Title

Designer substrate library for quantitative, predictive modeling of reaction performance

Journal

Proceedings of the National Academy of Sciences of the United States of America, 111(41)

ISSN

0027-8424

Authors

Bess, Elizabeth N
Bischoff, Amanda J
Sigman, Matthew S

Publication Date

2014-10-14

DOI

10.1073/pnas.1409522111

Peer reviewed

Designer substrate library for quantitative, predictive modeling of reaction performance

Elizabeth N. Bess, Amanda J. Bischoff, and Matthew S. Sigman[1]

Department of Chemistry, University of Utah, Salt Lake City, UT 84112

Edited by Eric N. Jacobsen, Harvard University, Cambridge, MA, and approved August 28, 2014 (received for review May 22, 2014)

Assessment of reaction substrate scope is often a qualitative endeavor that provides general indications of substrate sensitivity to a measured reaction outcome. Unfortunately, this field standard typically falls short of enabling the quantitative prediction of new substrates' performance. The disconnection between a reaction's development and the quantitative prediction of new substrates' behavior limits the applicative usefulness of many methodologies. Herein, we present a method by which substrate libraries can be systematically developed to enable quantitative modeling of reaction systems and the prediction of new reaction outcomes. Presented in the context of rhodium-catalyzed asymmetric transfer hydrogenation, these models quantify the molecular features that influence enantioselection and, in so doing, lend mechanistic insight to the modes of asymmetric induction.

asymmetric catalysis | free-energy relationships | computational chemistry

Human brains are highly experienced at recognizing patterns in observed data. Organizing information and drawing connections between data enables general conclusions to be made, whether fast or slow, good or bad, or high or low. Although these qualitative assessments are routinely crafted they are subject to biases, causing evaluations to differ from one individual to another (1). The examination of a reaction's substrate scope often takes on a similarly qualitative air (2–5). A substrate scope for a developed synthetic method typically provides an indication of functional group tolerance and general trends in reaction outcomes for sterically and/or electronically varied substrates. This qualitative approach, which lacks quantitation of how substrate features will influence a reaction's outcome, particularly product selectivity, often limits a reaction's application to contexts with high degrees of similarity to the initial scope library. Additionally, it can be difficult to predict, beyond generalities such as poorly versus well-behaved, how a new substrate will perform under the reaction conditions. Addressing this limitation through quantitative prediction of reaction outcomes would significantly affect how one both develops and applies a new synthetic method while simultaneously imparting fundamental mechanistic insight (6).

To accomplish this goal, an entirely new approach to examining a reaction's substrate scope is required. Because the ultimate goal is to mathematically predict a broad range of reaction outcomes, an initial library of substrates would need to be carefully designed to represent many of the impactful features influencing the reaction. Specifically, one would need to include systematic variation of steric and electronic features of a given substrate class while also limiting the initial size of the substrate library to make this a practical venture. With this in mind, the tenets of design of experiments (DoE) and regression modeling will need to be exploited, where broadly descriptive models are built from data that systematically sample the experimental space to be described (7, 8). The results of this sampling will then be correlated to chemical descriptors—which are hypothesized to define relevant interactions between a substrate and catalyst—using linear regression algorithms for the prediction of reaction outcomes (9–17).

In the context of enantio- or site-selective reactions, we anticipate that this type of strategy would have two far-reaching effects: (*i*) provide mechanistic information of substrate properties that are essential for differentially engaging a catalyst, where one product is favored over others, and (*ii*) enable quantified predictions of how future substrates will behave under the reaction conditions, before performing the experiment. Numerical depiction of patterns in reaction outcomes expands the applicability of developed reaction methods, adding a quantitatively accurate and precise dimension to qualitative expectations of chemical behavior (18). Herein, we describe an approach for constructing a substrate library of ketones that is sterically and electronically varied according to DoE principles and, thus, is amenable to descriptive, predictive quantitative modeling, wherein mechanistic patterns in reaction outcomes are robustly delineated. Particularly, it is the aim of this work to develop a ketone library defined by molecular descriptors that broadly represent properties that are influential in a variety of mechanistically distinct reactions.

The following sections describe a four-step process to designing a substrate scope library that is suited to eventual quantitative modeling of reaction outcomes. These steps are (*i*) identifying parameters to describe reaction sensitivities and define the virtual experimental space, (*ii*) organizing ketones that systematically sample the experimental space and evaluating these ketones' performance in enantioselective reactions, (*iii*) connecting molecular descriptors of ketones to reaction outcomes via linear regression modeling, and (*iv*) applying models to quantitatively predict the performance of new ketones.

Identify Substrate Parameters and Define the Experimental Space

Quantitative modeling initiates with consideration of the experimental ketone space to be described. What substrate changes affect reaction outcomes, and why? Hypothesizing a reaction's

Significance

Product distributions of chemical reactions are dictated by a myriad of interactions between molecular species. Identifying which of these features affects reaction selectivity is a key facet for mechanistically understanding a transformation. Such insight often facilitates optimization as well as indicates which types of substrates (substrate scope) are well suited to the method. Unfortunately, the assessment of impactful features is frequently a qualitative endeavor that would significantly benefit from quantitation. We demonstrate a robust method for developing a varied substrate scope library of ketones, identifying quantitative descriptors of mechanistic significance, and applying these descriptors to mathematically elucidate trends in enantioselective reaction outcomes of rhodium-catalyzed asymmetric transfer hydrogenation. The developed mathematical relationships were used to predict future outcomes of new ketone substrates.

Author contributions: E.N.B. and M.S.S. designed research; E.N.B. and A.J.B. performed research; E.N.B. and M.S.S. analyzed data; and E.N.B. and M.S.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: sigman@chem.utah.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1409522111/-/DCSupplemental.

sensitivity to substrate modifications, and identifying a system by which these changes are numerically depicted, sets the stage for a systematic substrate analysis focused on facets of mechanistic relevance. This process was recently demonstrated in our analysis of Nozaki–Hiyama–Kishi (NHK) propargylation of alkyl ketones (Fig. 1) (11). Measuring dimensional information according to Sterimol parameters, a library of sterically varied methyl ketones and ligands, was prepared according to DoE principles (Fig. 1 *B* and *C*) (19–21). Subjected to NHK propargylation reaction conditions, enantiomeric ratios resulting from each ligand/ketone combination were measured. Linear regression modeling was used to develop a mathematical relationship that related steric features of the substrates (R′ Sterimol B_1 and B_5) and ligands (R″ Sterimol B_1) to $\Delta\Delta G^{\ddagger}$ [$\Delta\Delta G^{\ddagger} = -RT\ln$(enantioselectivity), Fig. 1*D*]. This model enabled the robust prediction of enantioselectivity afforded for new methyl ketones, ligands, and combinations thereof. Additionally, this model quantifies specific aspects of steric differentiation that are key features of ketone facial discrimination.

A constraint of this study was its inability to predictively describe electronically perturbed and more complex ketone substrates, because the training set (data points used for model development) was limited to methyl alkyl ketones. Thus, we aimed to build an expanded, DoE-founded library of ketones bearing steric and electronic variation at both ketone substituent sites, which adds significant complexity to the experimental design (Fig. 2). Two salient molecular features of ketones that we hypothesized to be important, general considerations for describing them are differential steric size between R^1 and R^2, which can enable discrimination between ketone faces, and carbonyl electrophilicity, which likely modulates the early or late nature of the corresponding transition states involved in determining selectivity (22). Owing to the demonstrated effective use of Sterimol parameters for describing the relative size of methyl ketones (discussed above), this multidimensional steric measure—particularly B_1 and B_5, which measure the minimum and maximum sterics, respectively, near the reactive carbonyl—seemed an appropriate choice for describing differential steric bulk.

An interest in describing ketone variation at both R-group sites, for alkyl and aryl substituents, limits the range of parameters that could be effective electronic descriptors. For instance, the commonly used Hammett σ value (the acidity of benzoic acid derivatives) could not be used to parameterize this library, because σ values are limited to describing electronic changes on phenyl rings at *meta* and *para* positions, also precluding description of heteroaromatic arenes and alkyl chains (23). Although σ values are too limited a descriptor for this library, these values are well correlated to carbonyl IR stretching frequencies (24, 25). IR bond vibrational frequencies originate from differential charges and masses across a bond and therefore inherently describe, without limitation, carbonyl electronics and the groups bonded to the carbonyl carbon (9, 18, 26).

Using Sterimol parameters and carbonyl IR stretching frequencies as general descriptors of the anticipated reaction of ketones, it was next necessary to assess the sensitivity limits of these parameters. This determination was undertaken by first virtually populating three categories of ketone substituents: aliphatic, halogenated/oxygenated aliphatic, and arene (Fig. 2*A*). The resulting 52 groups were each included as a substituent on a methyl ketone. Energy minimization and frequency calculations (M06-2X/TZVP) were carried out for this diverse set to evaluate each ketone's carbonyl IR stretching frequency (27–31). From structurally related sets, carbonyl IR stretches' sensitivity to R-group variation was determined. For instance, comparison of *i*Pr, $CH(Et)_2$, and $CH(Pr)_2$ demonstrate little change in carbonyl IR stretching frequency for the latter two methyl ketones (*SI Appendix*, Fig. S1). Thus, of this trio only the *i*Pr group was determined to be within the descriptive range of the carbonyl IR stretch parameter. The results of similar analyses for other R-group sets are presented in Fig. 2*A*, and graphical depiction is given in *SI Appendix*, Fig. S1.

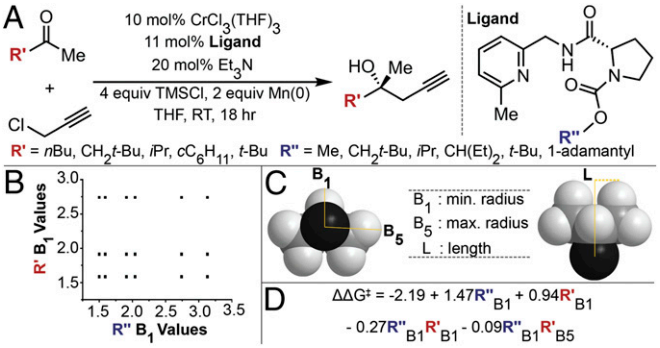

Fig. 1. Using design of experiments to evaluate steric effects. (*A*) NHK propargylation of alkyl ketones. Steric variation at R′ and R″. (*B*) DoE approach to the relatively even distribution of substrate and ligand variation according to B_1 values. (*C*) Depiction of an isopropyl substituent's Sterimol measurements. (*D*) Mathematical model of NHK propargylation.

Once complete, this sensitivity analysis reduced the initial 52-membered ketone set to 32 methyl ketones, each hypothesized to be distinct according to Sterimol and carbonyl stretching parameters. The next stage was to consider how the experimental space for multisite, multieffect ketone variation could be identified. To accomplish this, eight methyl ketone R groups were selected in DoE fashion (Fig. 2*B*) to representatively sample the methyl ketone library. All unique combinations of these eight groups at R^1 and R^2 yielded 28 ketones for which differential Sterimol values ($Sterimol_{R1} - Sterimol_{R2}$) and carbonyl IR stretch frequencies were measured from energy-minimized structures. The identified ketone experimental space is given in Fig. 3*A*.

Organize and Evaluate Ketones That Systematically Sample the Experimental Space

After defining the substrate scope space, this bounded region was populated with ketones that broadly span its dimensions (Fig. 3 *A* and *B*). All but one of these ketones, **16**, are commercially available, enabling ready evaluation of this designer DoE library. The conceptual framework upon which the DoE library was built is best evaluated by subjecting the ketones to a reaction. Then, linear regression models can be developed to describe the observed outcomes as related to ketone changes. Finally, the model's validity is determined by assessing how well the model predicts the reaction outcomes of new ketones.

To limit factors that would confound assessment of the described approach's effectiveness for DoE library development, it was desirable to subject the library to reaction conditions where measured reaction outcomes are (*i*) highly reproducible, (*ii*) sensitive to structural changes to ketones, and (*iii*) rationalized according to previous mechanistic work. These requirements allow the method of DoE library development to be assessed in a manner that is not contingent upon the reaction itself. Additionally, corroborating the developed model with prior mechanistic work adds credence to the use of future DoE-founded modeling approaches for lending mechanistic insight in the absence of computational transition state models.

Rhodium-catalyzed asymmetric transfer hydrogenation (ATH) is a reaction that satisfies the requirements for DoE library assessment (32–34). Of particular note, computational models of the ruthenium ATH variant's selectivity-determining transition state have been investigated (35). These models suggest that the favored diastereomeric transition state benefits from a key stabilizing C–H/π interaction between a C–H bond of the pentamethylcyclopentadienyl ligand's methyl group and the pi cloud of a ketone substrate's arene (Fig. 3*C*). This proposed transition-state model suggests that aryl/alkyl ketones, which can engage with the catalyst through a C–H/π interaction, and alkyl/alkyl ketones, which cannot participate in this intermolecular interaction,

CHEMISTRY

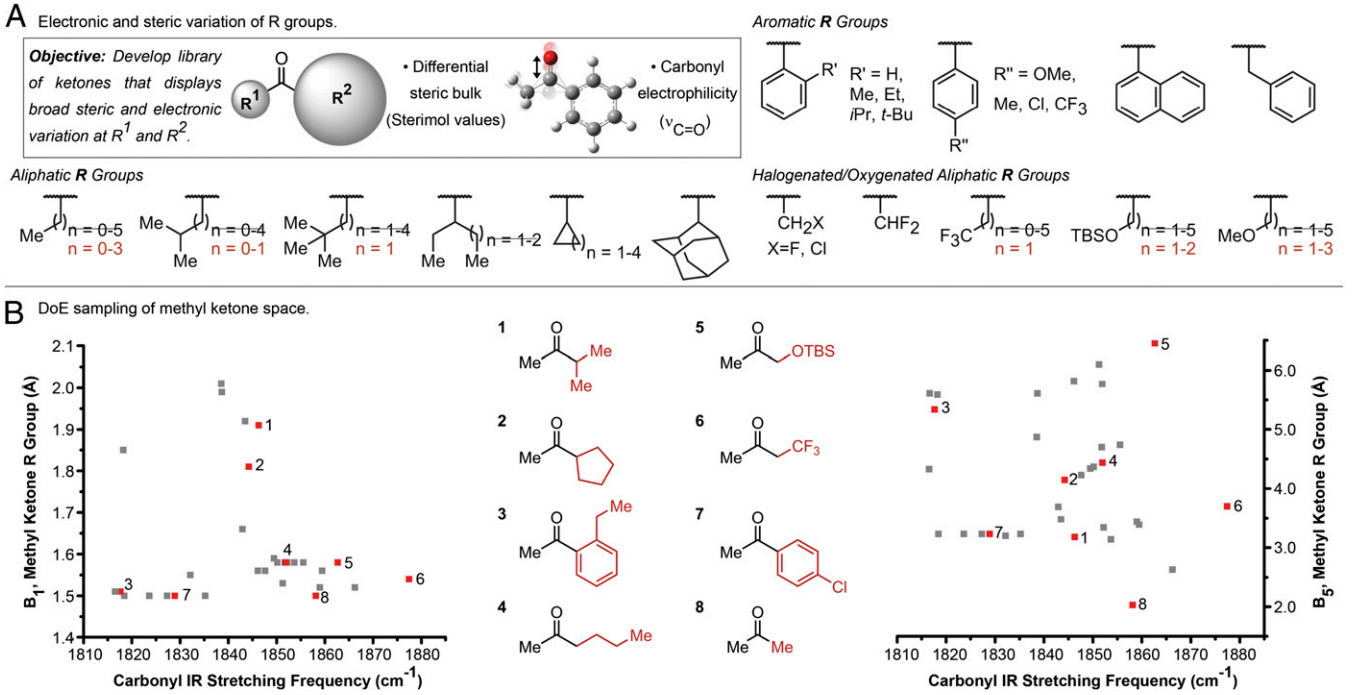

Fig. 2. Building the experimental space. (*A*) A 52-membered set of methyl ketones sampling the steric and electronic ketone space was reduced to a 32-membered library. The reduction was performed by assessing putatively relevant ketone descriptors (Sterimol and carbonyl IR stretching frequency values) via a sensitivity analysis (depicted in *SI Appendix*, Fig. S1), the results of which are given in red in *A*. (*B*) Plots of the 32-membered set of methyl ketones that was identified via sensitivity analyses. Red data points represent a reasonably even, DoE-type sampling of these sterically and electronically described spaces.

may behave as two distinct substrate classes in the ATH reaction (36).

Supposing that these two types of ketones rely on different modes of asymmetric induction, we determined that each ketone class should be independently modeled. This two-pronged approach allows for an optimal description of the unique features relevant to each ketone class's selectivity-determining interactions. Interestingly, assessment of the DoE library's graphical depiction according to carbonyl IR stretching frequencies and Sterimol values demonstrates a natural divide between the two ketone classes (Fig. 3*A*).

Quantitatively Connect Reaction Outcomes and Molecular Descriptors of Ketones

The aryl/alkyl and alkyl/alkyl ketone DoE libraries were each subjected to ATH conditions, and the resulting enantiomeric ratios, in the form of $\Delta\Delta G^{\ddagger}$ [$\Delta\Delta G^{\ddagger} = -RTln([S]/[R])$], were tabulated (*SI Appendix*, Tables S1 and S2) (10, 37). To quantitate and interpret the key selectivity determinants in each ketone class's library a set of ketone parameters that is capable of detailing the selective process was required. Measured from computationally energy-minimized ketone structures, parameters were included in this set based on hypotheses of their mechanistic significance. These hypotheses are continually informed and refined in an iterative process according to the failures and successes of the explanatory parameters that are evaluated via regression. For instance, and as described in greater detail below, previously established steric parameters were ineffective descriptors of the enantioselective role of differential steric effects in the ATH system. This deficiency fueled development of a new steric measure better suited to the demands of the ATH system (and potentially other systems) and provided insight into the steric dimensions of mechanistic relevance. Although there are many possible parameter combinations with potential descriptive relevance, below is described the logic used to arrive at the parameter set that was used for regression modeling.

To begin constructing the parameter set, the relevant electronic details of ketones were proposed to be described by three descriptors (Fig. 4*A*). First, the vibrational frequency of the carbonyl IR stretch ($\nu_{C=O}$), used to define the library, was included. The intensity of this stretch ($I_{C=O}$) was also considered for its representation of electronic effects. An alternative measure of electronic nature was incorporated through point charges at the four atoms that are conserved throughout the ketone library.

Turning to structural features of the library, it is plausible that asymmetric induction for aryl/alkyl ketone substrates is influenced by the degree of torsion (Tor) between the carbonyl and the arene. This parameter may describe the energy expenditure/stabilization balance between various torting of aryl/alkyl ketones from their energetic minima to conformations where transition state-stabilizing C–H/π interactions can occur. Indeed, in the absence of this term, robust models were not developed (discussed below).

Although this torsion angle cannot be used to describe alkyl/alkyl ketones, where facial discrimination cannot arise from C–H/π stabilizing interactions, a surrogate was envisioned in a scissoring IR vibrational mode. Termed $\nu_{scissor}$ and $I_{scissor}$, the frequency and intensity of this vibration, which induces a compression of the two carbons alpha to the carbonyl (Fig. 4*A*), were added to the alkyl/alkyl ketone descriptor set. Although several other vibrations, and combinations thereof, might describe similar molecular dynamics, this scissoring term was selected owing to the confidence with which it could be consistently identified for all computed alkyl/alkyl ketones.

Finally, the initial DoE library descriptors of sterics, Sterimol B_1 and B_5 values, were revisited for inclusion in the parameter set. Sterimol values provided an ineffective representation of ATH enantioselection, which precluded identification of robust models and, consequently, instigated a reanalysis of steric-effect treatment. Because size proximal to the reactive carbonyl moiety is likely to play a different role in enantioselection than distal steric effects, we assessed means of partitioning steric measures into these distinct units. Sterimol measurements of ethyl and

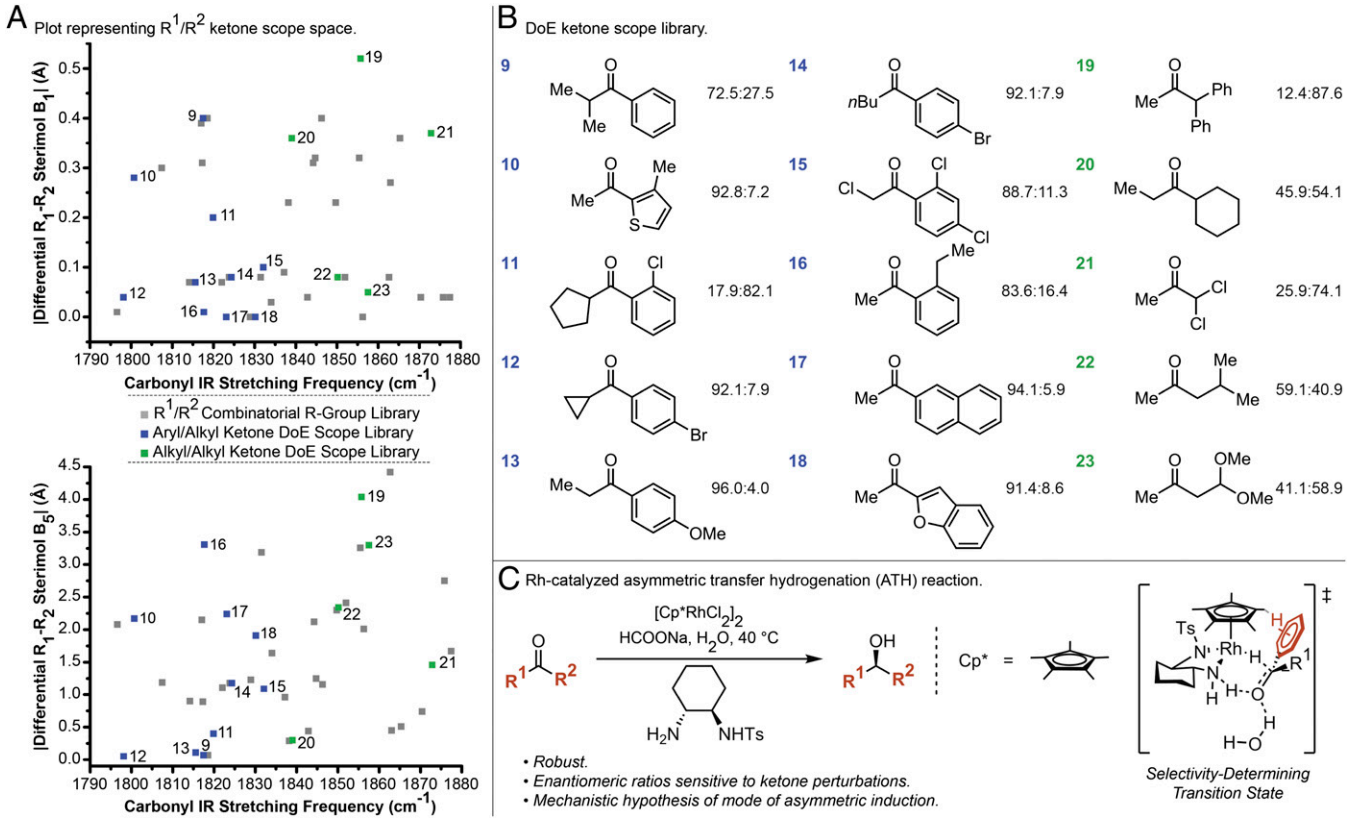

Fig. 3. Identify ketones that systematically sample the experimental ketone scope space. (*A*) Plot of the R^1/R^2 combinations of the eight R groups represented in Fig. 2*D* (gray points). Presumably, this space defines a relevant ketone scope and was evenly sampled (blue and green points) according to DoE principles. (*B*) Substrates that systematically populate the ketone scope. *R*:*S* enantiomeric ratios (er) are given, except for substrate 15, where the *S*:*R* er is given. (*C*) Rhodium-catalyzed ATH reaction, to which conditions the DoE library was subjected, and the transformation's proposed selectivity-determining transition state.

*n*butyl substituents highlight the necessity of this distinction. Whereas the B_1 measures of ethyl and *n*butyl are nearly identical, the B_5 measures differ substantially. Ethyl's B_5 measurement is 3.15 Å, whereas for *n*butyl this parameter measures 4.45 Å—the width to the distal end of the aliphatic chain. Practically, each substituent's steric dimensions that are proximal to the carbonyl are more similar to one another than these measurements indicate.

We addressed the discrepancy between the Sterimol measure of size and practically relevant proximal sterics by slicing R groups into two portions. For aliphatic groups, the proximal effect is defined as the first two carbons and its associated hydrogens; any atoms beyond the proximal segment comprise the distal fragment (Fig. 4*A*). B_1 and B_5 were measured for the proximal unit. B_5 and L were measured for the distal unit. Arenes were segmented in a similar fashion. Proximal sterics is defined as positions spatially equivalent to the *ortho* position on a phenyl ring, beyond which is distal sterics. For each steric fragment two measurements of arene width were obtained, as depicted in Fig. 4*A*. The Sterimol length parameter, L, was measured for the distal steric slice.

From this set of ketone parameters, the combinations thereof that describe the observed trends in the DoE library's enantioselectivity were identified through an iterative process of constructing and assessing various combinations of parameters via MATLAB stepwise regression algorithms [MATLAB Version 8.1.0.604 (R2013a); The MathWorks, Inc.]. This automated mathematical process involves evaluating *P* value statistical measures for each parameter to determine whether the term is an appropriate descriptor of the system. From starting models including either no parameters or all parameters, terms are added to ($P < 0.05$) or removed from ($P > 0.1$) the models according to their *P* values (*SI Appendix*). Through this process, a model for each ketone class was developed (Fig. 4 *B* and *C*). The majority of terms in this model that describe the aryl/alkyl ketones were anticipated to be effective descriptors, because the carbonyl IR stretching frequency ($\nu_{C=O}$) and differential steric bulk were two design parameters. However, with the developed mathematical model these terms' mechanistic significance regarding the reaction's enantioselectivity can now be more precisely, quantitatively understood.

The carbonyl/aryl torsion angle (Tor) was hypothesized to be relevant owing to a proposed, computationally supported C–H/π interaction (35). The normalized regression model (Fig. 4*B*) conveys the relative importance of the torsion term via the magnitude of its coefficient. Bearing the third largest coefficient, the torsion term plays a significant role in the prediction of $\Delta\Delta G^{\ddagger}$. In the raw regression model (*SI Appendix*), Tor bears a negative coefficient, indicating that increases in torsion angle erode enantioselectivity. This correlation is graphically represented in Fig. 4*B*, where deviation from this pattern is described by the model's other parameters. The relationship between torsion and enantioselectivity does directly support that a C–H/π interaction is operative in face selection, which is consistent with computational structural models (35). Tor's relevance in the mathematical model emphasizes the capability of this linear regression approach for delineating distinct mechanistic features among a multitude of potential effects.

Description of the various steric and electronic effects playing roles in enantioselection for the alkyl/alkyl library was afforded with $\nu_{C=O}$ and steric measures (Fig. 4*C* and *SI Appendix*). However, the greatest predictor of enantioselectivity is the cross-term (largest parameter coefficient) describing the synergistic influence of the scissoring vibrational frequency and intensity. Although these parameters are difficult to mechanistically deconvolute, the

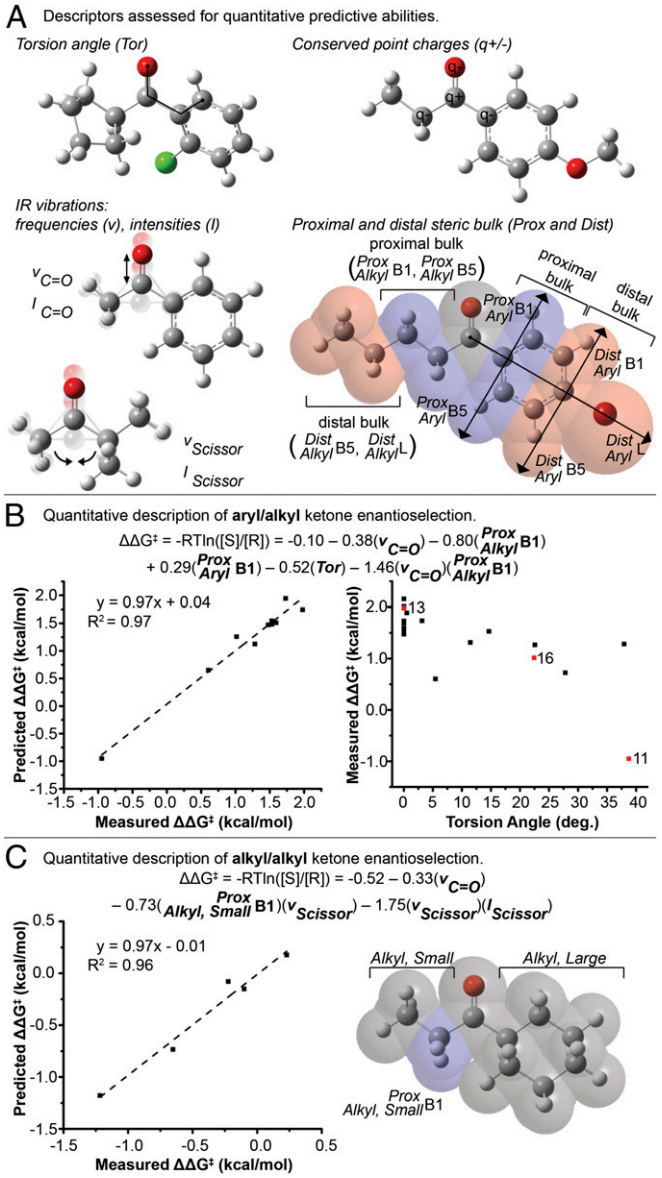

Fig. 4. Quantitatively connecting reaction outcomes to molecular descriptors. (*A*) Numeric molecular descriptors hypothesized to be relevant predictors of ATH enantioselectivity. Mathematical models were developed to describe how attributes of (*B*) aryl/alkyl and (*C*) alkyl/alkyl ketones modulate enantioselective ATH reaction outcomes.

origin of vibrational frequency and intensity in differential mass and charge across a bond indicate the appropriateness of such a parameter.

Robustness of the models for describing the steric and electronic variation in the DoE libraries is determined by correlating experimentally measured enantioselectivity to the enantioselective reaction outcomes predicted by the developed models. Plotting these comparisons (Fig. 4 *B* and *C*) demonstrates that both models exhibit an ability to accurately (slope near 1; *y* intercept near 0) and precisely (R^2 near 1) predict enantioselection for the ketones used in model development.

Predict the Performance of New Substrates

An important application of the developed models lies with their potential to predict the enantioselective outcomes of new ketone substrates. Demonstrating the models' predictive power provides a validation of model robustness and a measure of its broad applicability. That is, optimal models represent generalized patterns in reaction outcomes. The models' reliability was evaluated through external validation experiments. Nine aryl/alkyl ketones (24–32) and three alkyl/alkyl ketones (33–35) were subjected to ATH conditions and enantiomeric ratios were subsequently determined. Using the developed models for each ketone class, predictions of enantioselectivity were made. Fig. 5*A* demonstrates the excellent agreement between predicted and measured enantioselectivities.

Of note, neither the aryl/alkyl substrate 2,2-dimethyl-1-phenylpropan-1-one ($\Delta\Delta G^{\ddagger}$: measured, −0.31 kcal·mol^{-1}; predicted, 13.66 kcal·mol^{-1}) nor the alkyl/alkyl substrate 1-((*tert*-butyldimethylsilyl)oxy)propan-2-one ($\Delta\Delta G^{\ddagger}$: measured, 0.06 kcal·mol^{-1}; predicted 1.59 kcal·mol^{-1}) could be modeled. This may be attributed to inadequate description of the significant steric bulk that each of these substrates bears, which represents large extrapolations from the DoE library. Expanding the DoE library to include similarly bulky substrates would enable elucidation of these outliers' origin and inform further development of parameters finely tuned for describing these substrates.

Conclusions

In the absence of quantitative models, the expected levels of asymmetric induction, that is, the mechanistic influences of multieffect substrate variation, are difficult to forecast beyond the generalities of good, average, or poor. Certainly, qualitatively predicting reversal of face selection, such as observed

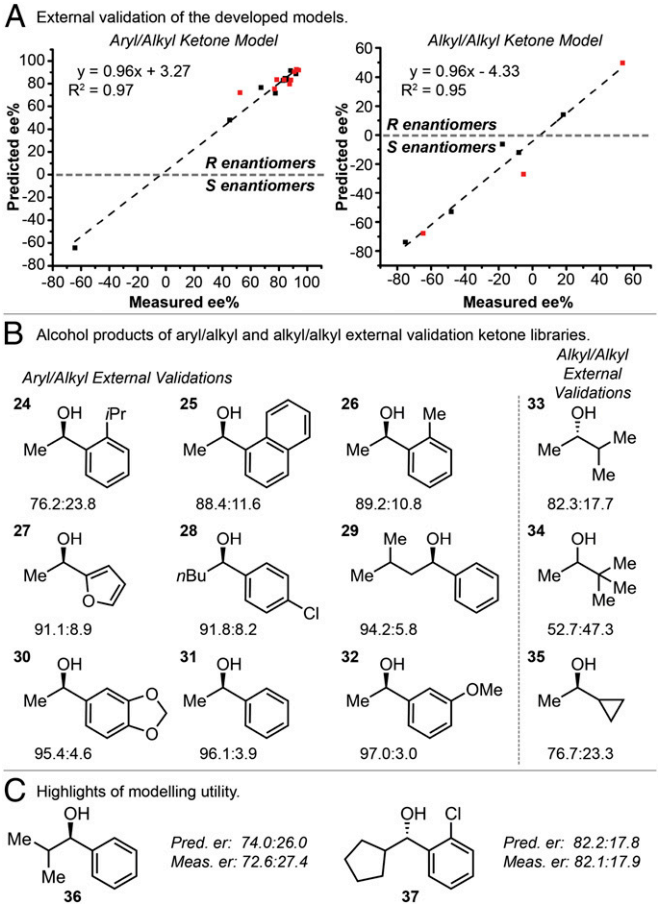

Fig. 5. Predicting the enantioselective performance of new ketones. (*A*) Plots depicting robust external validation of the aryl/alkyl and alkyl/alkyl ketone models. (*B*) Products of aryl/alkyl and alkyl/alkyl external validation, with associated enantiomeric ratios. (*C*) Predicted and measured enantiomeric ratios for two substrates, where the aryl/alkyl model predicted the observed, yet unexpected, reversal of enantioselection.

for the enantiomeric products (36, 37), is even more challenging. Asymmetric catalysis, as a field, has numerous demonstrated apparent outliers, similar to the one described above, which may be predicted more accurately by using the techniques described herein. The presented approach offers an important advancement for comprehensively modeling electronically and structurally diverse substrates. Classic electronic (Hammett) and steric (Charton) linear free-energy relationship analyses are unable to describe the developed multivariate ketone scope libraries, because the substrates boast significant steric and electronic variability (23, 38). From *ortho*-substituted phenyl rings to heteroaromatics to electronically perturbed alkyl substituents, each of these R groups represents a limitation of classical descriptors. An ability to design this broadly diverse ketone library that is also amenable to quantitative modeling demonstrates a new, information-rich approach to reaction scope assessment. Presenting a reaction's substrate scope with a robust quantitative model renders the substrate assessment greater than the sum of its experimentally analyzed constituents. Robust quantitative models enable recognition of patterns by which the reaction outcomes of novel substrates can be predicted, effectively expanding a substrate scope.

Although a significant amount of both intellectual and applied effort was required to develop the approach by which an appropriate DoE-founded library is constructed, following the outlined process will enable libraries of new substrate classes to be readily developed. We have demonstrated that with model training sets of only 5 to 10 substrates, on par with or even less than scope breadths of modern synthetic reports, robust models were developed. Additionally, it is often simple and rapid to perform ground-state computations (completed in a matter of hours) to tabulate data for substrate-specific parameters. Thus, applying the developed methodology requires only a modest effort beyond the standard timeline for reaction method development. Each unique investigation is enabled by the ability to develop parameter sets that are tailored to the features of hypothesized mechanistic importance to afford information-rich quantitative models. Simply through thoughtful designs of substrate scope interrogations the extent of meaningful data obtained from this necessary aspect of reaction development can be significantly enhanced without a correspondingly substantial time investment.

Accordingly, the ongoing goals of this program are now focused on applying this library (and those of other substrate types) to mechanistically diverse reactions. Developing and comparing appropriate predictive models for a variety of reactions of ketones will ultimately reveal the interactions between substrates and catalysts that are conserved throughout each application and those that are uniquely important, broadly affecting mechanistic analyses on a substrate class-wide scale.

Methods

Descriptive quantitative models were developed using MATLAB stepwise linear regression algorithms. Beginning with a set of parameters of hypothesized mechanistic significance, stepwise linear regression was performed and mathematical models were identified. Model robustness was assessed via external validation. Provided in *SI Appendix* is a detailed description of the model development process and rationale.

ACKNOWLEDGMENTS. We thank Anat Milo and Sarah Vilardi for important initial discussions. Support and resources from the Center for High Performance Computing at the University of Utah are gratefully acknowledged. This work was supported by National Science Foundation Grants CHE-0749506 and CHE-1361296 (to M.S.S.).

1. Linhares A, Chada DM (2013) What is the nature of the mind's pattern-recognition process? *New Ideas Psychol* 31(2):108–121.
2. Collins KD, Glorius F (2013) A robustness screen for the rapid assessment of chemical reactions. *Nat Chem* 5(7):597–601.
3. Collins KD, Rühling A, Glorius F (2014) Application of a robustness screen for the evaluation of synthetic organic methodology. *Nat Protoc* 9(6):1348–1353.
4. Collins KD, Rühling A, Lied F, Glorius F (2014) Rapid assessment of protecting-group stability by using a robustness screen. *Chemistry* 20(13):3800–3805.
5. Malik HA, et al. (2014) Non-directed allylic C-H acetoxylation in the presence of Lewis basic heterocycles. *Chem. Sci.* 5(6):2352–2361.
6. Trost BM (1983) Selectivity: A key to synthetic efficiency. *Science* 219(4582):245–250.
7. Carlson R (1992) *Design and Optimization in Organic Synthesis* (Elsevier, Amsterdam).
8. Deming SN, Morgan SL (1993) *Experimental Design: A Chemometric Approach* (Elsevier, Amsterdam).
9. Milo A, Bess EN, Sigman MS (2014) Interrogating selectivity in catalysis using molecular vibrations. *Nature* 507(7491):210–214.
10. Harper KC, Sigman MS (2013) Using physical organic parameters to correlate asymmetric catalyst performance. *J Org Chem* 78(7):2813–2818.
11. Harper KC, Vilardi SC, Sigman MS (2013) 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(7):2482–2485.
12. Harper KC, Bess EN, Sigman MS (2012) Multidimensional steric parameters in the analysis of asymmetric catalytic reactions. *Nat Chem* 4(5):366–374.
13. Harper KC, Sigman MS (2011) Predicting and optimizing asymmetric catalyst performance using the principles of experimental design and steric parameters. *Proc Natl Acad Sci USA* 108(6):2179–2183.
14. Harper KC, Sigman MS (2011) Three-dimensional correlation of steric and electronic free energy relationships guides asymmetric propargylation. *Science* 333(6051):1875–1878.
15. Donoghue PJ, Helquist P, Norrby P-O, Wiest O (2009) Prediction of enantioselectivity in rhodium catalyzed hydrogenations. *J Am Chem Soc* 131(2):410–411.
16. Kozlowski MC, Ianni JC (2010) Quantum molecular interaction field models of substrate enantioselection in asymmetric processes. *J Mol Catal Chem* 324(1-2):141–145.
17. Ianni JC, Annamalai V, Phuan PW, Panda M, Kozlowski MC (2006) A priori theoretical prediction of selectivity in asymmetric catalysis: Design of chiral catalysts by using quantum molecular interaction fields. *Angew Chem Int Ed Engl* 45(33):5502–5505.
18. Bess EN, et al. (2014) Analyzing site selectivity in Rh2(esp)2-catalyzed intermolecular C-H amination reactions. *J Am Chem Soc* 136(15):5783–5789.
19. Verloop A, Hoogenstraaten W, Tipker J (1976) Development and application of new steric substituent parameters in drug design. *Drug Design*, ed Ariens EJ (Academic Press, New York), Vol 7, pp 165–207.
20. Verloop A, Tipker J (1977) A comparative study of new steric parameters in drug design. *Pharmacochem Libr* 2:63–81.
21. Verloop A, Tipker J (1987) Physical basis of STERIMOL and related steric constants. *Pharmacochem Libr* 10:97–102.
22. Jacobsen EN, Zhang W, Guler ML (1991) Electronic tuning of asymmetric catalysts. *J Am Chem Soc* 113(17):6703–6704.
23. Hammett LP (1935) Some relations between reaction rates and equilibrium constants. *Chem Rev* 17(1):125–136.
24. Jones RN, Forbes WF, Mueller WA (1957) The infrared carbonyl stretching bands of ring substituted acetophenones. *Can J Chem* 35(5):504–514.
25. Brown HC, Okamoto Y (1958) Electrophilic substituent constants. *J Am Chem Soc* 80(18):4979–4987.
26. Coates J (2000) Interpretation of infrared spectra, a practical approach. *Encyclopedia of Analytical Chemistry*, ed Meyers RA (Wiley, Chichester, UK), pp 10815–10837.
27. Frisch MJ, et al. (2009) *Gaussian 09, Revision C.01* (Gaussian, Inc., Wallingford, CT).
28. Zhao Y, Truhlar DG (2008) 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(1–3):215–241.
29. Valero R, Gomes JRB, Truhlar DG, Illas F (2008) 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(12):124710.
30. Schäfer A, Horn H, Ahlrichs R (1992) Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J Chem Phys* 97(4):2571–2577.
31. Schafer A, Huber C, Ahlrichs R (1994) Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *J Chem Phys* 100(8):5829–5835.
32. Ohkuma T, et al. (1998) Asymmetric hydrogenation of alkenyl, cyclopropyl, and aryl ketones. RuCl2(xylbinap)(1,2-diamine) as a precatalyst exhibiting a wide scope. *J Am Chem Soc* 120(51):13529–13530.
33. Murata K, Ikariya T, Noyori R (1999) New chiral rhodium and iridium complexes with chiral diamine ligands for asymmetric transfer hydrogenation of aromatic ketones. *J Org Chem* 64(7):2186–2187.
34. Wu X, Xiao J (2007) Aqueous-phase asymmetric transfer hydrogenation of ketones—a greener approach to chiral alcohols. *Chem Commun (Camb)* (24):2449–2466.
35. Yamakawa M, Yamada I, Noyori R (2001) CH/π Attraction: The origin of enantioselectivity in transfer hydrogenation of aromatic carbonyl compounds catalyzed by chiral η6-arene-ruthenium(II) complexes. *Angew Chem Int Ed* 40(15):2818–2821.
36. Ahlford K, Lind J, Maler L, Adolfsson H (2008) Rhodium-catalyzed asymmetric transfer hydrogenation of alkyl and aryl ketones in aqueous media. *Green Chem* 10(8):832–835.
37. Bess EN, Sigman MS (2012) Linear free energy relationships (LFERs) in asymmetric catalysis. *Asymmetric Synthesis II: More Methods and Applications*, eds Christmann M, Bräse S (Wiley-VCH, Weinheim, Germany).
38. Charton M (1975) Steric effects. I. Esterification and acid-catalyzed hydrolysis of esters. *J Am Chem Soc* 97(6):1552–1556.