

CrossMark
click for updatesCite this: *Chem. Sci.*, 2015, 6, 3057

Using IR vibrations to quantitatively describe and predict site-selectivity in multivariate Rh-catalyzed C–H functionalization†

Elizabeth N. Bess,‡^a David M. Guptill,‡^b Huw M. L. Davies*^b and Matthew S. Sigman*^a

Achieving selective C–H functionalization is a significant challenge that requires discrimination between many similar C–H bonds. Yet, reaction systems employing $\text{Rh}_2(\text{DOSP})_4$ and $\text{Rh}_2(\text{BPCP})_4$ were recently demonstrated to afford high levels of selectivity in the C–H insertion of carbenes into toluene-derived substrates. Herein, we explore the origin of this selectivity through a systematic analysis of substrate and reagent features that alter levels of selectivity from 20 : 1 to 1 : 610 for secondary (or tertiary)-to-primary benzylic C–H functionalization of toluene derivatives. Describing this variation using infrared vibrations and point charges, we have developed a mathematical model from which are identified features of the systems that determine levels of site-selectivity and are applied as predictive factors to describe the selectivity behavior of new substrate/reagent combinations.

Received 29th January 2015
Accepted 18th March 2015

DOI: 10.1039/c5sc00357a

www.rsc.org/chemicalscience

Introduction

The abundance of C–H bonds in any given molecule presents a major challenge to the development of site-selective C–H functionalization methods.¹ Out of this challenge arises an objective to identify and understand the chemical precepts and interactions that govern site-selectivity in such reactions.² With increasingly precise knowledge of these outcome-defining molecular features, a methodology can be better refined and optimized, with the ultimate goal of developing predictive models and tailoring systems to yield prescribed reaction outcomes.

An impressive example of site-selective C–H functionalization is the C–H insertion reaction of donor/acceptor rhodium-carbenes.^{3,4} This process is proposed to proceed through a hydride transfer-like event from the substrate to the carbene, with subsequent C–C bond formation.⁵ It is generally observed in these systems that more substituted (or activated) C–H bonds are electronically favored and primary C–H bonds are sterically preferred. With the most established catalyst, $\text{Rh}_2(\text{S-DOSP})_4$, these effects are balanced so as to favor functionalization of secondary C–H bonds (Fig. 1).³ This tendency can be overcome to favor functionalization of primary C–H bonds when using a

more sterically demanding catalyst, $\text{Rh}_2(\text{R-BPCP})_4$ (Fig. 1).^{6,7} Nevertheless, the effect of the ester substituent⁷ and functionality on the donor group⁸ remain relatively unexplored as elements of control over site-selectivity. Understanding the origin of selectivity is thus a major goal in the design and development of new reagents and catalysts.^{6,7}

While there are many diverse means of exploring the molecular features at the origin of selectivity,⁹ the Sigman group has focused on a method that derives from linear free-energy relationship analysis. In short, mathematical relationships are developed that equate the differential transition state free energy between pathways affording isomeric products ($\Delta\Delta G^\ddagger$) to numeric depictions of molecular features.^{10,11} Of the various parameters available, computationally measured infrared (IR) bond vibrations have recently emerged as particularly apt descriptors of the molecular characteristics that impact selectivity.¹⁰ The intrinsic ability of IR vibrations to accurately

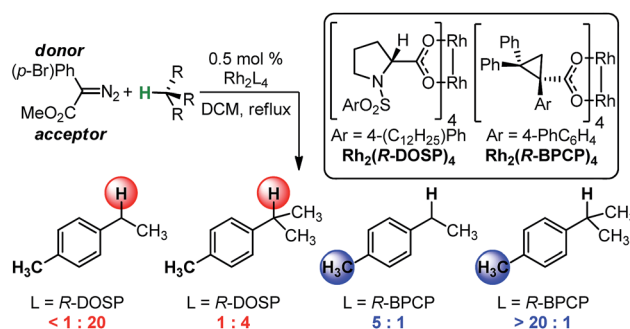


Fig. 1 Davies and colleagues' Rh-catalyzed C–H functionalization reaction, demonstrating site-selectivity sensitivity. Preferred C–H bond for functionalization under given conditions is highlighted.

^aDepartment of Chemistry, University of Utah, 315 South 1400 East, Salt Lake City, UT 84112, USA. E-mail: sigman@chem.utah.edu

^bDepartment of Chemistry, Emory University, 1515 Dickey Drive, Atlanta, GA 30322, USA. E-mail: hmdavie@emory.edu

† Electronic supplementary information (ESI) available: Experimental procedures, tabulated descriptors, and model development MATLAB commands. See DOI: 10.1039/c5sc00357a

‡ These authors contributed equally.

represent a molecule's multifaceted, fundamental structure has enabled the mathematical description of complex systems.¹² However, the density of information manifested in an IR vibration can also make this descriptor difficult to interpret, clouding the origin of a vibration's importance for mathematically describing trends in selectivity.

Herein, we investigate a method of decoding the controlling elements of rhodium-catalyzed C–H insertion of donor/acceptor carbenes *via* free-energy mathematical models (Fig. 1).⁶ Using IR vibrations, we have developed quantitative relationships that describe a substantial range (20 : 1 to 1 : 610) in ratios of secondary (or tertiary)-to-primary benzylic C–H functionalization of toluene derivatives. These relationships begin to inform the design of the optimum reagent/catalyst combination for reaction at a specific C–H bond and for the prediction of site-selective outcomes.

Methodology

Experimental design

Evaluation of the $\text{Rh}_2(\text{R-BPCP})_4$ - and $\text{Rh}_2(\text{S-DOSP})_4$ -catalyzed C–H insertion reactions of toluene derivatives commenced with construction of a design of experiments (DoE)-founded matrix to systematically evaluate diazo ester features that impact site selection (Fig. 2).^{13,14} This matrix was built upon the hypothesis that electron-donating and -withdrawing arene substituent effects (R) will influence reaction outcomes in tandem with steric and/or electronic contributions from the ester moiety (R'). Thus, the arene of **1** was assessed according to a Hammett series of electronically varied substituents, and the ester was evaluated *via* a set of sterically and electronically diverse groups. Finally, the substrate was considered to be an additional dimension. As such, the catalysts' abilities to discriminate

between primary and secondary or tertiary C–H bonds, adjacent to different steric environments, could be gauged by installation of ethyl, isopropyl, and isobutyl substituents at the *para*-position in toluene substrates.

To systematically and efficiently assess combinations of these three variables, a face-centered cubic design matrix was implemented to define experimental evaluation of the reaction (Fig. 2b).¹⁴ The resulting suite of arene, ester, and substrate combinations was assessed in the $\text{Rh}_2(\text{S-DOSP})_4$ and $\text{Rh}_2(\text{R-BPCP})_4$ catalytic systems, affording an array of secondary (or tertiary)-to-primary benzylic C–H functionalization ratios that range from 20 : 1 to 1 : 610, respectively (Tables 1 and 2). In the $\text{Rh}_2(\text{S-DOSP})_4$ -catalyzed reactions (Table 1), insertion into a secondary C–H bond preferentially occurs, as previously observed, but there is a considerable difference in the selectivity depending on the nature of the aryl substituent and the ester group. In the $\text{Rh}_2(\text{R-BPCP})_4$ -catalyzed reactions (Table 2), insertion into a primary C–H bond is strongly favored, although a range is also measured.

Reactions performed on 4-isobutyltoluene with $\text{Rh}_2(\text{R-BPCP})_4$ lead to the detection of only primary C–H insertion products. Changing to $\text{Rh}_2(\text{S-DOSP})_4$, insertion was favored at the primary site, but tertiary (rather than secondary, benzylic) insertion products were also observed. These results preclude the use of 4-isobutyltoluene in the remainder of these studies.

Correlation of IR parameters to classical measurements

To initiate mathematical modelling of the selectivity ratios as a function of molecular features of the diazo ester and toluene substrate, potentially relevant mathematical descriptors were gathered. While σ may describe the appropriate changes to the arene ring (*vide infra*), it was anticipated that simultaneous variation of both the ester and arene would influence the reactivity and selectivity of the resultant carbene, which could not be described by Hammett σ or σ^+ values alone. Consequently, it was of particular interest to identify an analogous parameter that would simultaneously embody both dimensions of modulation. Thus, the $\text{N}\equiv\text{N}$ diazo IR stretching frequency and intensity were considered, as this functional group is positioned at the fusion of the two modulated moieties and the site of the eventual reactive carbene. It was hypothesized that the IR vibrational properties of the diazo might appropriately describe the electronic and, perhaps, steric influences that coalesce at this position.

Due to the ease and fidelity with which computational IR measurements can be made and identical vibrational modes can be identified, IR frequencies and intensities were computed from energy-minimized diazo esters (M06-2X/TZVP, Fig. 3a).¹⁵ Although the diazo ester is not the active species that engages in C–H insertion, it likely shares significant structural similarities to the implicated rhodium-carbene. In analogy to the classical application of Hammett σ -parameters, we hypothesized that the rhodium-carbene's *relative* steric and electronic features would be conserved in its closely related precursor, a diazo ester. Additionally, performing computations on this simple, ground-state organic molecule is a facile exercise that obviates the

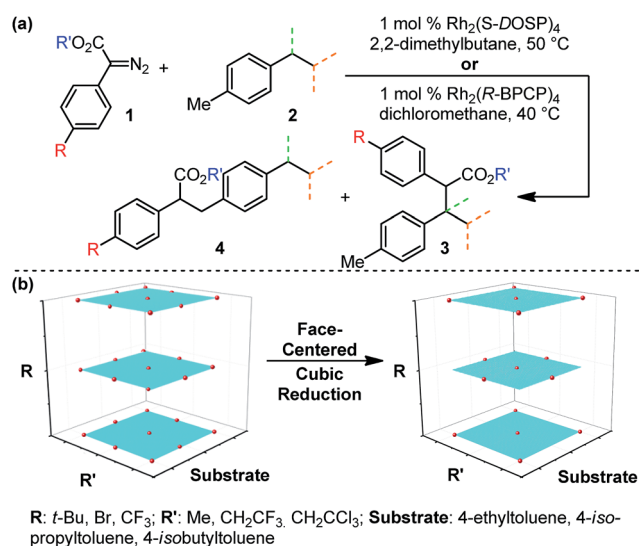


Fig. 2 (a) Identification of three reaction aspects for each Rh-catalyzed reaction that can be systematically varied to determine their influence on site selectivity. (b) Full DoE matrix and face-centered cubic reduction matrix, which defines a simplified approach to the systematic evaluation of each Rh-catalyzed system.



Table 1 Results of $\text{Rh}_2(\text{S-DOSP})_4$ -catalyzed carbene insertion reaction used to develop the descriptive model in Fig. 5a

Entry ^a	R	R'	Toluene Substrate	Meas. 3 : 4	Meas. $\Delta\Delta G^\ddagger$ (kcal mol ⁻¹)	Pred. $\Delta\Delta G^\ddagger$ (kcal mol ⁻¹)
1	<i>t</i> -Bu	Me	4-Ethyl	20.0 : 1.0	1.92	1.90
2	CF ₃	Me	4-Ethyl	10.0 : 1.0	1.48	1.64
3	Br	Me	4-Isopropyl	1.9 : 1.0	0.41	0.68
4	Br	CH ₂ CF ₃	4-Ethyl	11.0 : 1.0	1.54	1.50
5	<i>t</i> -Bu	CH ₂ CF ₃	4-Isopropyl	4.5 : 1.0	0.97	0.34
6	Br	CH ₂ CF ₃	4-Isopropyl	1.8 : 1.0	0.38	0.31
7	CF ₃	CH ₂ CF ₃	4-Isopropyl	1.4 : 1.0	0.22	0.10
8	<i>t</i> -Bu	CH ₂ CCl ₃	4-Ethyl	9.0 : 1.0	1.41	1.28
9	CF ₃	CH ₂ CCl ₃	4-Ethyl	4.8 : 1.0	1.01	1.05
10	Br	CH ₂ CCl ₃	4-Isopropyl	1.0 : 2.1	-0.48	0.07
11	Cl	Me	4-Ethyl	13.0 : 1.0	1.65	1.80
12	Br	Et	4-Ethyl	12.3 : 1.0	1.61	1.80
13	Br	CH ₂ CBr ₃	4-Ethyl	5.8 : 1.0	1.13	1.15
14	OMe	Me	4-Isopropyl	4.7 : 1.0	0.99	0.73
15	OMe	CH ₂ CCl ₃	4-Isopropyl	1.3 : 1.0	0.17	-0.03
16	Br	CH ₂ CCl ₃	4-Ethyl	6.6 : 1.0	1.21	1.26

^a Entries 1–10, training set; entries 11–16, external validation set.

greater computational load required for transition metal computation, especially when there is only one characterized example for this complex.¹⁶

As the goal of this study was not only to develop comprehensive models for prediction purposes but also to better understand the origin of site-selectivity, it was of interest to investigate what the $\text{N}\equiv\text{N}$ diazo IR stretching frequency and intensity may represent in terms of classical physical organic parameters. Therefore, a MATLAB stepwise linear regression algorithm was used to find the optimal combinations of more simple parameters that describe the observed changes in ν_{diazo} and i_{diazo} .¹⁷ These parameters include Sterimol measures of

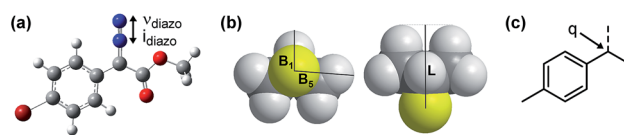


Fig. 3 (a) Representation of computationally measured diazo IR stretch in energy-minimized structures. (b) Depiction of Sterimol values used to quantitate sterics of an isopropyl substituent. B_1 is a minimum and B_5 is a maximum radial bulk; L is substituent length. (c) NBO point charge used to describe electronic variation in the toluene-derived substrates investigated.

Table 2 Results of $\text{Rh}_2(\text{R-BPCP})_4$ -catalyzed carbene insertion reaction used to develop the descriptive model in Fig. 5b

Entry ^a	R	R'	Toluene Substrate	Meas. 3 : 4	Meas. $\Delta\Delta G^\ddagger$ (kcal mol ⁻¹)	Pred. $\Delta\Delta G^\ddagger$ (kcal mol ⁻¹)
1	<i>t</i> -Bu	Me	4-Ethyl	1.0 : 3.7	-0.81	-0.68
2	CF ₃	Me	4-Ethyl	1.0 : 5.2	-1.03	-0.99
3	Br	Me	4-Isopropyl	1.0 : 95.0	-2.83	-2.80
4	Br	CH ₂ CF ₃	4-Ethyl	1.0 : 5.1	-1.01	-1.15
5	<i>t</i> -Bu	CH ₂ CF ₃	4-Isopropyl	1.0 : 101.0	-2.87	-3.21
6	Br	CH ₂ CF ₃	4-Isopropyl	1.0 : 182.0	-3.24	-3.24
7	CF ₃	CH ₂ CF ₃	4-Isopropyl	1.0 : 213.0	-3.33	-3.49
8	<i>t</i> -Bu	CH ₂ CCl ₃	4-Ethyl	1.0 : 10.0	-1.43	-1.41
9	CF ₃	CH ₂ CCl ₃	4-Ethyl	1.0 : 14.0	-1.64	-1.69
10	Br	CH ₂ CCl ₃	4-Isopropyl	1.0 : 610.0	-3.99	-3.52
11	Cl	Me	4-Ethyl	1.0 : 3.8	-0.83	-0.80
12	Br	Et	4-Ethyl	1.0 : 4.4	-0.92	-0.79
13	Br	CH ₂ CBr ₃	4-Ethyl	1.0 : 14.0	-1.64	-1.57
14	OMe	CH ₂ CCl ₃	4-Isopropyl	1.0 : 55.0	-2.49	-3.65
15	F	CH ₂ CCl ₃	4-Isopropyl	1.0 : 437.0	-3.78	-3.76
16	Br	CH ₂ <i>t</i> -Bu	4-Ethyl	1.0 : 7.4	-1.24	-1.05
17	Br	CH ₂ CCl ₃	4-Ethyl	1.0 : 11.1	-1.50	-1.44
18	H	Me	4-Ethyl	1.0 : 3.0	-0.68	-0.99
19	Br	Me	4-Ethyl	1.0 : 4.1	-0.88	-0.71

^a Entries 1–10, training set; entries 11–19, external validation set.



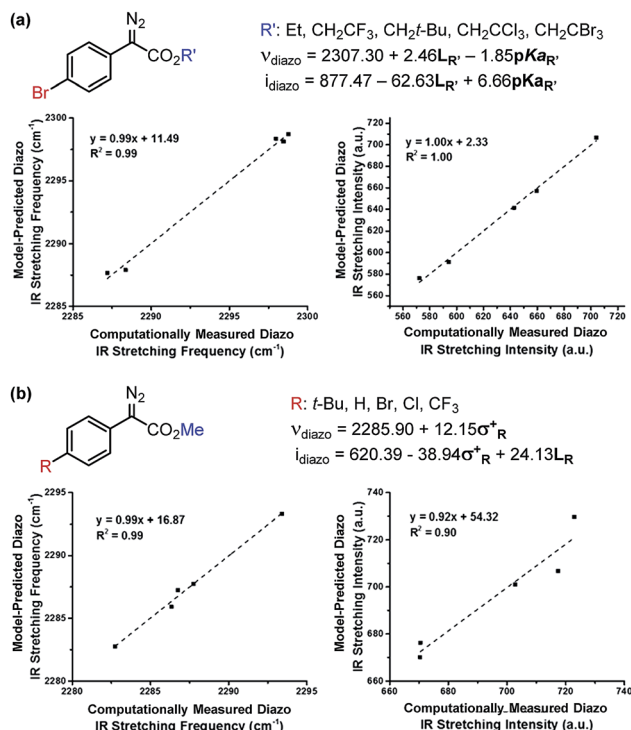


Fig. 4 Determination of which factors contribute to diazo IR stretching frequencies and intensities upon independent variation of the (a) ester and (b) arene moieties.

substituents' steric effects (B_1 , minimum radial bulk; B_5 , maximum radial bulk; L , substituent length—Fig. 3b),¹⁸ molecular weight of R' , $\text{p}K_a$ of the corresponding R' alcohol ($R'\text{-OH}$), and arene σ and σ^+ .

The results of these analyses are given in Fig. 4a. The effect of the ester substituent R' on both ν_{diazo} and i_{diazo} is best related by the Sterimol length ($L_{R'}$) and $R'\text{-OH}$ $\text{p}K_a$, which combine to effectively describe the observed trends in both IR measurements, although with different weights of importance. Of specific interest, groups with enhanced length increase the frequency whereas groups that are less able to stabilize negative charge result in decreased ν_{diazo} . The opposite scenario is described by the i_{diazo} model: vibrational intensity is decreased as the length increases but raised when R' substituents bear higher $\text{p}K_a$ values. Arene influences on the diazo IR vibration were similarly assessed (Fig. 4b). A positive, nearly one-to-one correlation between ν_{diazo} and σ^+ is demonstrated, depicting a solely electronic contribution of *para*-substituted arenes on this IR frequency (Fig. 4b).¹⁹ Alternatively, an inverse correlation exists between i_{diazo} and σ^+ , which is attenuated by mass at R , as described by Sterimol L .

Model development

With a more detailed understanding of the factors that contribute to ν_{diazo} and i_{diazo} variation, linear regression was performed using the training sets presented in Tables 1 and 2 ($\text{Rh}_2(\text{S-DOSP})_4$ and $\text{Rh}_2(\text{R-BPCP})_4$ systems, respectively). From the parameters ν_{diazo} and i_{diazo} , describing the diazo ester, and

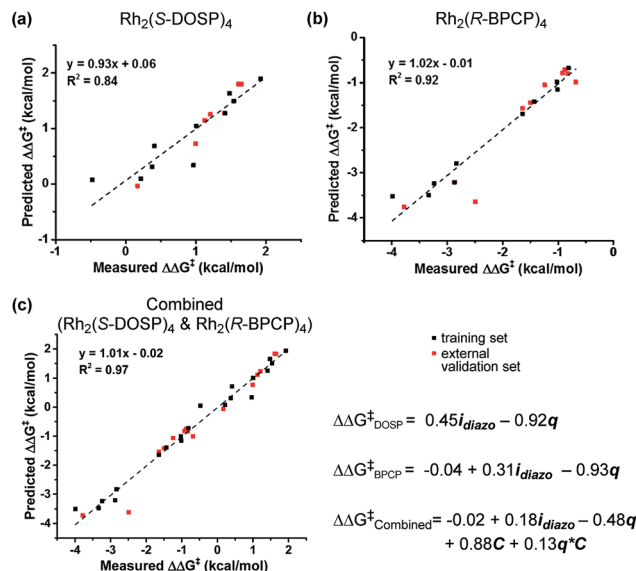


Fig. 5 Normalized descriptive mathematical models and associated model-predicted versus experimentally measured $\Delta\Delta G^\ddagger$ values for the (a) $\text{Rh}_2(\text{S-DOSP})_4$ and (b) $\text{Rh}_2(\text{R-BPCP})_4$ systems, and (c) the combination thereof.

an NBO point charge (q , Fig. 3c), describing electronic differences between 4-ethyl versus 4-isopropyl toluene substrates, stepwise linear regression was performed to identify the relationships given in Fig. 5. For both systems, i_{diazo} and the substrate point charge, q , combine to effectively describe the selectivity trends observed. The high degree of similarity between these models suggested that it might be possible to unite them into one mathematical equation. Thus, stepwise linear regression was performed on the 20 data points that result from the combination of both catalytic systems' DoE-defined training sets and the descriptive parameters i_{diazo} , q , and a binary catalyst designation, C ($\text{Rh}_2(\text{S-DOSP})_4 = 1$; $\text{Rh}_2(\text{R-BPCP})_4 = -1$). The resulting comprehensive model and external validation-assessed (Tables 1, entries 11–16; 2, entries 11–19) robustness measure is given in Fig. 5c.²⁰ This combined model effectively describes the system for both catalysts.

Analysis

The developed combined model bears information that quantitatively illustrates the origin of selectivity in $\text{Rh}(\text{II})$ -carbene C–H insertion reactions. The fact that both the $\text{Rh}_2(\text{R-BPCP})_4$ (favoring primary) and $\text{Rh}_2(\text{S-DOSP})_4$ (disfavoring primary) systems can be combined into one model demonstrates that both catalytic systems lie on a continuum with the same molecular features impacting selectivity (Fig. 6). While catalyst bias is the overriding selectivity determinant (which we surmise may be broadly attributable to steric differences), site selection is highly sensitive to the diazo reagent and substrate. Intuitively and irrespective of the catalyst, larger ester groups favor insertion at smaller primary C–H bonds. This is numerically evidenced in the interpretation of the i_{diazo} parameter. Greater substituent bulk, as defined with $L_{R'}$, diminishes the intensity of



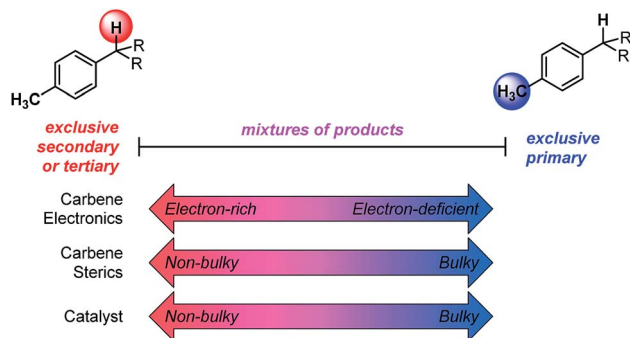


Fig. 6 Spectrum of site-selectivity for primary vs. secondary/tertiary C–H functionalization. The effects of various features, such as steric and electronic influences of the carbene and the influence of the catalyst, are shown.

the diazo stretch (Fig. 4a). As all three models in Fig. 5 indicate, lower values of i_{diazo} coincide with increased incidence of the primary insertion product. An explanation for these effects is offered: larger ester groups result in a bulkier carbene, which can better distinguish the more accessible primary C–H bonds from those at more sterically demanding secondary and tertiary positions.

Also represented in the i_{diazo} term is the electronic influence of the diazo, which originates in both the ester (R') and aryl (R) substituents. From the model in Fig. 4a, it is demonstrated that more electron-deficient esters (lower pK_a) diminish i_{diazo} and, correspondingly, erode the secondary (or tertiary)-to-primary product ratio (Fig. 5). Similarly, in the arene dimension (Fig. 4b), electron-withdrawing R substituents reduce i_{diazo} (e.g., for $R = \text{CF}_3$, $\sigma^+ = 0.61$) and yield product ratios favoring insertion at primary C–H bonds. Taken together, these electronic effects can be explained by considering that electron-deficient groups will destabilize the electrophilic carbene, rendering it more reactive. As a result of its augmented reactivity, it proceeds with C–H insertion *via* the pathway leading to the kinetic product, *i.e.*, the more easily accessible, albeit stronger, primary C–H bond.

Conclusions

In summary, the concept of having a toolbox of reagents/catalysts to control site selectivity by adjusting the carbene is exemplified in the ability to alter selectivity from a primary/tertiary selectivity ratio of 610 : 1 (Table 2, entry 10) to a primary/secondary ratio of 1 : 20 (Table 1, entry 1). The models developed demonstrate how these changes in selectivity ratios can be predictively afforded through a quantitative understanding of the systems' aspects that were investigated. Additionally, the use of correlative analysis to understand the origin of such disparate observations provides a platform for both prediction and design. Understanding such trends will be foundational in developing effective catalysts and reagents to control selective C–H functionalization, without relying on inherent substrate bias or directing groups to influence the site selectivity. Future experimental efforts are aimed at controlling the site-selectivity

of C–H functionalization of unactivated hydrocarbons, where the concepts revealed within this study will play an important role in the design of effective systems.

Acknowledgements

This work was supported by NSF under the CCI Center for Selective C–H Functionalization, CHE-1205646. The support and resources from the Center for High Performance Computing at the University of Utah is gratefully acknowledged.

Notes and references

- 1 K. M. Engle, T.-S. Mei, M. Wasa and J.-Q. Yu, *Acc. Chem. Res.*, 2011, **45**, 788–802; S. R. Neufeldt and M. S. Sanford, *Acc. Chem. Res.*, 2012, **45**, 936–946; D. A. Colby, A. S. Tsai, R. G. Bergman and J. A. Ellman, *Acc. Chem. Res.*, 2011, **45**, 814–825; H.-Y. Thu, G. S.-M. Tong, J.-S. Huang, S. L.-F. Chan, Q.-H. Deng and C.-M. Che, *Angew. Chem., Int. Ed.*, 2008, **47**, 9747–9751; D. Zalatan and J. D. Bois, in *C-H Activation*, ed. J.-Q. Yu and Z. Shi, Springer, Berlin, Heidelberg, 2010, vol. 292, ch. 19, pp. 347–378; H. Lu and X. P. Zhang, *Chem. Soc. Rev.*, 2011, **40**, 1899–1909; J. F. Hartwig, *Acc. Chem. Res.*, 2011, **45**, 864–873; M. C. White, *Science*, 2012, **335**, 807–809; P. E. Gormisky and M. C. White, *J. Am. Chem. Soc.*, 2013, **135**, 14052–14055.
- 2 M. A. Bigi, P. Liu, L. Zou, K. N. Houk and M. C. White, *Synlett*, 2012, 2768–2772.
- 3 H. M. L. Davies and D. Morton, *Chem. Soc. Rev.*, 2011, **40**, 1857–1869; H. M. L. Davies and Y. Lian, *Acc. Chem. Res.*, 2012, **45**, 923–935.
- 4 M. Doyle, M. McKervey and T. Ye, *Modern Catalytic Methods for Organic Synthesis with Diazo Compounds: From Cyclopropanes to Ylides*, Wiley, New York, 1998; D. F. Taber in *Comprehensive Organic Synthesis*, ed. G. Pattenden, Oxford, U.K., 1991, vol. 3, ch. 4.2.
- 5 J. Hansen, J. Autschbach and H. M. L. Davies, *J. Org. Chem.*, 2009, **74**, 6555–6563.
- 6 C. Qin and H. M. L. Davies, *J. Am. Chem. Soc.*, 2014, **136**, 9792–9796.
- 7 D. M. Guptill and H. M. L. Davies, *J. Am. Chem. Soc.*, 2014, **136**, 17718–17721.
- 8 D. L. Ventura, Z. Li, M. G. Coleman and H. M. L. Davies, *Tetrahedron*, 2009, **65**, 3052–3061.
- 9 E. N. Bess and M. S. Sigman, in *Asymmetric Synthesis II: More Methods and Applications*, eds. M. Christmann and S. Bräse, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2012, ch. 45; P. H.-Y. Cheong, C. Y. Legault, J. M. Um, N. Çelebi-Ölçüm and K. N. Houk, *Chem. Rev.*, 2011, **111**, 5042–5137; K. C. Harper and M. S. Sigman, *J. Org. Chem.*, 2013, **78**, 2813–2818; C. D. Johnson, *Chem. Rev.*, 1975, **75**, 755–765; A. R. Katritzky, M. Kuanar, S. Slavov, C. D. Hall, M. Karelson, I. Kahn and D. A. Dobchev, *Chem. Rev.*, 2010, **110**, 5714–5789; M. C. Kozlowski and J. C. Ianni, *J. Mol. Catal. A: Chem.*, 2010, **324**, 141–145; E. S. Lewis, *J. Phys. Org. Chem.*, 1990, **3**, 1–8; P. R. Wells, *Chem. Rev.*, 1963, **63**, 171–219.



- 10 E. N. Bess, A. J. Bischoff and M. S. Sigman, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 14698–14703; E. N. Bess, R. J. DeLuca, D. J. Tindall, M. S. Oderinde, J. L. Roizen, J. Du Bois and M. S. Sigman, *J. Am. Chem. Soc.*, 2014, **136**, 5783–5789; A. Milo, E. N. Bess and M. S. Sigman, *Nature*, 2014, **507**, 210–214.
- 11 K. C. Harper, E. N. Bess and M. S. Sigman, *Nat. Chem.*, 2012, **4**, 366–374; K. C. Harper and M. S. Sigman, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 2179–2183; K. C. Harper and M. S. Sigman, *Science*, 2011, **333**, 1875–1878; K. C. Harper, S. C. Vilardi and M. S. Sigman, *J. Am. Chem. Soc.*, 2013, **135**, 2482–2485.
- 12 J. Coates, in *Encyclopedia of Analytical Chemistry*, ed. R. A. Meyers, John Wiley & Sons Ltd, Chichester, 2000, pp. 10815–10837; R. N. Jones, W. F. Forbes and W. A. Mueller, *Can. J. Chem.*, 1957, **35**, 504–514.
- 13 R. Carlson, *Design and Optimization in Organic Synthesis*, Elsevier, Amsterdam, 1992.
- 14 S. N. Deming and S. L. Morgan, *Experimental Design: A Chemometric Approach*, Elsevier, Amsterdam, 1993.
- 15 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. L. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09, Revision C.01*, Gaussian, Inc., Wallingford, CT, 2009; Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2007, **120**, 215–241; R. Valero, J. R. B. Gomes, D. G. Truhlar and F. Illas, *J. Chem. Phys.*, 2008, **129**, 124710; A. Schäfer, H. Horn and R. Ahlrichs, *J. Chem. Phys.*, 1992, **97**, 2571; A. Schäfer, C. Huber and R. Ahlrichs, *J. Chem. Phys.*, 1994, **100**, 5829.
- 16 K. P. Kornecki, J. F. Briones, V. Boyarskikh, F. Fullilove, J. Autschbach, K. E. Schrote, K. M. Lancaster, H. M. L. Davies and J. F. Berry, *Science*, 2013, **342**, 351–354.
- 17 *MATLAB, Version 8.1.0.604 (R2013a)*, The MathWorks, Inc., Natick, Massachusetts, 2013.
- 18 A. Verloop, W. Hoogenstraaten and J. Tipker, in *Drug Design*, ed. E. J. Ariens, Academic Press, New York, 1976, vol. 7, pp. 165–207; A. Verloop and J. Tipker, *Pharmacochem. Libr.*, 1977, **2**, 63–81; A. Verloop and J. Tipker, *Pharmacochem. Libr.*, 1987, **10**, 97–102.
- 19 H. C. Brown and Y. Okamoto, *J. Am. Chem. Soc.*, 1958, **80**, 4979–4987.
- 20 S. Arlot and A. Celisse, *Stat. Surv.*, 2010, **4**, 40–79.

