine learning classification model using logistic regression to predict the major C–H functionalization site based on intrinsic substrate reactivity and catalyst propensity for overriding it. This model enables prediction of the major product when applying these C–H functionalization methods to a new substrate of interest. Applying this model broadly, we have demonstrated its utility for guiding late-stage functionalization in complex settings and developed an intuitive visualization tool to assist synthetic chemists in such endeavors.

Introduction

To maximize the applications of C–H functionalization in synthesis, one needs to develop a synthetic logic akin to that of classic retrosynthetic disconnections.[1] Many classic reactions exploit carbonyl chemistry to access products with defined patterns of functional groups, such as 1,3-heteroatom disconnections using C–H bond forming aldol (or Mannich) reactions[2] and 1,5-disconnections via Michael additions.[3] Indeed, these patterns are so ingrained into the logic of organic synthesis that computationally-guided programs are available for identifying such disconnections.[4] In contrast, the rapidly developing field of C–H functionalization does not follow the same logic. Elegant advances have leveraged appropriately positioned directing groups to control which C–H bond is functionalized, including selective C(sp^2)–H activation at *ortho*, *meta*, or *para* positions[5] and C(sp^3)–H functionalization at specific positions along a hydrocarbon chain.[6] A complementary and more versatile approach would be the development of catalyst-controlled C–H functionalization without prior substrate coordination. This tactic diverges from both directed C–H functionalization and classic disconnection strategies in that particular patterns of functional groups are no longer required for formulating an effective retrosynthesis.[7] Instead, the challenge is developing catalysts capable of precisely and predictably reacting with specific C–H bonds in complex settings.

In this context, considerable recent progress has been achieved in designing small molecule[8] and enzymatic[9] catalysts for site-selective C–H functionalization. In particular, one of our labs has developed a wide range of dirhodium catalysts that enable selective intermolecular C–H functionalization via donor/acceptor carbenes.[10] Recent highlights include the ability to selectively functionalize unactivated primary, secondary, or tertiary sites by choosing the appropriate catalyst to override the innate reactivity profile of the substrate.[11] We have also previously demonstrated how carbene-induced reactions at activated C–H bonds can be considered as surrogates to classic reactions, such as the aldol reaction,[12] Mannich reaction,[13] Claisen condensation,[14] Michael addition,[15] Claisen rearrangement,[14a] vinylogous Mukaiyama reaction,[16] and vinylogous Michael addition.[17] Herein, we describe how site-selective C–H functionalization of alkyl silyl ethers, a traditionally problematic substrate class (*vide infra*), can affect either 1,3-, 1,5-, or 1,6-disconnections depending on catalyst choice (Scheme 1). This chemistry relies on suitable chiral catalysts that can control where C–H functionalization will occur. A sterically unencumbered catalyst would be expected to react at the α-position to the siloxy group, which is the electronically preferred position, leading to the equivalent of a 1,3-disconnection. In contrast, a bulkier catalyst would be expected to react at a less crowded distal position, leading to the equivalent of a 1,5- or 1,6-disconnection through a selective reaction at the most accessible methylene or methine C–H bond.

Scheme 1. Retrosynthetic disconnections from catalyst-controlled site-selective C–H functionalization of silyl ethers



Catalyst preferences, however, cannot completely invert substrate reactivity profiles for all conceivable substrates of interest. In practice, the intrinsic reactivity of each C–H site and the extent of the bias exerted by the catalyst must both be precisely understood to develop a more versatile synthetic logic for C–H functionalization. In this regard, we saw an opportunity to leverage the broad scope examined in this study to train a machine learning classification model for site-by-site assessment of intrinsic substrate reactivity[18] and the propensity of catalysts for overriding it. Herein, we demonstrate the robustness of our model and the accompanying intuitive visualization tool for predicting how any given C–H functionalization reaction may proceed when mediated by two catalysts with disparate properties, Rh_2(*R*-TCPTAD)$_4$ and Rh_2(*S*-2-Cl-5-BrTPCP)$_4$. Finally, we showcase the synthetic utility of our model by applying it to cholesteryl pelargonate,[11b] a molecule with many distinct C–H sites that could potentially be functionalized. More broadly, we anticipate that the complementary synthetic strategies developed, coupled with the visual framework presented for predicting reactivity with new substrates of interest, will provide the foundational synthetic logic for enabling more widespread application of late-stage C–H functionalization in increasingly complex settings.

Results and Discussion

We have previously conducted Rh_2(*S*-DOSP)$_4$-catalyzed reactions at activated sites *alpha* to the siloxy group in allylic and benzylic silyl ethers.[12a, 12c] However, our previous efforts to extend the reaction to unactivated silyl ethers were less successful.[12b] Although tetraalkoxysilanes were effective substrates, functionalizing silylated alcohols proved especially problematic.[12c] Substrate scope was primarily limited to labile TMS ethers, and the esoteric solvent 2,2-dimethylbutane was required to avoid deleterious reactivity that was observed in other hydrocarbon solvents. Even under these specialized conditions required for achieving site- and stereoselective reactivity for a limited scope, the yields obtained were low (≤45%).

Since these earlier studies, we have demonstrated that replacing methyl esters with trichloroethyl esters as the carbene acceptor group greatly enhances the efficiency of C–H functionalization reactions at unactivated C–H bonds.[19] We have also recently developed a wide range of chiral dirhodium catalysts,[20] among which Rh_2(*R*-TCPTAD)$_4$ and Rh_2(*S*-2-Cl-5-BrTPCP)$_4$ are two of the most notable. The carboxylate ligands in these catalysts self-assemble to form relatively rigid bowl shapes that are pseudo C_4 symmetric.[21] Computational studies indicate that the reactive dirhodium-carbene is generated inside the bowl, and that site-selectivity is influenced by the relative ease of substrate approach for functionalization at different positions.[11b, 21] As can be seen in Figure 1, the Rh_2(*R*-TCPTAD)$_4$ bowl has a wider aperture and larger volume than the Rh_2(*S*-2-Cl-5-BrTPCP)$_4$ bowl based on "Spatial Molding for Approachable Rigid Targets" (SMART) analysis.[22] This steric influence is a key reason why Rh_2(*R*-TCPTAD)$_4$ drives reactions to occur at more congested tertiary C–H sites, whereas Rh_2(*S*-2- Cl-5-BrTPCP)$_4$ biases reactivity toward less activated but more accessible sites.

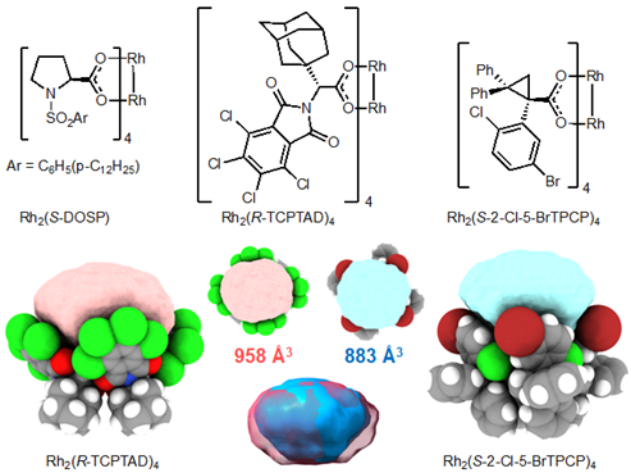
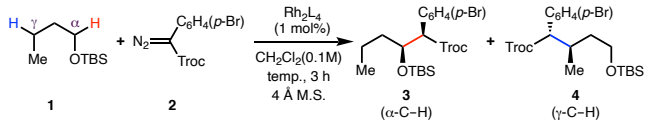


Figure 1. Chiral catalysts used in these studies and illustrations of their bowl-shaped structures. The SMART cavity volumes for Rh_2(*R*-TCPTAD)$_4$ and Rh_2(*S*-2-Cl-5-BrTPCP)$_4$ are 958 and 883 Å^3, respectively. The role of disparate electronic carbene stabilization in their divergent reactivity was also recently highlighted.[22]

These recent catalyst and methodological advances prompted us to re-evaluate the C–H functionalization of unactivated silyl ethers. We hypothesized that reactions may proceed in higher yields using trichloroethyl aryldiazoacetates and that catalyst choice could control site-selectivity. We began by examining the reaction of TBS-protected butanol (**1**) with 2,2,2-trichloroethyl-2-(4-bromophenyl)-2-diazoacetate (**2**, Table 1). The TBS group was expected to cause greater steric interference for functionalization *alpha* to the siloxy group compared to the previously used TMS group,[12c] but the resulting silylated products would be more readily isolated and purified.

These investigations revealed that reactions mediated by Rh_2(*R*-TCPTAD)$_4$ in CH_2Cl_2 at 5 °C resulted in good yields and site-selectivity for α-functionalized product **3** (entry 3). In contrast, Rh_2(*S*-2-Cl-5-BrTPCP)$_4$ gave high site-selectivity for the γ-functionalization product (**4**, entries 4-5). This divergent reactivity is consistent with the inherent differences between Rh_2(*R*-TCPTAD)$_4$ and Rh_2(*S*-2-Cl-5-BrTPCP)$_4$ (*vide supra*), where the latter's more constrained reactive site cavity and less electrophilic carbene moiety promotes reactivity at the most accessible secondary C–H bond. The Rh_2(*S*-2-Cl-5-BrTPCP)$_4$-catalyzed γ-selective C–H functionalization of **1** could be extended to a series of aryldiazoacetates, resulting in similar site-selectivity and diastereoselectivity (Table S2).

Table 1. Selected Catalyst Optimization Studies

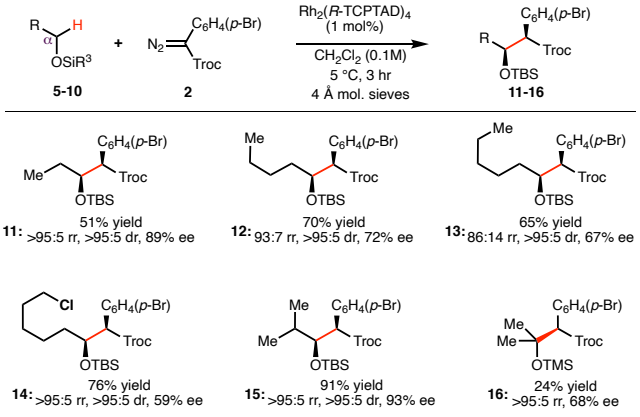


Entry	L	Temp. (°C)	Yield (%)[b]	r.r. (3:4)[c]	d.r.[d]	ee (%)[e]
1	*S*-DOSP	22	75	90:10	94:6	23
2	*R*-TCPTAD	22	70	92:8	>95:5	44
3	*R*-TCPTAD	5	82	>95:5	>95:5	64
4	*S*-2-Cl-5-BrTPCP	22	95	8:92	88:12	90
5	*S*-2-Cl-5-BrTPCP	5	92	7:93	89:11	90

[a]Reaction conditions: **1** (0.600 mmol), **2** (0.200 mmol), Rh_2L_4 catalyst (1 mol%), 3 h slow addition time of **2** under a nitrogen atmosphere. [b]Combined crude ^{1}H NMR yields of all regioisomers using trichloroethylene as internal standard. [c,d]Regioselectivity and diastereoselectivity determined from crude ^{1}H NMR. [d]Diastereoisomeric ratio of major regioisomer. [e]Enantiomeric excess determined from chiral HPLC traces.

Having established Rh_2(*R*-TCPTAD)$_4$ as the optimal catalyst for α-C–H functionalization, we proceeded to test a range of silyl ethers to generate products with oxygen functionality in a 1,3 arrangement (Scheme 2). The reactions of **2** with linear alkyl silyl ethers (**5**-**8**) catalyzed by Rh_2(*R*-TCPTAD)$_4$ afforded the α-C–H functionalization products **11**-**14** with excellent site- and diastereoselectivity (>95:5 r.r., >95:5 d.r.). The enantioselectivity was variable (59-89%), however, with the highest ee among linear substrates obtained for **11**.

Scheme 2. α-Selective functionalization of silyl ethers

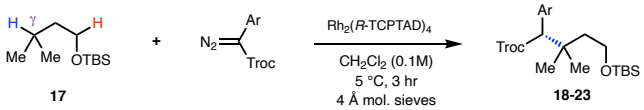


[a] Reaction conditions: silyl ether (0.600 mmol), **18** (0.200 mmol), Rh_2(*R*-TCPTAD)$_4$ (1 mol %), ~200 wt% of 4 Å mol. sieves, 3 h reaction time under a nitrogen atmosphere at 5 °C. Indicated yields are isolated yields.

The influence of substrate branching on enantioselectivity is evident for isobutyl silyl ether **9**, from which **15** is generated in excellent yield (91%) and selectivity (>95:5 r.r., >95:5 d.r., 93% *ee*). In the case where the silyl ether features a tertiary C–H bond *alpha* to the siloxy group (**10**), C–H functionalization is still observed when using the less bulky trimethylsilyl (TMS) protecting group, but the reaction is less efficient, as **16** was obtained in only 24% yield.

Beyond reactivity patterns for secondary C–H bonds α versus γ to a siloxy group, we next probed competition between a γ–tertiary site and a α-secondary site. For 3-methylbutyl silyl ether **17**, the γ-tertiary site would not be suitable for functionalization with a bulky catalyst such as Rh_2(*S*-2-Cl-5-BrTPCP)$_4$. Thus, the less hindered Rh_2(*R*-TCPTAD)$_4$ catalyst was examined. At the outset, it was uncertain which site(s) in **17** would be reactive because Rh_2(*R*-TCPTAD)$_4$ is capable of functionalizing either the tertiary C–H bond or the electronically activated methylene site *alpha* to the siloxy group. We were pleased to observe that highly site-selective reactions occurred for all diazo precursors examined with **17**, favoring γ-methine C–H functionalization to form 1,5-products **18-23** with a quaternary center at the γ-position (Table 2). Furthermore, enantioselectivity was high for a range of aryldiazoacetates (84-96% ee), albeit lower for unsubstituted phenyldiazoacetate (63% ee).

Table 2. γ-methine selective functionalization



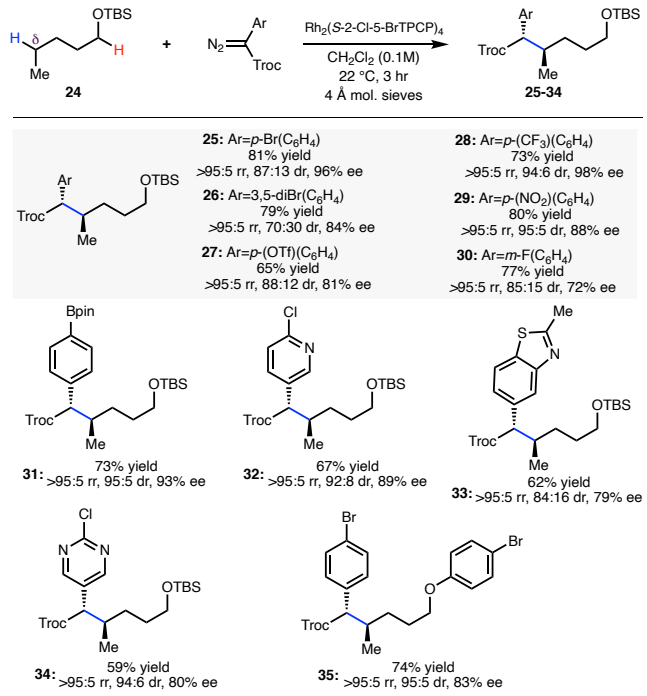
Product	Ar	Yield (%)[b]	r.r. (3:4)[c]	e.e. (%)[e]
18	C_6H_5	67	>95:5	63
19	*p*-Br(C_6H_4)	89	>95:5	96
20	*p*-I(C_6H_4)	66	>95:5	96
21	*p*-CF3(C_6H_4)	59	>95:5	90
22	*p*-OTf(C_6H_4)	62	>95:5	84
23	*p*-Ph(C_6H_4)	44	>95:5	84

[a]Reaction conditions: **17** (0.600 mmol), aryldiazoacetate (0.200 mmol), Rh_2L_4 catalyst (1 mol%), 3 h slow addition time of aryldiazoacetate at 5 °C under a nitrogen atmosphere.

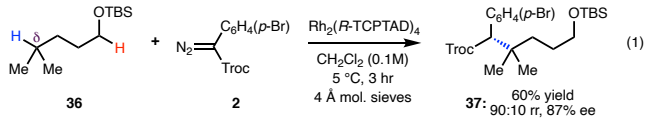
Selective functionalization of δ-C–H bonds was also explored, using pentyl silyl ether **24** as the model substrate. It was anticipated that subjecting **24** to catalytic conditions with Rh_2(S-2-Cl-5-BrTPCP)$_4$ would result in selective reactions at the most accessible methylene site, which in this case would be the δ position. This indeed was found to be the case, and the reaction was extended to a variety of aryl and heteroaryldiazoacetates to form products **25**-**34** in high site-selectivity (>95:5 r.r.; Scheme 3).

Enantioselectivity was dependent on the nature of the aryl group (72 to 98% ee), and the highest diastereoselectivity was observed with electron-deficient *para*-substituted aryl groups (up to 95:5 d.r.). Notably, pinacolborane derivative **31** and heterocyclic derivatives **32**-**34** can also be readily formed in high d.r. using this chemistry. Even though TBS derivatives have been predominately used in this study, other ethers are also compatible, as illustrated by the formation of phenolic derivative **35**. Extension of the δ-C–H functionalization to 4-methylpentyl silyl ether **36** was possible by switching the catalyst to Rh_2(*R*-TCPTAD)$_4$ (eq 1). The reaction of **36** with **2** under these conditions generated **37** in 60% yield with good site- and enantioselectivity (90:10 r.r., 87% ee), although there was a small amount of competing C–H functionalization observed α to oxygen.

Scheme 3. δ-Selective functionalization of silyl ether 24

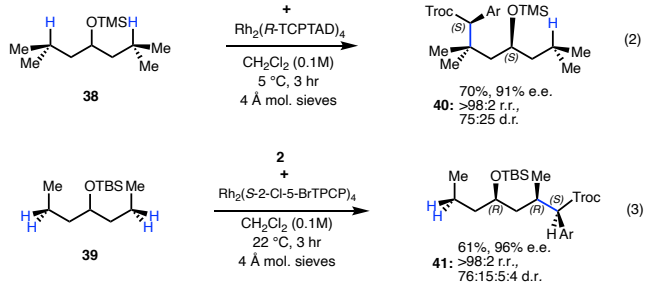


[a] Reaction conditions: **24** (0.600 mmol), aryldiazoacetate (0.200 mmol), Rh_2(*S*-2-Cl-5-BrTPCP)$_4$ (1 mol %), ~200 wt% of 4 Å mol. sieves, 3 h reaction time under a nitrogen atmosphere at 22 °C. Indicated yields are isolated yields.

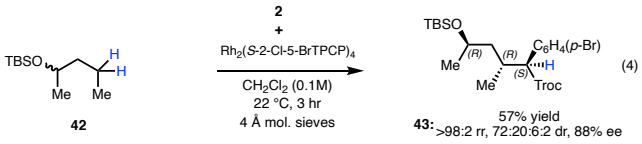


(1)

We next tested the possibility of using C–H functionalization to achieve the desymmetrization of meso silyl ethers **38** and **39** (eq 2-3), thereby enabling additional stereocenters to be set in a single step.[23] The reaction with **38** explored the competition between two enantiotopic γ-methine sites. The Rh_2(*R*-TCPTAD)$_4$-catalyzed reaction with **2** furnished the desired product **40** in 70% yield with excellent site selectivity (>98:2 r.r.) and high enantioselectivity (91% ee). The diastereoselectivity of the reaction was 75:25 d.r., indicating that even though the system is acyclic, Rh_2(*R*-TCPTAD)$_4$ is able to differentiate between the enantiotopic sites. The reaction of **2** with meso silyl ether **39** is more complicated because two sets of diastereomers could be formed: one through desymmetrization, and another because the reaction occurs at a methylene site. In this case, Rh_2(*S*-2-Cl-5-BrTPCP)$_4$ was the optimal catalyst, and the reaction was highly site-selective (>95:5 r.r.). The reaction favored the formation of the major diastereomer **41** in 96% ee, although in total four diastereomers were observed in a ratio of 76:15:5:4 (see the SI for details on the stereochemical assignments for **40**-**41**).

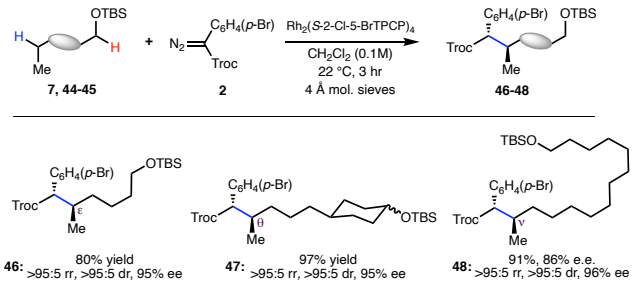


The reaction with racemic silylated 2-pentanol **42** was also examined to determine if the process could be susceptible to kinetic resolution (eq 4). In this case, our interest was the stereochemical outcome for product **43** rather than enantioenrichment of the starting material. The reaction with Rh_2(*S*-2-Cl-5-BrTPCP)$_4$ gave the expected γ-C–H functionalization product with excellent site-selectivity (>95:5 r.r.). The reaction gave rise to the preferred formation of one major diastereomer in 72:20:6:2 d.r., where the major diastereomer was produced in 88% ee. The observed stereochemical outcome was assigned through NMR analysis and by analogy to that of the desymmetrization reactions (see SI for details).



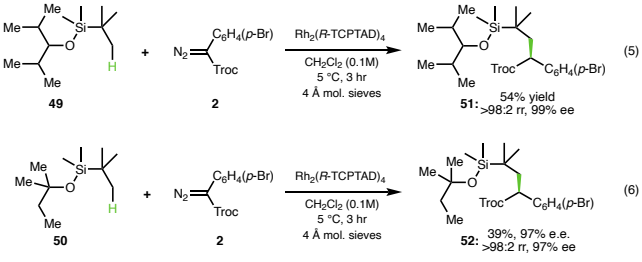
The strong propensity of Rh_2(*S*-2-Cl-5-BrTPCP)$_4$ to override the substrate electronic preference for the α-siloxy C–H site led us to pursue functionalization of remote C–H bonds beyond the δ-position (Scheme 4). Rh_2(*S*-2-Cl-5-BrTPCP)$_4$ maintained its inherent site-selectivity for the most accessible methylene C–H bond to afford stereoselective formation of products **46**-**48** (up to 95:5 r.r. and >95:5 d.r.). Enantioselectivity for these longchained aliphatic silyl ethers remained high in all cases (86-95 ee). For product **47**, both the *cis*- and *trans*- isomers of the cyclohexyloxy-*tert*-butyldimethylsilane reacted in a 1:1 ratio. However, carbene C–H insertion was exclusively observed at the terminal methylene site (>95:5 r.r.) to generate **47**.

Scheme 4. Catalyst-controlled site-selective functionalization beyond the δ-C–H site



[a] Reaction conditions: **44-46** (0.600 mmol), **2** (0.200 mmol), Rh_2(*S*-2-Cl-5-BrTPCP)$_4$ (1 mol %), ~200 wt% of 4 Å mol. sieves, 3 h reaction time under a nitrogen atmosphere at 22 °C. Indicated yields are isolated yields.

During the process of developing these transformations, we also observed that substrates lacking sterically accessible secondary or tertiary C–H bonds (**49** and **50**) were susceptible to Rh_2(*R*-TCPTAD)$_4$-catalyzed carbene insertion into the primary C–H bonds of the silyl protecting groups (eq 5-6). These reactions proceed to form **51** and **52** with very high levels of enantioinduction (97-99% ee). Intermolecular C–H functionalization of primary C–H bonds is electronically challenging, but presumably a favorable β-silicon effect enables these reactions to occur.[24]



Leveraging Scope Data to Train a Machine Learning Model for Site-Selectivity Classification

The general site-selectivity trends observed in these studies are summarized in Figure 2. Rh_2(*R*-TCPTAD)$_4$ typically biases C–H functionalization α to the siloxy group, but reactivity can also occur preferentially at a distal tertiary site (e.g., **17**, **36**, and **38**). In contrast, Rh_2(*S*-2-Cl-5-BrTPCP)$_4$ mediates preferential C–H functionalization at the most accessible methylene site. If the secondary and tertiary sites are too sterically crowded, then C–H functionalization can be diverted to primary C–H sites on the silyl protecting group (e.g., **49**-**50**).

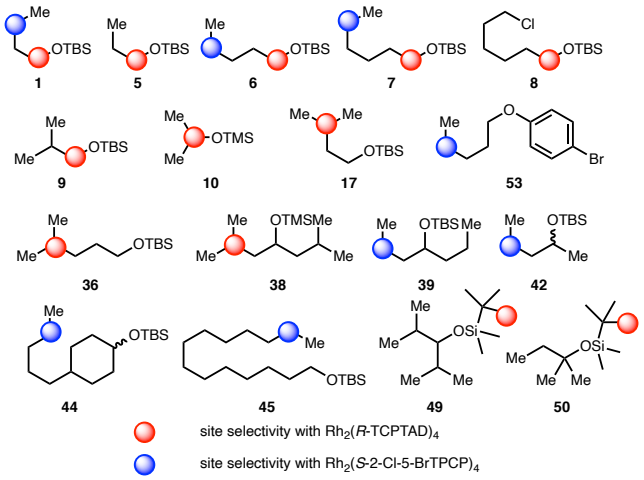


Figure 2. Site-selectivity trends for unactivated silyl ethers.

We recently developed a predictive model for C–H functionalization site-selectivity as a function of catalyst properties.[22] This model considered a diverse dirhodium catalyst set for a constant substrate (1-bromo-4-pentylbenzene) and diazo precursor (2,2,2-trichloroethyl-2-(4-bromo-phenyl)-2-diazoacetate). The data displayed in Figure 2 now allows us to analyze how intrinsic substrate reactivity profiles intersect with the divergent reactivity biases exerted by Rh_2(*R*-TCPTAD)$_4$ and Rh_2(*S*-2-Cl-5-BrTPCP)$_4$ catalysts. We envisioned that, by leveraging this scope data, a machine learning model could be trained to consider all possible C(sp^3)–H sites within a complex substrate and predict which is the most likely to be functionalized by a given catalyst based on the site-by-site[25] computed local chemical properties. A synthetic chemist could then use this tool to predict which late-stage intermediate in their synthesis would be most amenable for a desired selective C–H functionalization.

Among supervised learning algorithms, logistic regression is uniquely suited for providing intuitive insights into binary classifications for moderately sized data sets that are the norm in the chemical sciences.[26] Thus, we set out to develop a logistic regression model for predicting the probability that a given C(sp^3)–H site in a substrate would be the major site of functionalization. We began our analysis by calculating, at the M062X-D3/def2tzvp-SMD(CH_2Cl_2) // M062X-D3/6-311G** level of theory,[27] the structures, energetics, and chemical properties of representative conformers (MacroModel/OPLS3e)[28] of the 17 silyl ether substrates in Figure 2. To maximize the chemical space for which our model will hold predictive utility, additional substrate diversity was added by back-filling the data set with 18 substrates for which C–H functionalization data with Rh_2(*R*-TCPTAD)$_4$ and/or Rh_2(*S*-2-Cl-5-BrTPCP)$_4$ catalysts has previously been obtained (Figure S23).[11b,29] To quantify the typical spatial accessibility of each C–H site across the substrate conformations present in solution, the Boltzmann-averaged percent buried volume[30] ($\%V_{bur,\ Boltz}$) within a sphere of 2 Å radius iteratively centered at each hydrogen atom was used (Figure 3a, left).[31] For highly flexible substrates, however, the least sterically hindered conformation that the substrate can adopt (i.e., $\%V_{bur,\ min}$; Figure 3a, right) could also bely insight into describing the C–H insertion transition state (Figure S24).[32]

For electronic descriptors of C–H bonds,[33] we initially considered $\Delta E_{\sigma/\sigma^*}$ (C–H) values from Natural Bond Orbital (NBO) analysis,[34] where larger energy gaps between C–H σ-bonding and σ-antibonding orbitals are indicative of increased electronic activation.[35] This descriptor type was recently found to capture C(sp^3)–H bond electron-richness as it pertained to ease of oxidation for aliphatic substrates with remote electron-withdrawing groups.[36] However, upon diversifying the substrate scope in this study, we find that descriptions based on $\Delta E_{\sigma/\sigma^*}$ (C–H) are not necessarily generalizable. While $\Delta E_{\sigma/\sigma^*}$ (C–H) captures inductive deactivation (e.g., by an α-halogen in Figure 3b) and hyperconjugation effects that dictate relative partial carbocation stability (i.e., 3° > 2° > 1°),[37] its values for neutrally-charged substrate molecules do not capture resonance effects wherein proximal π-electrons can stabilize the buildup of partial positive charge in the C–H insertion transition state.[38] This is evident from the inability of $\Delta E_{\sigma/\sigma^*}$ (C–H) values to differentiate unactivated 2° C–H bonds from benzylic and allylic C–H bonds, as well as from C–H α to an oxygen atom (e.g., in Figure 3b).

Resonance stabilization effects can be recouped to some extent by combining computed $\Delta E_{\sigma/\sigma^*}$ (C–H) and ^{1}H NMR chemical shift (δ) values[39] into a composite descriptor, which we have termed the Electronic Activation Index (EAI, Figure 3b). Although ^{1}H NMR δ do depend intimately on inductive deshielding,[40] resonance stabilization effects on C–H insertion are also captured indirectly, since neighboring electronegative atoms invariably have lone-pairs that can stabilize positive charge buildup at α-carbon atoms through resonance (see SI for further discussion). Similarly, anisotropic effects of alkene and aryl π-systems[41] also result in relatively downfield ^{1}H NMR shifts that track with increased propensity for transition state stabilization in allylic and benzylic C–H insertions, respectively.[17b, 20]

The prediction of reactivity as a composite of electronic induction and resonance effects is well-precedented in physical organic chemistry, as exemplified by Hammett's foundational σ substituent constants (Figure 3c).[42] For substituted arenes, it is well understood that resonance effects are comparatively more influential relative to induction for substitution *para* (50:50) to the reaction center (rc) as opposed to *meta* (23:77), where the relative coefficients for resonance and induction contributions to σ constants were determined empirically to best predict reactivity (Figure 3c).[43] Taking inspiration from Hammett's σ constants, we sought to empirically determine EAI coefficients for contributions from $\Delta E_{\sigma/\sigma^*}$ (C–H) and ^{1}H

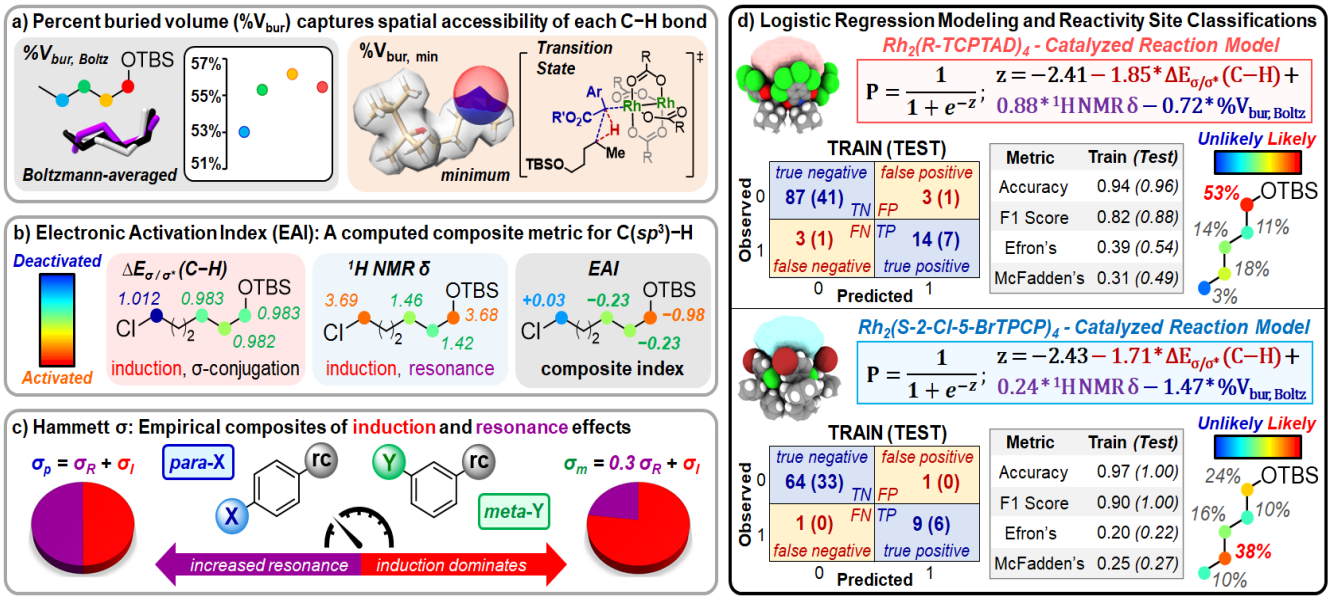


Figure 3. Logistic regression classification modeling results based on intuitive input descriptors for the relative reactivity of C(sp^3)–H bonds. Note that the EAI values shown in Figure 3b are based on the logistic regression model coefficients for input electronic descriptors in the Rh_2(*R*-TCPTAD)$_4$-catalyzed model. Training/test set splits (see SI) resulted in 107 of 157 and 75 of 114 individual C–H sites chosen for training the Rh_2(*R*-TCPTAD)$_4$ and Rh_2(*S*-2-Cl-5-BrTPCP)$_4$ models, respectively (i.e., 17:8 and 10:6 substrate splits). Note that for the selected validation set predictions shown in Figure 3d, the remainders of the normalized probability distributions (where ΣP_{NORM} = 100% within each substrate) that are not explicitly shown are for TBS protecting group functionalization (≤ 2% predicted in each case).

NMR δ descriptors such that the relative reactivity of C–H sites toward carbene insertion could be successfully predicted. Logistic regression provides the mathematical framework for empirically fitting these coefficients.[44] Thus, we included these two normalized electronic descriptors, along with $\%V_{bur,\ Boltz}$, to assess if intuitive chemical properties could enable robust classification of which C–H site in a given substrate would be the major site of functionalization.

Logistic regression modeling was performed separately for reactions mediated by Rh_2(*R*-TCPTAD)$_4$ and Rh_2(*S*-2-Cl-5-BrTPCP)$_4$ to highlight the sensitivity of each catalyst's reactivity preferences to C–H electronic activation and spatial accessibility. The predicted probability (*P*) of each C–H site being the major site for functionalization is given based on the sigmoidal function, where *P* approaches 1 as *z* approaches infinity (Figure 3d). A modified logistic regression classification protocol was employed, wherein the predicted probabilities that each C–H site will be primarily functionalized are compared and the site with maximum probability within a given substrate is classified as the predicted major site for reactivity (see SI for details). Thus, C–H sites with normalized local properties that result in large positive z values will have high predicted probabilities for being the most reactive site within a given substrate. Notably, the parsing of substrate scope data in terms of considering local C–H sites that either are (1), or are not (0), the major functionalization site enables an adequate observation to parameter ratio for modeling.[45] The resulting logistic regression models and classification confusion matrices for training and test sets,[46] along with model statistical performance metrics and selected validation predictions, are shown in Figure 3d.

Both the Rh_2(*R*-TCPTAD)$_4$-catalyzed and Rh_2(*S*-2-Cl-5-BrTPCP)$_4$-catalyzed models provide high accuracy, precision, and recall (i.e., F1 score) for the training set (Figure 3d). High F1 scores are especially important given the unbalanced nature of the data sets in favor of C–H sites that are not functionalized.[47] Additionally, high pseudo-R^2 values show that the models provide significantly improved predictions relative to those for Efron´s and McFadden´s definitions of the null hypothesis.[48] Compared to multivariate linear regression, where models often feature R^2 values > 0.9, pseudo-R^2 values > 0.2 for logistic regression are indicative of an excellent fit to the classification data.[49] Importantly, both models maintain their robust performance across their respective validation sets, thereby suggesting that they are not the result of overfitting the training set data.

Beyond the favorable statistical performance of the models, their coefficients contain chemical insights into the divergent reactivity mediated by Rh_2(*S*-2-Cl-5-BrTPCP)$_4$ and Rh_2(*R*-TCPTAD)$_4$ catalysts. The negative coefficients for $\%V_{bur,\ Boltz}$ in both models are consistent with our expectation that more spatially accessible C–H sites (i.e., negative normalized $\%V_{bur,\ Boltz}$ values) will be more likely to be the major site of functionalization (i.e., larger positive z and P values). The coefficient magnitude for $\%V_{bur,\ Boltz}$ in the Rh_2(*S*-2-Cl-5-BrTPCP)$_4$ model is approximately double that in the Rh_2(*R*-TCPTAD)$_4$ model, which indicates that the preferred site for functionalization using Rh_2(*S*-2-Cl-5-BrTPCP)$_4$ is significantly more sensitive to relative C–H spatial accessibility. Conversely, Rh_2(*R*-TCPTAD)$_4$ biases reactivity toward more electronically activated C–H sites, as it is less sensitive to C–H bond steric congestion. Among electronic descriptors, smaller $\Delta E_{\sigma/\sigma^*}$ (C–H) gaps were found to be more important than downfield ^{1}H NMR shifts for predicting the major site of C–H functionalization with both catalysts. Selected validation set model predictions are shown for silyl ether **6** (as normalized P values in Figure 3d), where subsequent classification according to the maximum predicted probability within **6** correctly predicts the major sites of reactivity for both catalysts. It is worth noting that a more conservative classification model, where a prediction is made only if the difference between the maximum functionalization probability and that of the next most likely C–H site surpasses a threshold of 0.2, provides an assessment of prediction certainty by eliminating all classification errors with 50% model decisiveness.

To illustrate the broader synthetic utility of our model for predicting reactivity in unseen and increasingly complex settings,[31, 50] we challenged it by considering the reported Rh_2(*R*-TCPTAD)$_4$-catalyzed functionalization of cholesteryl pelargonate,[11b] which has many possible C(sp^3)–H bonds that could potentially be functionalized. In doing so, we also sought to demonstrate that our model can be readily transformed into an intuitive visualization tool that clearly illustrates how catalyst bias interfaces with intrinsic substrate reactivity profile to dictate C–H susceptibility to functionalization. We have developed two-dimensional charts plotting relative spatial accessibility (%$V_{bur,\ Boltz}$) against relative electronic activation index (EAI) for any given substrate, where all possible C–H sites are represented as individual data points according to their computed local chemical properties. This visual analysis is conceptually similar to Tolman's seminal analysis of phosphine ligands,[51] and is presented for cholesteryl pelargonate in Figure 4.

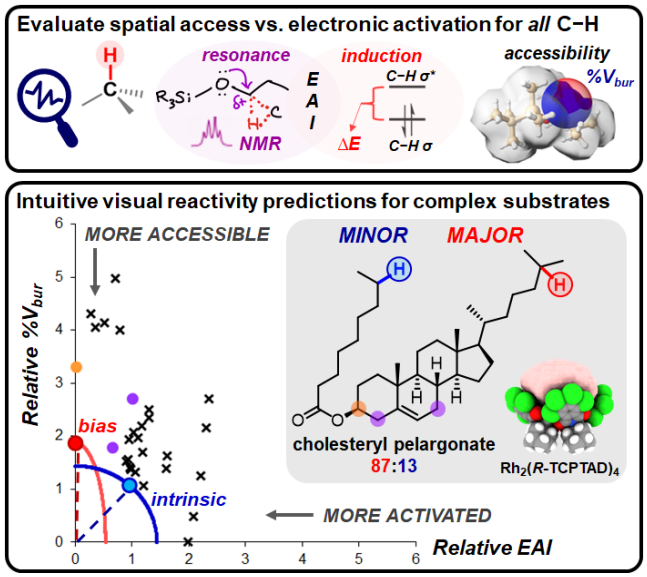


Figure 4. Plot showing the influence of Rh_2(*R*-TCPTAD)$_4$ catalyst bias in the C–H functionalization of cholesteryl pelargonate. EAI is based on logistic regression coefficients for input electronic descriptors, both of which capture inductive and resonance effects to some extent. The circular and oval projections are also based on the relative catalyst sensitivity to %V_{bur} and EAI obtained from the model. Each unit on the relative scales represents one standard deviation in that property across the C–H bond library (see Figures S22-S23). Sites not functionalized are denoted (x), and noteworthy allylic and tertiary sites that also were not functionalized are specifically highlighted.

In this visualization, the relative scales have been constructed such that a point at the origin would represent a C–H site that is both the most spatially accessible and the most electronically activated site within the substrate of interest. Given the inherent trade-off between spatial accessibility (1° > 2° > 3°) and electronic activation (3° > 2° > 1°) for C–H bonds (R^2 = 0.48;[52] Figure S25),[53] it is unlikely that any site will be simultaneously the most accessible and activated. In the absence of catalyst bias, if spatial accessibility and electronic activation were equally important for predicting the site of C–H insertion, then functionalization would be expected to occur at whichever C–H site has the smallest distance to the origin on this plot. For cholesteryl pelargonate, this would be the most accessible secondary C–H site shown in blue, where the distances from other C–H sites to the perimeter of the blue circle helps illustrate how much less intrinsically susceptible they each are to functionalization (Figure 4).

However, a catalyst may significantly bias reactivity away from this intrinsic substrate selectivity profile (Figure S26). Because Rh_2(*R*-TCPTAD)$_4$ can tolerate increased steric hindrance to functionalize tertiary C–H bonds, it will mediate preferential functionalization of the accessible tertiary site shown in red for cholesteryl pelargonate. This outcome was successfully predicted as an external validation using our logistic regression model (P=67%, P_{NORM} = 18%; see SI). Projecting an oval onto this plot can visually account for Rh_2(*R*-TCPTAD)$_4$ catalyst bias, where electronically activated but less accessible C–H sites are treated as effectively closer to the origin (Figure 4). Intuitively, one might expect the allylic C–H sites (purple) or the tertiary C–H bond alpha to oxygen (orange) in cholesteryl pelargonate to also be inherently reactive, but our visual analysis of local C–H bond properties clearly shows these sites to be less prone to functionalization. We propose that overlaying the computed site-by-site properties for a new target substrate containing many C–H bonds with bias ovals for reported catalysts can enable facile visual prediction of which site is likely to be functionalized. More broadly, we view the intersection of this substrate profile analysis and our previous catalyst descriptions using SMART parameters[22] as key steppingstones to enabling large-scale multi-site classification virtual screening. In this way, a matrix of probabilities that any given catalyst will functionalize each possible C–H site within any given substrate could be rapidly attainable for *a priori* guidance of chemical synthesis across extensive substrate/catalyst libraries.

Conclusion

In summary, we have demonstrated how carbene-induced C–H functionalization can be incorporated into the classic disconnection logic used in organic synthesis. By using catalyst-controlled reactions of alkyl silyl ethers, we have developed practical and mild methods for intermolecular asymmetric C(sp^3)–H bond functionalization that affords products with oxygen functionality located in 1,3-, 1,5-, or 1,6-arrangements. Intermolecular desymmetrization and kinetic resolution of alkyl silyl ethers has also been shown to be effective. Finally, we have harnessed this data set to highlight the synthetic utility of training a machine learning classification algorithm for predicting which C(sp^3)–H site within a substrate will be functionalized based on facile calculations of local, ground-state chemical properties. We hope that this study will spur interest in leveraging this chemistry for complex molecule assembly as a complementary tool to some of the classical disconnection strategies, with our intuitive visual analysis potentially aiding in the prediction of tractable synthetic targets for late-stage C–H functionalization.

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)

Computed properties and modeling input data is provided. (XLSX)

Accession Codes

CCDC 2168240 and 2168271 contain the supplementary crystallographic data for this paper. These data can be obtained free of charge via www.ccdc.cam.ac.uk/data_request/cif, or by emailing data_request@ccdc.cam.ac.uk, or by contacting The Cambridge Crystallographic DataCentre, 12 Union Road, Cambridge CB2 1EZ, UK; fax: +44 902 1223 336033.

AUTHOR INFORMATION

Corresponding Authors

Huw M. L. Davies – Department of Chemistry, Emory University, Atlanta, Georgia 30322, United States; orcid.org/0000-0001-6254-9398; Email: hmdavie@emory.edu

Matthew S. Sigman – Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, United States; orcid.org/0000-0002-5746-8830; Email: matt.sigman@utah.edu

Authors

Yannick T. Boni – Department of Chemistry, Emory University, Atlanta, Georgia 30322, United States; orcid.org/0000-0002-5040-3997

Ryan C. Cammarota – Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, United States: orcid.org/0000-0002-3089-9534

Kuangbiao Liao – Department of Chemistry, Emory University, Atlanta, Georgia 30322, United States; orcid.org/0000-0001-9089-0569

Present Addresses

#Ryan C. Cammarota: Department of Chemistry, Pomona College, Claremont, California 91711, United States.

$Kuangbiao Liao: Guangzhou Laboratory, Guangzhou 510005, Guangdong Province, China.

Author Contributions

‡ Y.T.B. and R.C.C 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). We thank Dr. Wenbin Liu for supplying some of the chiral dirhodium catalyst used in this study. We would also like to thank Hanna Clements for introducing us to the logistic regression tools used in this study and Dr. Jordan Dotson for helpful discussion. We thank Dr. John Bacsa, Emory X-ray Crystallography Center, for the X–ray structural analysis.

REFERENCES

(1) McCowen, S. V.; Doering, N. A.; Sarpong, R., Retrosynthetic strategies and their impact on synthesis of arcutane natural products. *Chem. Sci.* **2020**, *11*, 7538-7552.

(2) (a) Yoshikawa, N.; Yamada, Y. M.; Das, J.; Sasai, H.; Shibasaki, M., Direct catalytic asymmetric aldol reaction. *J. Am. Chem. Soc.* **1999**, *121*, 4168-4178. (b) Notz, W.; Tanaka, F.; Watanabe, S.-I.; Chowdari, N. S.; Turner, J. M.; Thayumanavan, R.; Barbas, C. F., The direct organocatalytic asymmetric Mannich reaction: unmodified aldehydes as nucleophiles. *J. Org. Chem.* **2003**, *68*, 9624-9634. (c) Trost, B. M.; Brindle, C. S., The direct catalytic asymmetric aldol reaction. *Chem. Soc. Rev.* **2010**, *39*, 1600-1632.

(3) Zhang, Y.; Wang, W., Recent advances in organocatalytic asymmetric Michael reactions. *Catal. Sci. Technol.* **2012**, *2*, 42-53.

(4) (a) Kim, D. E.; Zweig, J. E.; Newhouse, T. R., Total synthesis of paspaline A and emindole PB enabled by computational augmentation of a transform-guided retrosynthetic strategy. *J. Am. Chem. Soc.* **2019,** *141*, 1479-1483. (b) Shen, Y.; Borowski, J. E.; Hardy, M. A.; Sarpong, R.; Doyle, A. G.; Cernak, T., Automation and computer-assisted planning for chemical synthesis. *Nat. Rev. Methods Primers* **2021**, *1*, 1-23.

(5) (a) Lou, J.; Wang, Q.; He, Y.; Yu, Z., A Simple Aliphatic Diamine Auxiliary for Palladium-Catalyzed Arylation of Unactivated β-C(sp^3)–H Bonds. *Adv. Synth. Catal.* **2018**, *360*, 4571-4584. (b) Shi, H.; Lu, Y.; Weng, J.; Bay, K. L.; Chen, X.; Tanaka, K.; Verma, P.; Houk, K. N.; Yu, J.-Q., Differentiation and functionalization of remote C–H bonds in adjacent positions. *Nat. Chem.* **2020**, *12*, 399-404. (c) Xia, G.; Zhuang, Z.; Liu, L. Y.; Schreiber, S. L.; Melillo, B.; Yu, J. Q., Ligand-Enabled β-Methylene C(sp^3)–H Arylation of Masked Aliphatic Alcohols. *Angew. Chem.* **2020**, *132*, 7857-7861. (d) Sinha, S. K.; Guin, S.; Maiti, S.; Biswas, J. P.; Porey, S.; Maiti, D., Toolbox for Distal C–H Bond Functionalizations in Organic Molecules. *Chem. Rev.* **2022**, *122,* 5682-5841.

(6) (a) Ren, Z.; Mo, F.; Dong, G., Catalytic functionalization of unactivated sp^3 C–H bonds via exo-directing groups: synthesis of chemically differentiated 1, 2-diols. *J. Am. Chem. Soc.* **2012**, *134*, 16991-16994. (b) Xu, Y.; Yan, G.; Ren, Z.; Dong, G., Diverse sp^3 C– H functionalization through alcohol β-sulfonyloxylation. *Nat. Chem.* **2015**, *7*, 829-834. (c) Bunescu, A.; Butcher, T. W.; Hartwig, J. F., Traceless silylation of β-C(sp^3)–H bonds of alcohols via perfluorinated acetals. *J. Am. Chem. Soc.* **2018**, *140*, 1502-1507. (d) Tanaka, K.; Ewing, W. R.; Yu, J.-Q., Hemilabile benzyl ether enables γ-C(sp^3)–H carbonylation and olefination of alcohols. *J. Am. Chem. Soc.* **2019**, *141*, 15494-15497.

(7) (a) Gutekunst, W. R.; Baran, P. S., C–H functionalization logic in total synthesis. *Chem. Soc. Rev.* **2011,** *40*, 1976-1991. (b) Chen, D. Y. K.; Youn, S. W., C−H Activation: A complementary tool in the total synthesis of complex natural products. *Chem. Eur. J.* **2012**, *18*, 9452-9474. (c) Cernak, T.; Dykstra, K. D.; Tyagarajan, S.; Vachal, P.; Krska, S. W., The medicinal chemist's toolbox for late stage functionalization of drug-like molecules. *Chem. Soc. Rev.* **2016**, *45*, 546-576. (d) Davies, H. M. L.; Morton, D., Recent advances in C–H functionalization. *J. Org. Chem.* **2016,** *81*, 343-350. (e) Hartwig, J. F., Evolution of C–H Bond Functionalization from Methane to Methodology. *J. Am. Chem. Soc.* **2016**, *138*, 2-24.

(8) MacMillan, D. W., The advent and development of organocatalysis. *Nature* **2008**, *455*, 304-308.

(9) (a) Macdonald, D. S.; Garrabou, X.; Klaus, C.; Verez, R.; Mori, T.; Hilvert, D., Engineered artificial carboligases facilitate regioselective preparation of enantioenriched aldol adducts. *J. Am. Chem. Soc.* **2020**, *142*, 10250-10254. (b) 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., Engineering P450 TamI as an iterative biocatalyst for selective late-stage C–H functionalization and epoxidation of tirandamycin antibiotics. *ACS Catal.* **2021**, *11*, 8304-8316.

(10) Davies, H. M. L., Finding opportunities from surprises and failures. Development of rhodium-stabilized donor/acceptor carbenes and their application to catalyst-controlled C–H functionalization. *J. Org. Chem.* **2019**, *84*, 12722-12745.

(11) (a) 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*, 230-234. (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*, 609-613. (c) Liao, K.; Yang, Y.-F.; Li, Y.; Sanders, J. N.; Houk, K.; 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*, 1048-1055.

(12) (a) Davies, H. M. L.; Antoulinakis, E. G.; Hansen, T., Catalytic asymmetric synthesis of syn-aldol products from intermolecular C–H insertions between allyl silyl ethers and methyl aryldiazoacetates. *Org. Lett.* **1999**, *1*, 383-386. (b) Davies, H. M. L.; Antoulinakis, E. G., Asymmetric catalytic C−H activation applied to the synthesis of syn-aldol products. *Org. Lett.* **2000**, *2*, 4153-4156. (c) Davies, H. M. L.; Beckwith, R. E.; Antoulinakis, E. G.; Jin, Q., New strategic reactions for organic synthesis: Catalytic asymmetric C−H activation α to oxygen as a surrogate to the aldol reaction. *J. Org. Chem.* **2003**, *68*, 6126-6132.

(13) (a) Davies, H. M. L.; Hansen, T.; Hopper, D. W.; Panaro, S. A., Highly regio-, diastereo-, and enantioselective C−H insertions of methyl aryldiazoacetates into cyclic *N*-Boc-protected amines. Asymmetric synthesis of novel C_2-symmetric amines and threo-methylphenidate. *J. Am. Chem. Soc.*

1999, *121*, 6509-6510. (b) Davies, H. M. L.; Venkataramani, C., Catalytic enantioselective synthesis of β2-amino acids. *Angew. Chem. Int. Ed.* **2002**, *41*, 2197-2199. (c) Davies, H. M. L; Venkataramani, C.; Hansen, T.; Hopper, D. W., New strategic reactions for organic synthesis: catalytic asymmetric C–H activation α to nitrogen as a surrogate for the Mannich reaction. *J. Am. Chem. Soc.* **2003**, *125*, 6462-6468. (d) Davies, H. M. L.; Jin, Q., Double C–H Activation Strategy for the Asymmetric Synthesis of C_2-Symmetric Anilines. *Org. Lett.* **2004**, *6*, 1769-1772. (e) Davies, H. M. L.; Ni, A., Enantioselective synthesis of β-amino esters and its application to the synthesis of the enantiomers of the antidepressant Venlafaxine. *Chem. Commun.* **2006**, 3110-3112.

(14) (a) Davies, H. M. L.; Ren, P.; Jin, Q., Catalytic asymmetric allylic C–H activation as a surrogate of the asymmetric Claisen rearrangement. *Org. Lett.* **2001**, *3*, 3587-3590. (b) Davies, H. M. L.; Yang, J.; Nikolai, J., Asymmetric C–H insertion of Rh (II) stabilized carbenoids into acetals: A C–H activation protocol as a Claisen condensation equivalent. *J. Organomet. Chem.* **2005**, *690*, 6111-6124.

(15) Davies, H. M. L.; Ren, P., Catalytic asymmetric C–H activation of silyl enol ethers as an equivalent of an asymmetric Michael reaction. *J. Am. Chem. Soc.* **2001**, *123*, 2070-2071.

(16) Lian, Y.; Davies, H. M. L., Combined C–H functionalization/Cope rearrangement with vinyl ethers as a surrogate for the vinylogous Mukaiyama aldol reaction. *J. Am. Chem. Soc.* **2011**, *133*, 11940-11943.

(17) (a) Smith, A. G.; Davies, H. M. L., Rhodium-catalyzed enantioselective vinylogous addition of enol ethers to vinyldiazoacetates. *J. Am. Chem. Soc.* **2012**, *134*, 18241-18244. (b) Vaitla, J.; Boni, Y. T.; Davies, H. M. L., Distal allylic/benzylic C–H functionalization of silyl ethers using donor/acceptor rhodium(II) carbenes. *Angew. Chem. Int. Ed.* **2020**, *59*, 7397-7402.

(18) (a) Li, X.; Zhang, S. Q.; Xu, L. C.; Hong, X., Predicting regioselectivity in radical C–H functionalization of heterocycles through machine learning. *Angew. Chem.* **2020,** *132* (32), 13355-13361. (b) 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*, 896-902. (c) Gallegos, L. C.; Luchini, G.; St. John, P. C.; Kim, S.; Paton, R. S., Importance of engineered and learned molecular representations in predicting organic reactivity, selectivity, and chemical properties. *Acc. Chem. Res.* **2021**, *54*, 827-836. (d) Guan, Y.; Coley, C. W.; Wu, H.; Ranasinghe, D.; Heid, E.; Struble, T. J.; Pattanaik, L.; Green, W. H.; Jensen, K. F., Regioselectivity prediction with a machine-learned reaction representation and on-the-fly quantum mechanical descriptors. *Chem. Sci.* **2021**, *12*, 2198-2208.

(19) Guptill, D. M.; Davies, H. M. L., 2, 2, 2-Trichloroethyl aryldiazoacetates as robust reagents for the enantioselective C–H functionalization of methyl ethers. *J. Am. Chem. Soc.* **2014**, *136*, 17718-17721.

(20) Davies, H. M. L.; Liao, K., Dirhodium tetracarboxylates as catalysts for selective intermolecular C–H functionalization. *Nat. Rev. Chem.* **2019**, *3*, 347-360.

(21) (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.

(22) Cammarota, R. C.; Liu, W.; Bacsa, J.; Davies, H. M. L.; Sigman, M. S., Mechanistically guided workflow for relating complex reactive site topologies to catalyst performance in C–H functionalization reactions. *J. Am. Chem. Soc.* **2022,** *144 (4)*, 1881-1898.

(23) Borissov, A.; Davies, T.; Ellis, S.; Fleming, T.; Richardson, M.; Dixon, D., Organocatalytic enantioselective desymmetrisation. *Chem. Soc. Rev.* **2016**, *45*, 5474-5540.

(24) (a) Lambert, J. B., The interaction of silicon with positively charged carbon. *Tetrahedron* **1990**, *46*, 2677-2689. (b) Lambert, J. B.; Zhao, Y.; Emblidge, R. W.; Salvador, L. A.; Liu, X.; So, J.-H.; Chelius, E. C., The β effect of silicon and related manifestations of σ conjugation. *Acc. Chem. Res.* **1999**, *32*, 183-190.

(25) 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*, 3112-3116.

(26) (a) Kirasich, K.; Smith, T.; Sadler, B., Random forest vs logistic regression: binary classification for heterogeneous datasets. *SMU Data Science Review* **2018**, *1*, 9. (b) Levy, J. J.; O'Malley, A. J., Don't dismiss logistic regression: the case for sensible extraction of interactions in the era of machine learning. *BMC Med. Res. Methodol.* **2020**, *20*, 1-15. (c) Sarker, I. H., Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science* **2021**, *2*, 1-21.

(27) (a) 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*, 650-654. (b) McLean, A.; Chandler, G., Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z= 11–18. *J. Chem. Phys.* **1980**, *72*, 5639-5648. (c) Curtiss, L. A.; McGrath, M. P.; Blaudeau, J. P.; Davis, N. E.; Binning Jr, R. C.; Radom, L., Extension of Gaussian-2 theory to molecules containing third-row atoms Ga–Kr. *J. Chem. Phys.* **1995**, *103*, 6104-6113. (d) Glukhovtsev, M. N.; Pross, A.; McGrath, M. P.; Radom, L., Extension of Gaussian-2 (G2) theory to bromine-and iodine-containing molecules: Use of effective core potentials. *J. Chem. Phys.* **1995**, *103*, 1878-1885. (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*, 215-241. (f) 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*, 6378-6396. (g) Grimme, S.; Ehrlich, S.; Goerigk, L., Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456-1465. (h) M. J. Frisch, G. W. T., 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. (i) Beck, A. D., Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648-5652. (j) Becke, A. D.; Johnson, E. R., A density-functional model of the dispersion interaction. *J. Chem. Phys.* **2005**, *123*, 154101 (1-9). (k) 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*, 154104 (1-19). (l) Johnson, E. R.; Becke, A. D., A post-Hartree–Fock model of intermolecular interactions. *J. Chem. Phys.* **2005**, *123*, 024101 (1-7). (m) Johnson, E. R.; Becke, A. D., A post-Hartree-Fock model of intermolecular interactions: Inclusion of higher-order corrections. *J. Chem. Phys.* **2006**, *124*, 174104 (1-9). (n) 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*, 3297-3305. (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*, 785-789.

(28) (a) *MacroModel, version 11.8*, Schrödinger, LLC: New York, NY 2017. (b) 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*, 1063-1065. (c) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L., OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281-296.

(29) 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*, 12247-12255.

(30) (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*, 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*, 2286-2293.

(31) 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*, 15539-15547.

(32) Newman-Stonebraker, S. H.; Smith, S. R.; Borowski, J. E.; Peters, E.; Gensch, T.; Johnson, H. C.; Sigman, M. S.; Doyle, A. G., Univariate classification of phosphine ligation state and reactivity in cross-coupling catalysis. *Science* **2021**, *374*, 301-308.

(33) (a) Qu, X.; Latino, D. A.; Aires-de-Sousa, J., A big data approach to the ultra-fast prediction of DFT-calculated bond energies. *J. Cheminformatics* **2013**, *5*, 1-13. (b) 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*, 1296-1310. (c) John, P. C. S.; Guan, Y.; Kim, Y.; Kim, S.; Paton, R. S., Prediction of organic homolytic bond dissociation enthalpies at near chemical accuracy with sub-second computational cost. *Nat. Commun.* **2020**, *11*, 1-12. (d) Xue, X.-S.; Ji, P.; Zhou, B.; Cheng, J.-P., The essential role of bond energetics in C–H activation/functionalization. *Chem. Rev.* **2017**, *117* , 8622-8648.

(34) Glendening, E. D.; Badenhoop, J. K.; Reed, A. E.; Carpenter, J. E.; Bohmann, J. A.; Morales, C. M.; Karafiloglou, P.; Landis, C. R.; Weinhold, F., *NBO 7.0*, Theoretical Chemistry Institute, University of Wisconsin, Madison, 2018.

(35) Rogge, T.; Kaplaneris, N.; Chatani, N.; Kim, J.; Chang, S.; Punji, B.; Schafer, L. L.; Musaev, D. G.; Wencel-Delord, J.; Roberts, C. A.; Sarpong, R.; Wilson, Z. E.; Brimble, M. A.; Johansson, M. J.; Ackermann, L., C–H activation. *Nat. Rev. Methods Primers* **2021**, *1*, 43.

(36) Griffin, J. D.; Vogt, D. B.; Du Bois, J.; Sigman, M. S., Mechanistic guidance leads to enhanced site-selectivity in C–H oxidation reactions catalyzed by ruthenium bis(bipyridine) complexes. *ACS Catal.* **2021**, *11*, 10479-10486.

(37) (a) Houle, F.; Beauchamp, J., Photoelectron spectroscopy of methyl, ethyl, isopropyl, and tert-butyl radicals. Implications for the thermochemistry and structures of the radicals and their corresponding carbonium ions. *J. Am. Chem. Soc.* **1979**, *101*, 4067-4074. (b) Aue, D. H., Carbocations. Wiley Interdisciplinary Reviews: Computational Molecular Science **2011**, 1, 487-508. (c) Stoyanov, E. S.; Nizovtsev, A. S., Stabilization of carbocations CH_3^+, $C_2H_5^+$, $iC_3H_7^+$, *tert*-Bu^+, and cyclo-pentyl$^+$ in solid phases: experimental data versus calculations. *Phys. Chem. Chem. Phys.* **2017**, *19*, 7270-7279.

(38) We do not view this observation as an inherent limitation of $\Delta E_{\sigma/\sigma^*}$ (C–H) derived from NBO analysis, but rather as a consequence of attempting to describe transition states where positive charge accumulates with the molecular properties of neutrally-charged ground states. However, it is impractical to compute the structures and properties for all possible carbocations that can result from heterolytic cleavage of each C–H bond in each possible conformation. Furthermore, this approach was recently pursued by Maseras, Pérez, and co-workers for predicting the relative reactivity of alkane C–H bonds, and instead the authors ultimately opted for a model based on simple topological parameters when more chemically interpretable descriptors from this type of analysis using quantum mechanical calculations afforded only mild correlations with experiment (see ref. 25).

(39) (a) Ditchfield, R., Self-consistent perturbation theory of diamagnetism: I. A gauge-invariant LCAO method for NMR chemical shifts. *Mol. Phys.* **1974**, *27*, 789-807. (b) Wolinski, K.; Hinton, J. F.; Pulay, P., Efficient implementation of the gauge-independent atomic orbital method for NMR chemical shift calculations. *J. Am. Chem. Soc.* **1990**, *112*, 8251-8260.

(40) (a) Keeler, J., *Understanding NMR Spectroscopy*. John Wiley & Sons: Hoboken, NJ, **2011,** p. 5-22. (b) Kupka, T., Theory and computation of nuclear shielding. In *Nuclear Magnetic Resonance*. Vol. 46, **2020,** pp. 1-33. DOI: 10.1039/9781788010665-00001.

(41) (a) Fraser, R. R.; Gurudata; Renaud, R. N.; Reyes-Zamora, C.; Swingle, R. B., Effect of substituents on the chemical shift of benzylic protons. *Can. J. Chem.* **1969**, *47*, 2767-2773. (b) Abraham, R. J.; Canton, M.; Griffiths, L., Proton chemical shifts in NMR: Part 17. Chemical shifts in alkenes and anisotropic and steric effects of the double bond. *Magn. Reson. Chem.* **2001**, *39*, 421-431. (c) Baranac-Stojanović, M., New insight into the anisotropic effects in solution-state NMR spectroscopy. *RSC Advances* **2014**, *4*, 308-321.

(42) (a) Hammett, L. P., Some relations between reaction rates and equilibrium constants. *Chem. Rev.* **1935**, *17*, 125-136. (b) Hammett, L. P., Linear free energy relationships in rate and equilibrium phenomena. *Trans. Faraday Soc.* **1938**, *34*, 156-165. (c) Jaffe, H., A reexamination of the Hammett equation. *Chem. Rev.* **1953**, *53*, 191-261.

(43) (a) Hammett, L. P., The effect of structure upon the reactions of organic compounds. Benzene derivatives. *J. Am. Chem. Soc.* **1937,** *59* (1), 96-103. (b) Hansch, C.; Leo, A.; Taft, R., A survey of Hammett substituent constants and resonance and field parameters. *Chem. Rev.* **1991**, *91*, 165-195.

(44) (a) Long, J. S. Regression models for categorical and limited dependent variables. In *Advanced Quantitative Techniques in the Social Sciences*. Vol. 7. SAGE Publishing: London, UK, 1997, p. 34-83. (b) Long, J. S.; Freese, J., Regression models for categorical dependent variables using Stata. Third Ed. Stata Press: College Station, TX, **2014,** p. 1-328.

(45) van Smeden, M.; de Groot, J. A.; Moons, K. G.; Collins, G. S.; Altman, D. G.; Eijkemans, M. J.; Reitsma, J. B., No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *BMC Med. Res. Methodol.* **2016**, *16*, 1-12.

(46) (a) Lloyd, S., Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **1982,** *28*, 129-137. (b) Faber, V., Clustering and the continuous k-means algorithm. *Los Alamos Science* **1994,** *22*, 138-144. (c) 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.* Loizides, F. and Scmidt, B. (Eds.). In *Positioning and Power in Academic Publishing: Players, Agents and Agendas.* IOS Press: **2016,** p. 87-90 (DOI:10.3233/978-1-61499-649-1-87). (d) *Anaconda Software Distribution*, Anaconda Inc.: 2020. <https://docs.anaconda.com/>. Accessed 2021-07-01. (e) Moriwaki, H.; Tian, Y.-S.; Kawashita, N.; Takagi, T., Mordred: a molecular descriptor calculator. *J. Cheminformatics* **2018**, *10*, 1-14.

(47) (a) Saito, T.; Rehmsmeier, M., The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one* **2015**, *10*, e0118432. (b) Powers, D. M., Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061* **2020**. <https://arxiv.org/abs/2010.16061>. Accessed 2021-10-01.

(48) (a) McFadden, D., Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), Frontiers in Econometrics (pp. 105–142). Academic Press. **1974**. (b) Agresti, A., Applying R 2–type measures to ordered categorical data. *Technometrics* **1986**, *28*, 133-138. (c) Ash, A.; Shwartz, M., R^2: a useful measure of model performance when predicting a dichotomous outcome. *Stat. Med.* **1999**, *18*, 375-384.

(49) (a) McFadden, D., Chapter 15. Quantitative Methods for Analyzing Travel Behaviour on Individuals: Some Recent Developments. In *Behavioural Travel Modeling*, Hensher, D. A.; Stopher, P. R., Eds. Croom Helm: 1979; p 306. (b) Cox, D. R.; Wermuth, N., A comment on the coefficient of determination for binary responses. *Am. Stat.* **1992,** *46* (1), 1-4.

(50) Drew, K. L.; Reynisson, J., The impact of carbon–hydrogen bond dissociation energies on the prediction of the cytochrome P450 mediated major metabolic site of drug-like compounds. *Eur. J. Med. Chem.* **2012**, *56*, 48-55.

(51) (a) Tolman, C. A., Electron donor-acceptor properties of phosphorus ligands. Substituent additivity. *J. Am. Chem. Soc.* **1970**, *92*, 2953-2956. (b) Tolman, C. A., Phosphorus ligand exchange equilibriums on zerovalent nickel. Dominant role for steric effects. *J. Am. Chem. Soc.* **1970**, *92*, 2956-2965. (c) Tolman, C. A., Steric effects of phosphorus ligands in organometallic chemistry and homogeneous catalysis. *Chem. Rev.* **1977**, *77*, 313-348.

(52) Refer to the SI for a discussion of the intersection of chemically meaningful modeling with the norms for logistic regression. Also see: Midi, H.; Sarkar, S. K.; Rana, S., Collinearity diagnostics of binary logistic regression model. *J. Interdiscip. Math.* **2010**, *13*, 253-267.

(53) (a) 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*, 1857-1869. (b) Newhouse, T.; Baran, P. S., If C–H bonds could talk: Selective C–H bond oxidation. *Angew. Chem. Int. Ed.* **2011**, *50*, 3362-3374. (c) 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.

For Table of Contents Only: