LETTER

doi:10.1038/nature13019

Interrogating selectivity in catalysis using molecular vibrations

Anat Milo[1], Elizabeth N. Bess[1] & Matthew S. Sigman[1]

The delineation of molecular properties that underlie reactivity and selectivity is at the core of physical organic chemistry[1–5], and this knowledge can be used to inform the design of improved synthetic methods or identify new chemical transformations[6–9]. For this reason, the mathematical representation of properties affecting reactivity and selectivity trends, that is, molecular parameters, is paramount. Correlations produced by equating these molecular parameters with experimental outcomes are often defined as free-energy relationships and can be used to evaluate the origin of selectivity and to generate new, experimentally testable hypotheses[6,10–12]. The premise behind successful correlations of this type is that a systematically perturbed molecular property affects a transition-state interaction between the catalyst, substrate and any reaction components involved in the determination of selectivity[10,11]. Classic physical organic molecular descriptors, such as Hammett[4], Taft[3] or Charton[5] parameters, seek to independently probe isolated electronic or steric effects[3–6,13,14]. However, these parameters cannot address simultaneous, non-additive variations to more than one molecular property, which limits their utility. Here we report a parameter system based on the vibrational response of a molecule to infrared radiation that can be used to mathematically model and predict selectivity trends for reactions with interlinked steric and electronic effects at positions of interest. The disclosed parameter system is mechanistically derived and should find broad use in the study of chemical and biological systems.

A molecule's structural features are embodied in its unique vibration modes, which invoke core, inherent bond force constants and atomic masses[7,15,16]. Thus, to interrogate intricate selectivity trends, we turned to vibrational energies. Kinetic isotope effects, a common reaction probe in biology and chemistry, further highlight the usefulness of relative vibrational energies for assessing mechanistic hypotheses[7,17] (Fig. 1b). With this underpinning, we postulated that infrared vibrations could serve as mechanistically meaningful molecular descriptors in the study of catalytic reaction selectivity, allowing for correlations akin to free-energy relationships[1] (Fig. 1c; see Methods Summary for details).

Here we describe three case studies that substantiate the prospect of using molecular vibrations to predict and elucidate selectivity trends. We first consider the desymmetrization of bisphenols. In our recent reports of modelling catalytic systems, the catalyst or substrate libraries, or both, were specifically designed to avoid the complexity of integrated steric and electronic effects[9,14,18]. As an example, we studied the role of substrate steric effects in the peptide-catalysed desymmetrization of bisphenols[14,19] (Fig. 2a). A specific finding was that Verloop's Sterimol parameters B_1 (minimal radius) and L (length, Fig. 1a) could be used to successfully correlate enantioselectivity to the steric impact of the substituent **R** on the bisphenol[2]. Because this study was directed towards evaluation of steric effects, obvious electronic changes were initially avoided.

To determine whether the perturbation to enantioselectivity in this reaction is purely steric in origin, we evaluated a substrate containing $-CCl_3$, which is a $-CMe_3$ (where Me is methyl) homologue according to its Sterimol values. The $-CCl_3$-containing substrate yielded a much lower enantioselectivity than would be expected from purely steric considerations, suggesting that substituent electronic effects may affect the enantioselective outcome. To explore this outlier in the steric trend further, a model was developed on the basis of Sterimol values from an eight-membered training set of the original substrates (Fig. 2b). External validation for this model using four of the original sterically perturbed substrates and six new substrates containing concurrent steric and electronic modulation reveals a poor correlation (slope, 0.82; intercept, 0.18; $R^2 = 0.60$). The poorest performers were substrates with multifaceted structural features that cannot be accounted for by the parameter selection. The steric model's shortcomings provided an impetus to explore infrared vibrations as a reaction interrogation technique.

Thus, we chose mechanistically relevant infrared frequencies to produce molecular descriptors for the prediction of enantioselectivity[16].

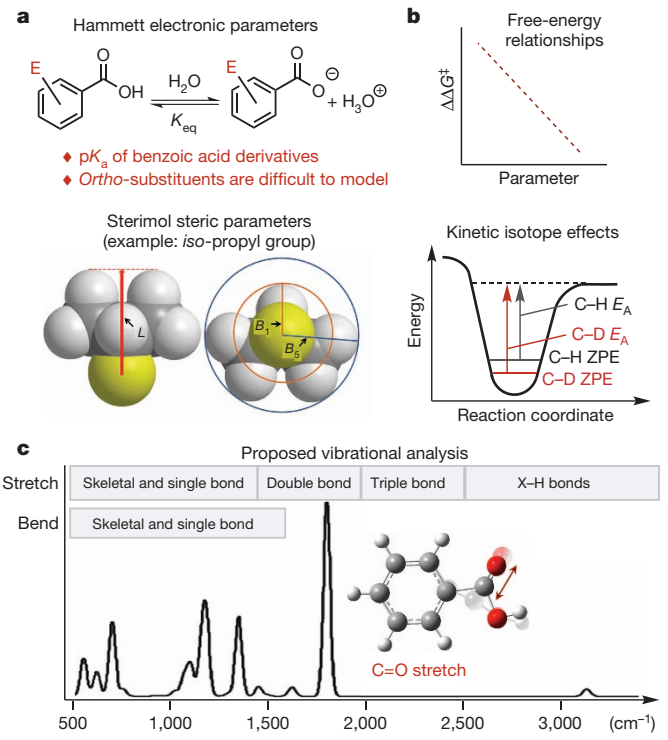
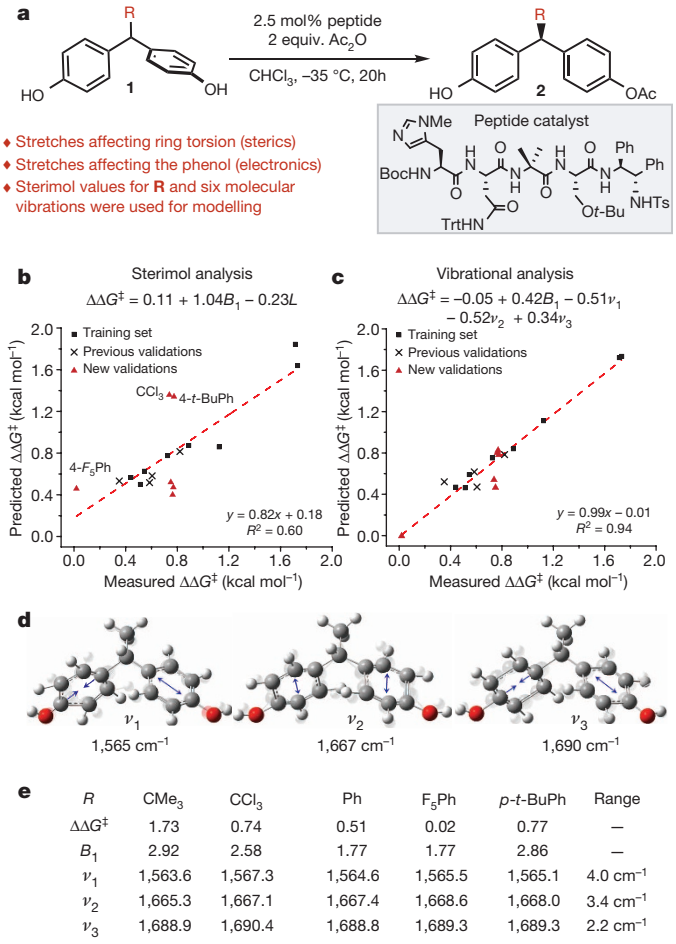


Figure 1 | Approaches to interrogating reaction mechanisms. a, Parameters used for free-energy correlations: Hammett electronic parameters based on the dissociation constant of substituted benzoic acids (top). Steric Sterimol parameters: substituent length (L), and minimal (B_1) and maximal (B_5) widths perpendicular to the length (bottom). **b**, Energetic considerations in free-energy relationships (top) and in isotopologue vibrational energies (bottom). $\Delta\Delta G^{\ddagger}$ is the difference in Gibbs free energy between the two selectivity-determining transition states, E_A is the activation energy barrier and ZPE is the zero-point energy. **c**, Simulated infrared spectrum of benzoic acid and functional group spectral ranges.

[1]Department of Chemistry, University of Utah, 315 South 1400 East, Salt Lake City, Utah 84112, USA.



e

R	CMe_3	CCl_3	Ph	F_5Ph	*p*-*t*-BuPh	Range
ΔΔG‡	1.73	0.74	0.51	0.02	0.77	—
B_1	2.92	2.58	1.77	1.77	2.86	—
ν_1	1,563.6	1,567.3	1,564.6	1,565.5	1,565.1	4.0 cm^{-1}
ν_2	1,665.3	1,667.1	1,667.4	1,668.6	1,668.0	3.4 cm^{-1}
ν_3	1,688.9	1,690.4	1,688.8	1,689.3	1,689.3	2.2 cm^{-1}

Figure 2 | Using infrared vibrations and Sterimol values to correlate enantioselectivity. a, Reaction scheme for the desymmetrization of bisphenols. Reactant torsion is proposed to have a role in the mechanism. **b**, Correlation between Sterimol values and enantioselectivity (normalized model), including the training set, previous sterically modulated validations, and new simultaneously sterically and electronically modulated validations. **c**, Correlation between Sterimol B_1 (minimal width), vibrations and enantioselectivity (normalized model). **d**, Illustration of the vibrational frequencies used for the correlation of enantioselectivity: ν_1, antisymmetric ring stretch with secondary C–H bends and a C–O stretch; ν_2, symmetric ring stretch with a secondary O–H bend; ν_3, antisymmetric ring stretch with secondary C–C bends and a C–O stretch. **e**, Parameter values for the steric model outliers (CCl_3, F_5Ph and *p*-*t*-BuPh) and for two sterically homologous but electronically divergent substituents (CMe_3 and Ph).

Because the steric and electronic features of **R** modulate enantioselectivity, we speculated that stretches of the bisphenol ring would be sensitive to enantioselectivity trends. These stretches are influenced by the mass and charge of **R**, incorporate various secondary C–H and C–C bends and affect the O–H group. Thus, the frequencies selected for modelling were six computationally derived, distinct sp^2 C=C stretches in the 1,700–1,500 cm^{-1} spectral region, which involve either one or both rings[20–24] (Methods and Supplementary Information). The most predictive, statistically significant model developed (slope, 0.99; intercept, –0.01; $R^2 = 0.94$) contains four parameters (Fig. 2c). These include Sterimol parameter B_1, which describes the minimal radius of **R**, and three computationally derived infrared frequencies (Fig. 2c, d). The derived model is highly predictive for both isolated steric effects and substrates containing concomitant steric and electronic changes.

The extreme outliers in the Sterimol analysis are **R** = $-CCl_3$, $-F_5Ph$ and –4*t*-Bu-Ph (where *t*-Bu is *tert*-butyl) (Fig. 2b). If we consider the Sterimol parameters as descriptors of repulsive steric interactions within the catalytic reaction site, a geometry-based parameter would not suffice to define the first two substituents. Because vibrational analysis takes charge distribution into account, and adds a directional aspect to steric interactions, it is able to address electronically diverse substituents. These results are consistent with a mechanistic hypothesis asserting a direct interaction between the peptide catalyst and the bisphenol substituent at the selectivity-determining step of the reaction[14,19]; such an interaction would be sensitive to substituent electronegativity. Inspecting the model reveals that in the cases of **R** = $-CCl_3$ and $-F_5Ph$, all three vibrations are shifted to a higher frequency relative to their respective non-halogenated steric homologues, $-CMe_3$ and –Ph (Fig. 2e). This generalized trend could indicate that the vibrational parameters are functioning as an electronic correction to the steric description.

It has been proposed that the minimum radius B_1 (Fig. 1a) describes steric effects proximal to the phenol rings, wherein the substituent applies torsion on the rings in a propeller-like strain[14] (Fig. 2a). However, the inability to predict the enantioselectivity of –4*t*-Bu-Ph using Sterimol analysis points to a limitation in applying this parameter to substituted aromatic rings. Sterimol values define this group similarly to a $-CMe_3$ group in terms of B_1. However, on the basis of both the distal location of the $-CMe_3$ group in –4*t*-Bu-Ph and the empirically observed enantioselectivity (enantiomeric ratio of 82:18 for –4*t*-Bu-Ph versus 97.5:2.5 for $-CMe_3$), this substituent more closely resembles an unsubstituted phenyl ring (enantiomeric ratio, 75:25). This significantly restricts the use of Sterimol values for evaluating steric effects in aromatic systems. Because infrared ring-stretching vibrations are modulated in response to substituent steric effects, they represent an auxiliary, directional aspect of the substituent geometry and allow for the prediction of groups such as –4*t*-Bu-Ph.

To explore the potential of infrared vibrational analysis further, as a second case study we evaluated our recently reported enantioselective hydrogenation of 1,1-diarylalkenes[25] (Fig. 3a). The original scope included 1,1-diarylalkenes in which ring substitution patterns were modulated to explore the origin of enantioselectivity (Fig. 3d). An unusual observation was that 3,5-dimethoxy substitutions were required to achieve high enantioselectivity (Fig. 3b). We began examining the origin of this observation using our new technology by including twelve of the originally reported substrates in the training set. This set was selected on the basis of both the rings' substitution patterns, in terms of substituent position (that is, *meta* or *para*), and the diversity of steric and electronic effects, as well as a significant enantioselectivity range. For external validation, four of the original substrates were combined with eight new substrates, which were specifically designed to introduce additional structural patterns.

Mechanistically relevant infrared vibrations were proposed according to structural features of the various substrates. Potential parameters for modelling enantioselectivity were six ring vibrations analogous to those used to describe the bisphenol rings in the previous case study. Additionally, three alkene vibrations were included, because the alkene is directly engaged with the catalyst during the transformation. Evaluating the minimized structures of the diarylalkenes showed that aryl substitution affects ring torsion. Therefore, the measured distance between adjacent *ortho*-carbons on the geminal aryl rings was introduced into the parameter set (Fig. 3c). Finally, considering conjugation between the aryl rings and the alkene, we incorporated intensities and frequencies of two vibration modes that involve both groups (Fig. 3c).

A predictive model (slope, 0.95; intercept, 0.06; $R^2 = 0.88$) with three parameters was determined (Fig. 3b, Supplementary Information). Of particular note is that vibrational intensities were identified as relevant descriptors, in contrast to the bisphenol model, where frequencies are sufficient. Intensity and frequency measure different, but interrelated, effects; frequency is dependent on force constants, bond energy and molecular distances, whereas intensity is a derivative of the dipole moment influenced by molecular symmetry and electronic structure[7,16,17,26]. The three parameters included in the model are the torsion distance, two infrared intensities and a cross-term between these intensities. The

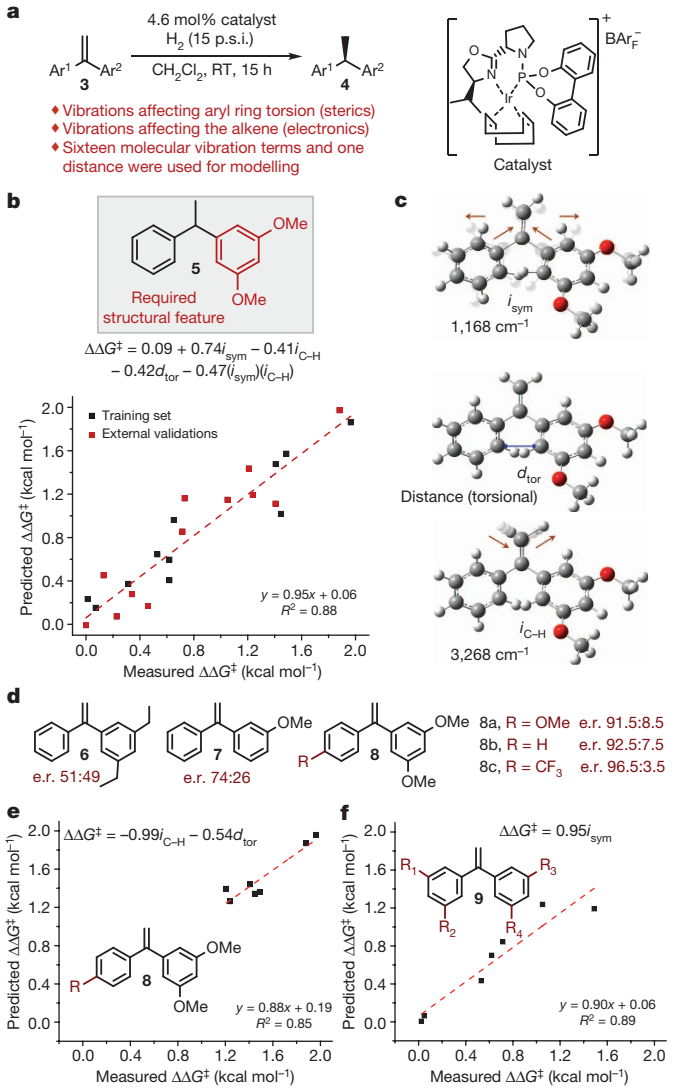


Figure 3 | Using infrared vibrations to correlate enantioselectivity. **a**, Reaction scheme for the enantioselective hydrogenation of 1,1-diarylalkenes. **b**, Mechanistically derived normalized model for the correlation of steric and electronic features of the entire set of substrates to enantioselectivity. **c**, Vibrational intensities and distance measurement used for the correlation of enantioselectivity: d_{tor}, tortional distance between *ortho*-positions; i_{sym}, symmetric central C–C stretch between rings and double bond; $i_{C–H}$, antisymmetric alkene C–H stretch. **d**, Selected substrates from the training set and their respective enantiomeric ratios (e.r.). **e**, Normalized model derived for a subset of all substrates with a 3,5-methoxy motif on one aryl ring and *para*-substitution on the geminal aryl. **f**, Normalized model derived for a subset of all substrates with various *meta*-substitution patterns on both rings.

first intensity, i_{sym}, belongs to the symmetric central C–C stretch between the geminal aryls and the alkene, which is proposed to describe conjugation, and the second, $i_{C–H}$, is the antisymmetric alkene C–H stretch (Fig. 3c).

To assess the effect of each parameter on the prediction of selectivity, two structurally distinct subsets of substrates were evaluated. Hammett σ-constants for *para*-substituents do not adequately describe the enantioselectivity in a first subset where one aryl ring contains a 3,5-dimethoxy motif and the *para*-position of the other is modulated (**8**; Fig. 3e). Deconstructing the original model reveals that this series is defined by only the torsion distance, d_{tor}, and the alkene stretch, $i_{C–H}$ (Fig. 3e). Although the source of the torsional effect may be electronic, these two observations imply that the role of *para*-substitution in determining enantioselectivity is not purely electronic. Interestingly, modulation of *meta*-substituents cannot be described by the same model, and a new model for this second substrate subset relies only on the i_{sym} vibration (**9**; Fig. 3f). It has been proposed[25] that *meta*-substituents could act as directing groups by pre-coordinating to the catalyst. On the basis of this mechanistic hypothesis, the identified intensity, i_{sym}, may denote a selectivity-imparting, direct interaction of *meta*-substituents with the catalyst. Together, these effects serve to define substrates in which both *meta*- and *para*-positions have been perturbed sterically and electronically. Notably, in the full model (Fig. 3b), a cross-term of the two intensities implies a relation between these components for the description of selectivity, suggestive of synergistic effects between the aryl substituents and the alkene.

In the final case study, we sought to expand this technique by exploring site selectivity and simultaneously interrogating two reactants. This is demonstrated in the context of our recently reported redox-relay oxidative Heck arylation, which is highly enantioselective for a broad range of reaction partners and uses a simple chiral ligand[27,28] (Fig. 4a). An interesting mechanistic aspect of this process is the site selectivity of the organometallic addition, leading to constitutional isomers (defined as γ and β). As observed empirically, the respective natures of the alkene and aryl sources both affect site selectivity. Specifically, a Hammett correlation has been observed as a function of the boronic-acid derivative (but only for *meta*- and *para*-substituted examples), with the most electron poor leading to high site selectivity for γ-addition[27]. Furthermore, as the distance between the alkene and the alcohol is increased, the selectivity is reduced. The latter result was initially correlated to the difference in ^{13}C NMR chemical shifts of the two alkene carbons, also suggesting an electronic origin to site selectivity. Finally, although we were unable to correlate the alkene substituent effect quantitatively, we observed a qualitative trend of larger substituents leading to higher site selectivity for γ-addition. These results suggest that both steric and electronic effects affect site selection, providing an exciting platform to utilize our new infrared-vibration-based parameters.

We elected to construct a model for this reaction in three stages. The first stage was aimed at determining whether infrared vibrations can be combined with Hammett analysis to model site selectivity as a function of both boronic-acid and alkene substituents. A specific hypothesis was that the alkene C=C stretching frequency would be sensitive to substituent changes at **R** (Fig. 4b). To test this, we designed[29] an eight-membered library by modulating the alkenol sterically (**R** = *i*-Pr (*iso*-propyl), Et (ethyl), Me) and the boronic acid electronically (**E** = CO_2Me, F, OMe). Using the C=C stretching frequency in combination with the Hammett σ-value provided an excellent model (slope, 0.98; intercept, 0.02; $R^2 = 0.98$; Fig. 4b). These results suggest that the C=C stretching frequency effectively describes the substituent effect on the alkene.

The second stage of this study was directed at the more profound question of describing simultaneous steric and electronic variation. Specifically, we wished to assess whether vibrational analysis could be a substitute for Hammett analysis, yet also allow for the modelling of multiple substituents on the ring or *ortho*-substitution—a significant deficiency of Hammett analysis. Inspired by this question and the classic Hammett analysis, we elected to use infrared vibrations of substituted benzoic acids as a generalized parameter system and as a mimic for arylboronic acids. We chose to examine a single homoallylic alcohol in combination with 12 distinctive aryl boronic acids, including *ortho*-substituted examples. An excellent model was derived, incorporating a ring stretch intensity, the carboxylic acid C=O stretch frequency and a cross-term between the two (slope, 0.95; intercept, 0.08; $R^2 = 0.94$; Fig. 4c). The cross-term is the most important term, with the largest regression coefficient, suggesting that, together, these two vibrations synergistically serve to describe the various effects induced by the boronic-acid component. Markedly, substrates with *ortho*-substitution were effectively predicted.

As the final stage, a model was desired to account for all the products with measured site-selectivity ratios, including those with different

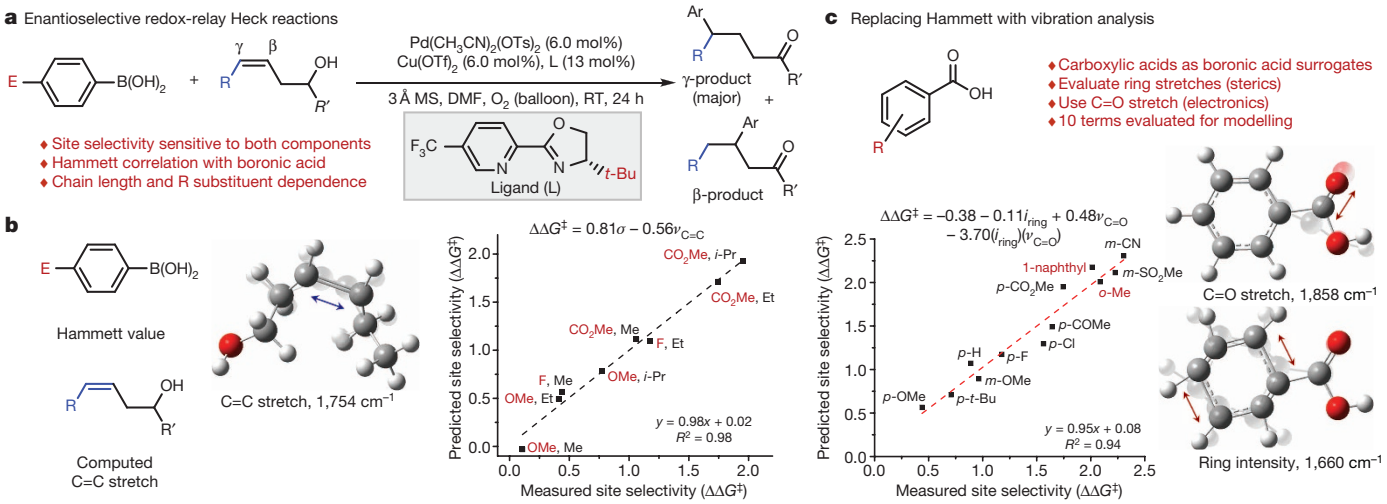


Figure 4 | Using infrared vibrations to correlate site selectivity. **a**, Enantioselective redox-relay Heck reaction scheme and mechanistic considerations. **b**, Correlation of site selectivity to the arene Hammett value and alkenol double-bond stretch (normalized model). **c**, Correlation between site selectivity with different aryl boronic acids and vibrational parameters that describe the aryl moiety (normalized model).

alkene geometries, chain lengths, **R** groups and arylboronic acids (Fig. 5a). A model was developed that effectively predicts 17 external validations (slope, 1.01; intercept, 0.03; $R^2 = 0.92$; Fig. 5b, c). The terms describing the alkene are the sp^2 C–H symmetric stretch and a cross-term between it and the C=C stretch. In this case, the difference relative to those above (Fig. 4b, c) is that this inclusive model must adequately describe the effect of chain length and alkene geometry. Impressively, only two examples of these variations are included in the training set. The arene is described by the same benzoic-acid ring stretch intensity as above, as well as its cross-term with the carbonyl stretch, but an additional ring frequency improves the model. Again, this is presumably due to the more complex interactions between the diverse alkenols and the boronic acids. An intriguing feature of this type of more comprehensive modelling is the potential ability to apply it in extrapolative fashion to predict the performance of new substrates. Because this is a reaction under further investigation, the present analysis will inform mechanistic hypotheses and reaction development.

In this work, we have introduced a method to model reaction selectivity in catalytic systems based on vibrational parameters. The value of this approach is that simultaneous changes to steric and electronic properties can be accounted for and successfully evaluated, overcoming a significant limitation of classical parameter systems.

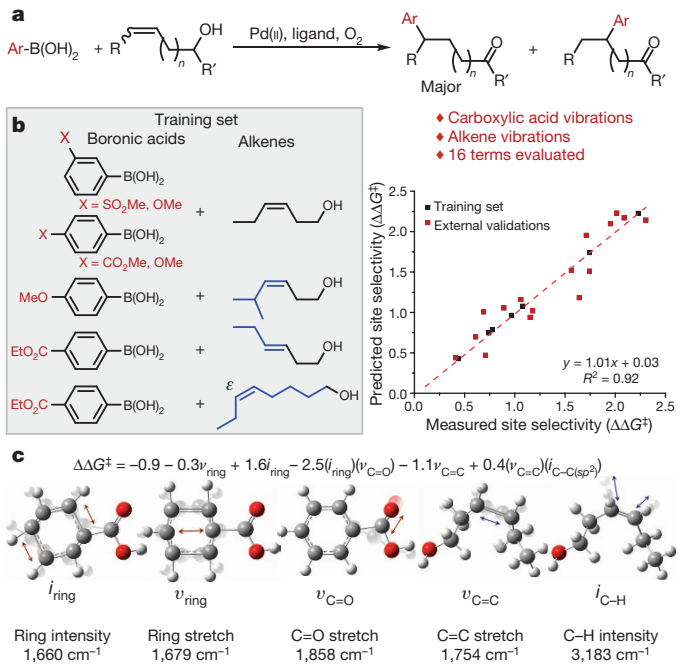


Figure 5 | Developing a comprehensive model. **a**, Enantioselective redox-relay Heck general reaction scheme. **b**, The training set used to develop a model describing site selectivity for products with different alkene geometries, chain lengths, **R** groups and aryl boronic acids. **c**, Normalized model and vibrational terms used for the correlation of enantioselectivity. The two terms on the right, C = C stretch and C–H intensity, are used to describe the alkenol, and the three on the left, C = O stretch, ring intensity and ring stretch, describe the aryl.

METHODS SUMMARY

To execute a strategy that involves computationally derived infrared vibrations as molecular parameters for the mathematical interpretation of selectivity in catalytic reactions, two steps were performed. The first step was identifying vibrations that are sensitive to substituent effects within a given reaction. Realizing this step requires an initial hypothesis for selecting mechanistically relevant vibrations, because complex molecules have an abundance of vibrational modes. Additionally, the selected vibration modes must be identified for all modulated substrates, thus representing the same effect throughout the data set. The abundance, superposition and coalescence of vibration modes lead to complex infrared spectra. This complexity stresses the utility of computed infrared spectra because individual vibrations can be visualized, allowing us to pick out the correct vibration terms for each set of molecules studied. Thus, to produce predictive linear regression models, in each case study we selected specific infrared vibrational modes with a mechanistic hypothesis in mind. Indeed, we found that a hypothesis-based interrogation of the experimental space is required to suggest parameters that can yield predictive, mechanistically relevant models. The second step requires design and evaluation of an empirical training set, which encompasses a systematic variation of substituent properties[29]. The experimental output, enantiomeric or site-selectivity ratios, were mathematically modelled through linear regression techniques to reveal which of the proposed parameters allow for the prediction of new outcomes[30] (Methods and Supplementary Information). The models produced were evaluated for their goodness of fit, and their robustness is demonstrated by external validations' goodness of fit. The nearer the R^2 and slope values are to one (indicating a tight, one-to-one correlation between predicted and measured outcomes) and the nearer the intercept is to zero (indicating minimal systematic error), the more robust the model. Of the potential models, those containing a minimal number of parameters were preferred, because this allows for a mechanistically informative interrogation.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 November 2013; accepted 7 January 2014.

1. Williams, A. *Free Energy Relationships in Organic and Bio-Organic Chemistry* (Royal Society of Chemistry, 2003).

2. Verloop, A., &. Tipker, J. Physical basis of Sterimol and related steric constants. *Pharmacochem. Libr.* **10,** 97–102 (1987).
3. Taft, R. W. Linear steric energy relationships. *J. Am. Chem. Soc.* **75,** 4538–4539 (1953).
4. Hammett, L. P. The effect of structure upon the reactions of organic compounds. Benzene derivatives. *J. Am. Chem. Soc.* **59,** 96–103 (1937).
5. Charton, M. Steric effects. I. Esterification and acid-catalyzed hydrolysis of esters. *J. Am. Chem. Soc.* **97,** 1552–1556 (1975).
6. Hansch, C., Leo, A. & Taft, R. W. A survey of Hammett substituent constants and resonance and field parameters. *Chem. Rev.* **91,** 165–195 (1991).
7. Anslyn, E. V. & Dougherty, D. A. *Modern Physical Organic Chemistry* (University Science Books, 2006).
8. Jacobsen, E. N., Zhang, W. & Guler, M. L. Electronic tuning of asymmetric catalysts. *J. Am. Chem. Soc.* **113,** 6703–6704 (1991).
9. Harper, K. C. & Sigman, M. S. Three-dimensional correlation of steric and electronic free energy relationships guides asymmetric propargylation. *Science* **333,** 1875–1878 (2011).
10. Curtin, D. Y. Stereochemical control of organic reactions differences in behaviour of diastereoisomers. *Rec. Chem. Prog.* **15,** 110–128 (1954).
11. Jaffe, H. H. A reexamination of the Hammett equation. *Chem. Rev.* **53,** 191–261 (1953).
12. Hammett, L. P. Some relations between reaction rates and equilibrium constants. *Chem. Rev.* **17,** 125–136 (1935).
13. Charton, M. The application of the Hammett equation to ortho-substituted benzene reaction series. *Can. J. Chem.* **38,** 2493–2499 (1960).
14. Harper, K. C., Bess, E. N. & Sigman, M. S. Multidimensional steric parameters in the analysis of asymmetric catalytic reactions. *Nature Chem.* **4,** 366–374 (2012).
15. Peng, C. S., Baiz, C. R. & Tokmakoff, A. Direct observation of ground-state lactam-lactim tautomerization using temperature-jump transient 2D IR spectroscopy. *Proc. Natl Acad. Sci. USA* **110,** 9243–9248 (2013).
16. Coates, J. in *Encyclopedia of Analytical Chemistry* (ed. Meyers, R. A.) 10815–10837 (Wiley, 2000).
17. Meyer, M. P. New applications of isotope effects in the determination of organic reaction mechanisms. *Adv. Phys. Org. Chem.* **46,** 57–120 (2012).
18. Harper, K. C., Vilardi, S. C. & Sigman, M. S. Prediction of catalyst and substrate performance in the enantioselective propargylation of aliphatic ketones by a multidimensional model of steric effects. *J. Am. Chem. Soc.* **135,** 2482–2485 (2013).
19. Gustafson, J. L., Sigman, M. S. & Miller, S. J. Linear free-energy relationship analysis of a catalytic desymmetrization reaction of a diarylmethane-bis(phenol). *Org. Lett.* **12,** 2794–2797 (2010).
20. Valero, R., Gomes, J. R. B., Truhlar, D. G. & Illas, F. Good performance of the M06 family of hybrid meta generalized gradient approximation density functionals on a difficult case: CO adsorption on MgO(001). *J. Chem. Phys.* **129,** 124710 (2008).
21. 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.* **120,** 215–241 (2008).
22. Schäfer, A., Huber, C. & Ahlrichs, R. Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *J. Chem. Phys.* **100,** 5829 (1994).
23. Schäfer, A., Horn, H. & Ahlrichs, R. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.* **97,** 2571 (1992).
24. Gaussian 09 rev. C.01 (Gaussian Inc., Wallingford, 2009).
25. Bess, E. N. & Sigman, M. S. Distinctive meta-directing group effect for iridium-catalyzed 1,1-diarylalkene enantioselective hydrogenation. *Org. Lett.* **15,** 646–649 (2013).
26. Dunitz, J. D. & Ibberson, R. M. Is deuterium always smaller than protium? *Angew. Chem. Int. Ed.* **47,** 4208–4210 (2008).
27. Mei, T.-S., Werner, E. W., Burckle, A. J. & Sigman, M. S. Enantioselective redox-relay oxidative heck arylations of acyclic alkenyl alcohols using boronic acids. *J. Am. Chem. Soc.* **135,** 6830–6833 (2013).
28. Werner, E. W., Mei, T. S., Burckle, A. J. & Sigman, M. S. Enantioselective Heck arylations of acyclic alkenyl alcohols using a redox-relay strategy. *Science* **338,** 1455–1458 (2012).
29. Livingstone, D. *Data Analysis for Chemists* (Oxford Univ. Press, 1995).
30. Draper, N. R. & Smith, H. *Applied Regression Analysis* (Wiley, 1998).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the US National Science Foundation (CHE-0749506). We thank S. J. Miller and S. Yoganathan for discussions and for providing the peptide catalyst used in these studies. The support and resources of the Center for High Performance Computing at the University of Utah are gratefully acknowledged.

Author Contributions A.M. and M.S.S. had the idea for the work; A.M. and E.N.B. performed the experiments; A.M. carried out computational and modelling analyses; A.M., E.N.B. and M.S.S. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.S.S. (sigman@chem.utah.edu).

METHODS

In this work, we develop linear regression models using parameters that pertain to steric and electronic properties of the substrates in each case study. For the development of a model, mechanistically relevant parameters were equated to the enantioselectivity or site selectivity in terms of relative energy—$\Delta\Delta G^{\ddagger}$ (kcal mol^{-1}). The first step requires selection of an adequate library that could be used as a training set for the identification of a model that describes selectivity. An oversized, poorly designed training set could lead to the inclusion of parameters that describe idiosyncrasies of the training set rather than true reactivity or selectivity trends[31]. Design-of-experiment principles guide training set selection, that is, systematic variation of substituent properties[29].

For the case study on the desymmetrization of bisphenols, steric models were previously reported[14,19] and, hence, we were able to minimize the substrate training set by comparing the Sterimol B_1 and enantioselectivity values of these studies. It was assumed that because the B_1 values have the greatest effect on enantioselectivity, substrates with similar B_1 and enantioselectivity values are redundant in the training set. In the subsequent case studies, we applied chemical intuition concerning steric and electronic perturbations to select a training set. A general consideration in picking training sets is choosing a minimal number of substrates that exemplify many of the structural variations present in the entire data set. Having this in mind, it should be noted that we purposefully excluded some substrates from the training sets, with the intention of interrogating the ability of a model to externally predict these substrates. Such an external validation establishes the generality of a model, because it is able to predict variations that were not explicitly incorporated in the training set. One such deliberate omission of substrates was *ortho*-substituted boronic acids in the training set in the case study on Heck arylation.

The second step includes four stepwise regression algorithms (for details, see Supplementary Information) that assess the significance of each parameter by applying statistical criteria. To realize this assessment, each set of parameter values is normalized by subtracting its respective mean and dividing by its standard deviation. The four stepwise regression algorithms are built into the MATLAB statistical toolbox and add or remove normalized parameters from an initial model according to a P value threshold[32,33]. Additional suggestions for models are inputted manually on the basis of the results of the four preliminary algorithm runs or on mechanistic hypotheses. A linear fit is performed to probe each manually suggested model, and each suggestion is examined as an initial model for a subsequent stepwise regression iteration to seek a more effective model. After identifying several statistically probable models, an external validation is carried out to determine the predictive nature of the proposed models for substrates that are not included in the training set. This procedure allows us to propose statistically probable models that describe the training set, and then to assess the predictive efficacy of each model towards an external validation set.

The theoretical context of this work is epitomized by Hammett's seminal observation that the acidity of benzoic acid derivatives can be correlated to new equilibria and, ultimately, to reaction rates[4,6,12] (Fig. 1a). Such types of free-energy relationships have been an extraordinarily revealing tool for reaction study[1,8,9,14,18,34–38]. However, a significant limitation is the inability of a single parameter to describe a substituent that introduces perturbations to more than one molecular property. As an example, Hammett parameters can be used only to describe the *ortho*-position on a benzene ring in cases where there is no steric effect involved, limiting their applicability to data sets that contain solely electronic trends[13]. Conversely, steric molecular descriptors, which relate to the spatial arrangement of atoms in a molecule, such as Taft[3,39], Charton[5,40–43] or Sterimol[2,44–46] parameters (Fig. 1a), can be used only to interrogate a reaction trend in the absence of significant electronic effects.

To overcome the aforementioned limitations, we applied a parameter set of molecular vibrations for the quantitative analysis of concurrent steric and electronic structural perturbations. Therefore, in each of the disclosed case studies, we selected and applied several infrared frequencies and intensities, as well as additional mechanistically relevant parameters that are presumed to have a bearing on selectivity. This approach is reminiscent of developing quantitative structure activity relationship (QSAR) regression models for the exploration of chemical or biological systems[47,48]. Yet, so far, infrared vibrations have not been an established parameter of choice in QSAR[49–51]. As an example, the evaluation of the fingerprint region of experimental infrared spectra as a whole (between 1,500 and 600 cm^{-1}), underperformed relative to models derived from typical QSAR parameters[51]. Nevertheless, by proposing infrared vibrational-energy-derived parameters that represent mechanistically pertinent features of the substrates and reactions at hand, it is possible to produce predictive, meaningful correlations.

31. Baldi, P. & Brunak, S. *Bioinformatics: the Machine Learning Approach* 2nd edn, 5–7 (MIT Press, 2001).
32. Goodman, S. A dirty dozen: twelve *P*-value misconceptions. *Semin. Hematol.* **45,** 135–140 (2008).
33. MATLAB student version v.7.14.0.739 (R2012a) (MathWorks Inc., 2012).
34. Jensen, K. H. & Sigman, M. S. Evaluation of catalyst acidity and substrate electronic effects in a hydrogen bond-catalyzed enantioselective reaction. *J. Org. Chem.* **75,** 7194–7201 (2010).
35. Sigman, M. S. & Miller, J. J. Examination of the role of Taft-type steric parameters in asymmetric catalysis. *J. Org. Chem.* **74,** 7633–7643 (2009).
36. Miller, J. J. & Sigman, M. S. Quantitatively correlating the effect of ligand-substituent size in asymmetric catalysis using linear free energy relationships. *Angew. Chem. Int. Ed.* **47,** 771–774 (2008).
37. Harper, K. C. & Sigman, M. S. Using physical organic parameters to correlate asymmetric catalyst performance. *J. Org. Chem.* **78,** 2813–2818 (2013).
38. Bess, E. N. & Sigman, M. S. in *Asymmetric Synthesis II: More Methods and Applications* (eds Christmann, M. & Bräse, S.) 363–370 (Wiley-VCH, 2012).
39. Taft, R. W. Polar and steric substituent constants for aliphatic and o-benzoate groups from rates of esterification and hydrolysis of esters. *J. Am. Chem. Soc.* **74,** 3120–3128 (1952).
40. Charton, M. Steric effects. II. Base-catalyzed ester hydrolysis. *J. Am. Chem. Soc.* **97,** 3691–3693 (1975).
41. Charton, M. Steric effects. III. Bimolecular nucleophilic substitution. *J. Am. Chem. Soc.* **97,** 3694–3697 (1975).
42. Charton, M. Steric effects. 7. Additional V constants. *J. Org. Chem.* **41,** 2217–2220 (1976).
43. Charton, M. Steric effects. 8. Racemization of chiral biphenyls. *J. Org. Chem.* **42,** 2528–2529 (1977).
44. Verloop, A. in *Drug Design* Vol. III (ed. Ariens, E. J.) 133 (Academic, 1976).
45. Verloop, A. & Tipker, J. A comparative study of new steric parameters in drug design. *Pharmacochem. Libr.* **2,** 63–81 (1977).
46. Verloop, A. in *Pesticide Chemistry, Human Welfare and Environment* Vol. 1 (eds Miyamoto, J. & Kearney, P. C.) 339–344 (Pergamon, 1983).
47. Shahlaei, M. Descriptor selection methods in quantitative structure–activity relationship studies: a review study. *Chem. Rev.* **113,** 8093–8103 (2013).
48. Dearden, J. C. & Cronin, M. T. D. in *Smith and Williams' Introduction to the Principles of Drug Design and Action* (ed. Smith, H. J. & Williams, H.) 185–209 (2006).
49. Pino, A., Giuliani, A. & Benigni, R. Toxicity mode-of-action: discrimination via infrared spectra and eigenvalues of the modified adjacency matrix. *QSAR Comb. Sci.* **22,** 191–195 (2003).
50. Benigni, R., Giuliani, A. & Passerini, L. Infrared spectra as chemical descriptors for QSAR models. *J. Chem. Inf. Model.* **41,** 727–730 (2001).
51. Benigni, R., Passerini, L., Livingstone, D. J., Johnson, M. A. & Giuliani, A. Infrared spectra information and their correlation with QSAR descriptors. *J. Chem. Inf. Model.* **39,** 558–562 (1999).