Published in final edited form as:

Science. 2015 February 13; 347(6223): 737–743. doi:10.1126/science.1261043.

A data-intensive approach to mechanistic elucidation applied to chiral anion catalysis

Anat Milo^{1,†}, Andrew J. Neel^{2,†}, F. Dean Toste^{2,*}, and Matthew S. Sigman^{1,*}

¹Department of Chemistry, University of Utah, 315 South 1400 East, Salt Lake City, UT 84112

²Chemical Sciences Division, Lawrence Berkeley National Laboratory, and Department of Chemistry, University of California, Berkeley, CA 94720

Abstract

Knowledge of chemical reaction mechanisms can facilitate catalyst optimization, but extracting that knowledge from a complex system is often challenging. Here we present a data-intensive method for deriving and then predictively applying a mechanistic model of an enantioselective organic reaction. As a validating case study, we selected an intramolecular dehydrogenative C-N coupling reaction, catalyzed by chiral phosphoric acid derivatives, in which catalyst-substrate association involves weak, non-covalent interactions. Little was previously understood regarding the structural origin of enantioselectivity in this system. Catalyst and substrate substituent effects were probed by systematic physical organic trend analysis. Plausible interactions between the substrate and catalyst that govern enantioselectivity were identified and supported experimentally, indicating that such an approach can afford an efficient means of leveraging mechanistic insight to optimize catalyst design.

Catalyst discovery and development often relies on empirical observations gained through the laborious evaluation of multiple potential reaction variables (1). While high-throughput methods can streamline this process (1–9), the ability to rationally design catalysts that affect chemical reactions in a predictable manner would be a transformative step forward. In catalysis, the principal challenge lies in inferring how catalyst structural features affect the mechanistic aspects of a given chemical reaction, including those that govern selectivity when multiple products are possible (10–12). While mechanistic studies are able to guide the rational design of catalytic systems, traditional approaches are not often suited to address the complexity of modern catalytic transformations. This limitation is especially apparent in cases where selectivity is affected by subtle catalyst and substrate structural features (13), and/or the product-determining step of the reaction occurs after the rate-determining step. In order to address such systems, we envisioned a strategy for mechanistic study involving the application of modern data analysis techniques. This approach relies on the generation of mathematical correlations between quantifiable properties describing the interacting reaction partners' molecular structures (i.e., molecular descriptors) and a measurable outcome of the reaction (e.g., enantioselectivity, represented as the energy difference between transition

*Correspondence to: sigman@chem.utah.edu, fdtoste@berkeley.edu.

†Anat Milo and Andrew J. Neel contributed equally to this work.

states leading to either enantiomer $G^\ddagger = -RT\ln[(S)/(R)], kcal/mol$ (14). Combining appropriate experimental design, data organization, and trend analysis techniques provides the basis to distinguish causal relations, producing testable hypotheses regarding the structural origin of the reaction outcome. New catalysts can be designed and the ability of the models to predict new experimental outcomes can be used as validation of the mechanistic hypotheses. Herein, we demonstrate that this approach enables in-depth mechanistic analysis of interactions that govern enantioselectivity, affording non-intuitive insight into the origin of asymmetric induction and guiding rational catalyst design.

Choice of Case Study

To assess the applicability of this data-driven approach toward mechanistic analysis, a system was sought with a poorly understood mechanism that would be difficult to probe using standard techniques. Accordingly, the field of chiral anion catalysis was appealing, because diverse catalyst-substrate interactions likely contribute to enantioselectivity, but their distinctive effects are difficult to deconvolute. Particularly, the oxoammonium salt (**3**) - mediated enantioselective cross dehydrogenative coupling (CDC) catalyzed by chiral BINOL-based phosphoric acids (PAs, **1**, Fig. 1) bearing triazoles at the 3 and 3' positions reported in 2013 by Toste and coworkers (15) was identified as a prototypical example. This type of reaction could benefit from such an analysis, due to the following challenges. First, high levels of enantioselectivity were necessarily predicated on the oxidant's (16–19) insolubility under the reaction conditions, precluding rigorous kinetic analyses. Second, the enantioselectivity trends with respect to both catalyst and substrate were not obvious, with even modest structural modifications resulting in substantial differences (Fig. 1C). Finally, we hypothesized that enantioselectivity was governed by attractive non-covalent interactions (15). These subtle interactions are ubiquitous in biological and catalytic systems (13), but are difficult to study, or apply toward rational catalyst design, especially if several such interactions are involved in determining reaction outcomes.

The distinct mechanistic features of triazole-PA catalysts are highlighted by the observation that they result in opposite and enhanced enantioselectivities (Fig. 1A) relative to more conventional PAs such as C₈-TRIP (**5**) and TCYP (**6**), representative of BINOL-based PAs that have seen the most extensive use (20–22). Additionally, electronically distinct pyrazolyl (*pyr-1e*) and imidazolyl (*imid-1e*) PAs afford products with significantly reduced enantioselectivities relative to the parent triazolyl (**1e**, Fig. 1A), despite having nearly identical steric profiles. While these data allude to selectivity determination *via* attractive, non-covalent interactions between the catalyst and substrate, such interactions are difficult to further characterize. This limitation is not uncommon in enantioselective catalysis. Thus, our goal was to develop a general, data-driven technique for the evaluation of how subtle structural features impact selectivity, using this reaction as a challenging case study.

Kinetic Isotope Effects

Prior to any mechanistic study focused on the origins of selectivity, we sought to establish the enantioselectivity-imparting step(s) in the catalytic cycle. Bearing in mind the aim of mathematically relating structural features of the reacting components to enantioselectivity,

this knowledge would reveal the elementary step that is being represented by the catalyst and substrate molecular descriptors (*vide infra*). With respect to the general mechanism in Fig. 1B, we sought to distinguish two scenarios: i) enantioselectivity is determined during oxidation of substrate **2** or ii) enantioselectivity is determined during the cyclization of an oxidized intermediate (**A**, Fig. 2A). With respect to the former scenario, it was conceivable that although the stereogenic center is formally set in the cyclization from the oxidized intermediate (**A**), the interactions between the substrate and catalyst during the oxidation event may pre-organize the system for effective enantioselection.

To distinguish between these possibilities, a set of kinetic isotope (KIE) experiments was performed using **2a-d₁** (90 % D incorporation, 74:26 er). We expected that, if the chiral phosphate were involved in substrate oxidation, different KIEs would be observed for the formation of **4a-d₁** when using enantiomeric catalysts. Indeed, (*R*)- and (*S*)-**1e** promoted the reaction with KIEs of 3.42 and 1.08 respectively, suggesting the involvement of the catalyst in the oxidation (Fig. 2B). If enantioselectivity were also established in this step, different enantiomeric enrichment (ee) values would be expected using the enantiomeric catalysts. The observation that the final products exhibited equal but opposite levels of enantioselectivity is consistent with enantioselection occurring during the bond-forming event from an oxidized intermediate such as **A**. This result indicated that the two key steps of this reaction (oxidation and cyclization) proceed under independent Curtin-Hammett control (10, 11), with similar interactions likely governing selectivity in both cases (Fig. 2C). The KIE values resulting from the less-selective catalysts **pyr-1e** and **6** (Fig. 1A and 2B) do not display the same enantiomer-dependency as those from **1e**. This effect is consistent with the assertion that similar catalyst-substrate interactions are involved throughout the mechanism, as well as the triazole substituent's superior ability to interact with substrate **2**. However, the specific nature of this interaction remained undefined. To this end, a thorough analysis of an extensive dataset containing structural perturbations to the catalyst and substrate could serve to illuminate these enantioselectivity-directing interactions.

Experimental Design and Analysis

The collection and organization of diverse datasets is at the foundation of data-driven analysis strategies (23). Therefore, an effective experimental library should include rational changes to various structural features that impact the reaction outcome of interest. To this end, substrates (**2**) were modified at positions hypothesized to influence enantioselectivity (i.e., at the 2-, 4-, and 6-, positions of the benzyl ring and the distal aryl ring, Fig. 1D), using substituents with varied electronic and steric properties (according to their Hammett σ_{para} and Sterimol B₁ values, respectively, see Supplementary Material (SM) p4–8 for additional details). Similarly, catalysts (**1**) were modified at the 2-, 4-, and 6- positions of the aryl ring attached to the triazole. Adamantyl-substituted catalysts **1e**, **pyr-1e**, and **imid-1e** were also included to explore the effect of changes to the heterocyclic ring. In total, 12 substrates and 11 triazolyl catalysts were selected (Fig. 1D). These libraries were then synthesized and the enantioselectivity of each catalyst-substrate combination was obtained. Simultaneously, a diverse array of molecular descriptor values was collected to describe the structural features of each catalyst and substrate, including Sterimol parameters (24), length measurements

from geometry optimized structures, and computationally-derived vibrational frequencies and intensities (see SM p4–8 for details, Fig. 1D) (25). Linear regression algorithms were then applied to various subsets of the data to identify correlations between molecular structure and the experimentally determined enantioselectivity. Subsequently, analysis and refinement of the resulting models were used to produce explicit mechanistic hypotheses that were then tested and validated experimentally.

Modeling Catalyst Heterocyclic Rings

Given the clear importance of the catalyst heterocyclic ring in enantioselectivity determination (*vide supra*), we initially sought to understand the subset of results obtained using catalysts **1e**, **pyr-1e**, and **imid-1e**. Accordingly, using linear regression techniques, the correlation depicted in Fig. 3B was identified from a training set of ten different substrate-catalyst combinations (Fig. 3A and 3C, black squares). Of the large number of steric (26) and vibrational (25) terms initially investigated as molecular descriptors, four discrete vibrational parameters were sufficient to produce a correlation with enantioselectivity: one catalyst descriptor ($\nu\text{Y-N}$, a stretching frequency on the heterocyclic ring), and two substrate descriptors (the stretching frequency of the amide C=O bond, $\nu\text{C=O}$, and stretching frequency/intensity of the C–H bond undergoing oxidation, $\nu/\text{iC-H}$, Fig. 3B and 3D). A cross-term between the catalyst and substrate descriptors improves the overall quality of the model ($\nu\text{Y-N} \times \text{iC-H}$), suggesting a synergistic structural effect. The model was evaluated by plotting measured vs. predicted $G^\ddagger$ values (Fig. 3C), and as a validation of its robustness, the enantioselectivities of ten catalyst-substrate combinations not included in the training set were well-predicted (Fig. 3A and 3C, red stars). A slope approaching unity and intercept approaching zero over the training and validation sets indicate an accurate and predictive model, while the R^2 value of 0.90 demonstrates a high degree of precision. The largest coefficient in this normalized model belongs to the heterocyclic ring vibrational frequency, signifying its substantial role in the quantification of enantioselectivity.

This model is capable of predicting results whose origins are not obvious upon inspection. For example, comparison of the reaction outcomes employing **1e** and **pyr-1e** with substrate **2a** (entries 1 vs 13, Fig. 3) may lead to the conclusion that **pyr-1e** generally affords inferior selectivity. Indeed, experimental results for several additional substrates support this notion, and are accurately predicted by the model (e.g., 2-OMe benzyl substrate **2e**, entries 5 vs 16). However, with substrate **2i** ($R_1 = 2\text{-OMe}$, $R_2 = \text{Ph}$, entries 9 vs 17), the triazolyl and pyrazolyl PAs afford the product with similar levels of enantioselectivity. This counterintuitive result is precisely predicted, indicating that the divergent enantioselectivity displayed by **1e**, compared to **pyr-1e** and **imid-1e**, is adequately addressed by the model.

Trend Analysis

Although the model in Fig. 3B establishes the capacity of the chosen parameters to describe subtle aspects of this system, the ultimate goal of this approach was to discern underlying mechanistic phenomena. This objective could not be achieved using merely the above correlation, as it was produced using a small subset of data, in which the catalyst

heterocyclic rings bore the same substituent (i.e. adamantyl). We hypothesized that the complete dataset contained invaluable information to this end, as it was produced using strategically modified catalysts and substrates with substituents intentionally introduced to probe subtle effects, resulting in 132 enantioselectivities between –54 and 94 % ee, which corresponds to a $33G^\ddagger$ range of 2.8 kcal/mol. In accordance with a data-intensive strategy, none of these measured values was discarded, because even low enantioselectivity carries information regarding the catalyst-substrate interactions at the origin of asymmetric induction.

Yet, prior to producing a global, predictive model, we considered that a series of focused correlations, coupled with an evaluation of overall trends, might serve to reveal fundamental features of the system. To this end, for each individual substrate, a correlation was produced relating its observed enantioselectivity values for the entire set of catalysts with parameters describing the catalyst structure (**2a-2l**, 12 models in total, SM Fig. S3). The same strategy was applied to all aryl-substituted triazole catalysts using parameters describing the substrate structure (**1b-1d**, **1f-1k**, 9 models total, SM Table S7). This organizational scheme was viewed as a means to facilitate the identification of catalyst features that are important for particular substrate types (and vice versa). Substrates or catalysts with similar structures behave analogously not only in a qualitative manner, but also in terms of the molecular descriptors that effectively serve to predict their enantioselectivities (see SM, Fig. S2, S3, S4, and Table S7 for individual substrate and catalyst measured *vs.* predicted $G^\ddagger$ plots and equations). These quantitative correlations, together with systematically organized trends of experimental outcomes, can guide the development of testable mechanistic hypotheses.

To simultaneously inspect multiple aspects of large and intricate datasets, a communicative visualization of data is crucial (27). Thus, we elected to present information gained from these focused mathematical models, alongside multiple observed enantioselectivity results, organized according to trends in catalyst or substrate structural features. Demonstrating this visualization technique, the enantioselectivity trend for each catalyst (each line represents a catalyst, Fig. 4) was plotted as a function of the substrates (each *x* axis tick-line represents a substrate, Fig. 4), and vice versa (Fig. 5). To afford a quantitative trend analysis, the plots were arranged according to which positions were modified on the catalyst or substrate structures, and the corresponding R_1 or R_2 substituent's Sterimol B_1 value (see Fig. S5–S6 for additional visualizations) (24). For example, in the purple frame in Fig. 4, 2-substituted benzyl substrates are displayed from the largest to smallest substituent. The catalyst molecular descriptors required for correlating each subset of substrates are also depicted (Fig. 4), along with the substrate descriptors for each subset of catalysts (Fig. 5). For example, the catalyst molecular descriptors used as parameters for the correlation of enantioselectivity obtained for 2-substituted benzyl substrates are presented below the blue frame in Fig. 4 ($vN-N$, $\angle_{tor}$).

Analysis of this systematic data arrangement reveals that, in general, catalyst performance correlates with the aryl substitution pattern in the order 2,4,6 > 2,6 > 4 (Fig. 5, grey, orange, and blue frames, respectively). Additionally, by juxtaposing Fig. 4 and 5, it appears that the reaction is mainly under catalyst control as catalyst features affect enantioselectivity in a

more considerable and systematic manner (Fig. 5), while, for each substrate, the spread of observed enantioselectivity is broader (Fig. 4). All substrates bearing a 2-substituted benzyl group, even those with substitution on the distal aryl ring (Fig. 4, purple and blue frames, respectively), can be modeled using the torsion angle between the triazole and its substituent, and a triazole vibration frequency ($\angle_{\text{tor}}$, $\nu\text{N-N}$, for the individual models see SM Table S7). The torsion angle represents a steric effect, yet also contains information concerning the conjugation between the triazole ring and its substituent. The vibration frequency can serve as a correction to both of these effects, as it takes into account non-additive features of the substituents' charge and mass distribution (25).

The models for substrates with 4-benzyl substitution (Fig. 4, green frame) contain the same two terms ($\angle_{\text{tor}}$, $\nu\text{N-N}$), and an additional steric descriptor (the catalyst aryl ring minimal width, B_1). Similar interactions with the triazole ring should still be present for these 4-benzyl substrates, but the presence of a B_1 term suggests an additional steric interaction between the substrate and catalyst substituents, avoided in the case of hydrogen at the 4-benzyl position. This claim is supported by the lower enantioselectivities observed for substrates with larger 4-benzyl substituents, especially using catalysts with larger 2,6-groups. Thus, the lack of the catalyst B_1 term in the models describing 2- relative to 4-benzyl substrates, and their overall higher enantioselectivities, are thought to arise from a better accommodation of the former substrates' shapes in the catalyst active site.

The individual catalyst trends and models carry complementary information, revealing the substrate descriptors relevant to each catalyst subset (Fig. 5). The parameters that correlate with 4-substituted aryl catalysts' enantioselectivity (Fig. 5, blue frame), include the substrate carbonyl stretching vibrational frequency and intensity ($\nu/\text{iC=O}$), the amide N-H vibrational frequency ($\nu\text{N-H}$), and a cross-term between the two ($\nu\text{C=O} \times \nu\text{N-H}$). These values vary in response to substitution on the substrate benzyl ring or the distal aryl ring, with the former having a greater effect (see SM, Table S2). These same parameters effectively correlate with the enantioselectivity observed for 2,6-disubstituted aryl catalysts (Fig. 5, orange frame), along with an additional vibrational term (νRing3) describing a benzyl ring stretch. The enantioselectivity spread of catalysts with larger substituents at the 2,4,6-position (e.g., 38–93 % ee for 2,4,6-(Cy)₃-Ph catalyst **1d**, Fig. 5, grey frame) can be described using two terms associated with the benzyl ring (νRing3 , iN-H), stressing this ring's significance in determining enantioselectivity.

Trend Interpretation

Collectively, these results suggest that a π interaction is established between the triazole ring and substrate during the enantioselectivity determining step, the strength of which is modulated by local steric and electronic structural features of both interacting partners (13, 28–34). Furthermore, π interactions are often strengthened by heteroatoms (35–37), which could explain the reduced enantioselectivity values obtained using *imid-1e*, *pyr-1e* and TCYP catalysts compared to **1e** (Fig. 1A), as well as the similar KIE values obtained when using both the (*R*)- and (*S*)- enantiomers of *pyr-1e* and TCYP as catalysts, compared to the divergent ones displayed by (*R*)- and (*S*)-**1e** (Fig. 2B). In relation to the substrate, participation of both the benzyl group and the distal aryl group in putative π interactions are

supported by the presence of molecular descriptors that are sensitive to substitution on these rings in every catalyst model ($vC=O$, $iC=O$, $vN-H$, $vRing3$, Fig. 5, SM Table S2).

The energy stabilization gained from π interactions is affected by the distance and geometry of the rings involved (13, 30–37). If a π interaction between the substrate and triazole is at the origin of enantioselectivity determination, the directionality of the triazole, represented by the torsion angle between the triazole and its substituent, is expected to directly affect enantioselectivity. In agreement with this hypothesis, catalysts with more pronounced torsional effects lead to higher enantioselectivity values for substrates with relatively small substituents at the benzyl 4-position. The torsion angle approaches perpendicularity (90°) due to larger 2,6-substituents on the catalyst aryl ring connected to the triazole (Fig. 4, purple and blue frames, blue lines). Moreover, large catalyst 2,6-aryl substituents are presumed to serve as a steric barrier, docking the substrate in place for an improved overlap with the catalyst triazole ring. Correspondingly, substrates with elongated 4-substituents ($R = Me, OMe$) lead to lower enantioselectivities using catalysts with large substituents at the 2,6-position (Fig. 4 green frame, blue lines). For these substrates, the steric repulsion exerted by the large 2,6-substituents, affords a weaker or less directing π interaction, and subsequently, lower enantioselectivity. Thus, the significance of the torsion angle and vibration parameters for correlating enantioselectivity in the individual models and overall trends (Fig. 4 and S2, Table S7) is proposed to reflect the angle at which the triazole engages the substrate, and the steric role of the catalyst aryl group. Lending further credence to this proposal, catalysts with reduced torsion angles (e.g. catalysts with triazole R substituents: Ph, 4- NO_2 Ph, 4-OMePh, 4- SO_2 MePh) that do not introduce the proposed directional and steric effects, lead to diminished enantioselectivities overall (Fig. 5).

Comprehensive Model and Probes of Mechanistic Hypotheses

Based on these hypotheses, we set out to design a series of new catalysts to specifically probe putative interactions. To facilitate catalyst development, a predictive model (Fig. 6) was generated for the entire substrate set with the aryl-substituted catalysts (**1b-1d** and **1f-1k**). This model contains 108 combinations (9 catalysts times 12 substrates) from the initial library of experiments, where half were used as a training set (Fig. 6B, black squares) and the other half as external validations (red crosses). New catalysts were proposed to address hypotheses raised by the focused models and trend analysis, and their enantioselectivity was predicted prior to synthesis by the comprehensive model (Fig. 6).

First, to probe whether the aryl substituent on the triazole ring plays a primarily steric role, rather than directly engaging the substrate in a π interaction, perfluorophenyl catalyst **1l** was designed and evaluated. Substituent local sterics and charge distribution have been shown to strongly affect non-covalent π interactions (28–30, 32–39). Therefore, we expected that if enantioselectivity were predominantly dependent on the aryl substituent directly engaging as a partner in a π interaction (as opposed to taking an ancillary role in π interactions involving the triazole), perfluorophenyl catalyst **1l** would deviate significantly from its Ad (**1e**), 2,6-(F)₂-Ph (**1k**) or 2,6-(MeO)₂-Ph (**1j**) counterparts. However, all four catalysts behave similarly with respect to the magnitude and sign of enantioselectivity (entries 1–9, 24–26,

Fig. 6C). This result is well predicted by the model (Fig. 6B) and is consistent with the hypothesis that the main function of the aryl substituent is steric.

Next, catalyst **1m**, bearing a single isopropyl group at the 2-position of the triazole aryl substituent was prepared to probe the hypothesis that steric repulsion exists between larger catalyst 2,6-substituents and elongated substrate 4-substituents. We anticipated that an isopropyl group would induce the torsion necessary to enforce the proposed benzyl-triazole π interaction, while avoiding a direct steric interaction between the substrate benzyl 4-position and the catalyst aryl 2,6-substituents. Indeed, for all 4-substituted substrates tested (entries 10–12, Fig. 6C), **1m** provided the corresponding product in higher enantioselectivity than **1c**, which bears isopropyl groups at both *ortho* positions of the triazole aryl substituent (entries 13–15, Fig. 6C). For 4-NO₂-Bn substrate **2d**, the 2-*i*Pr catalyst **1m** resulted in the highest enantioselectivity observed to date (entry 12, Fig. 6B–C), as predicted by the model.

Finally, in order to evaluate the capacity to obtain improved enantioselectivity as a result of a data-intensive approach, and the hypothesis that torsion leads to enhanced enantioselectivity for the 2-substituted substrates, several proposed catalysts were evaluated using the model in Fig. 6. Catalyst **1n** was selected as it is synthetically feasible, accommodates a torsion angle close to 90°, and was predicted to give improved enantioselectivity for all substrates bearing hydrogen at the 4-benzyl position. This prediction was verified in practice for all eight substrates evaluated, affording the highest enantioselectivities observed to date (Fig. 6C, entries 16–23). These results confirm that a perpendicular geometry of the triazole and the aryl ring can indeed lead to higher enantioselectivities, supporting the premise that the orientation of the triazole ring coupled with its R group's steric constraints control triazole π interactions.

The overall analysis of the triazole-PA case study demonstrates the complementary manner in which classical physical organic techniques and modern data analysis strategies can be merged toward a more complete mechanistic assessment (40). This approach is based on the use of empirical data, which is often a prerequisite for a rational reaction optimization process, to concomitantly conduct a mechanistic investigation. Information of this sort that could be used for an in-depth analysis is often omitted from reports in the field of catalysis, as only results leading to the desired outcomes are generally presented and pursued. Yet, as high-throughput (2, 4, 8), automated (7) methods for reaction development are now common, data analysis strategies could be applied in parallel to optimization procedures, allowing for simultaneous mechanistic and structural analysis.

Creatively collecting and organizing data to examine proposed hypotheses affords improved generalizations, particularly as datasets become larger and more complex (23). This idea holds true for the analysis of reaction trends by parameters that reflect structural modification. Indeed, the focused catalyst and substrate models, and their organization according to fundamental, quantitative, physical-organic trends, provided non-intuitive insights regarding interactions involved in enantioselectivity determination. Although this approach is general for the prediction and study of chemical reaction outcomes, this case study was chosen as a stringent benchmark because it employs weak, non-covalent interactions for the induction of enantioselectivity. These interactions are typically in the

energy range required to distinguish a highly enantioselective reaction from its racemic counterpart (2–3 kcal/mol) (12, 13, 32, 37), providing seemingly endless approaches to rational catalyst design. However, controlling non-covalent interactions represents a notable challenge in the design of catalytic systems due to multiple energetically accessible orientations (13). Complemented with rigorous experimental analysis, the disclosed data-intensive approach is suited to addressing such intricacies, and holds potential for the analysis of increasingly complicated catalytic systems streamlining both reaction and catalyst development.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the NSF (CHE-0749506 and CHE-1361296) and the National Institute of General Medical Sciences (R01 GM104534) for partial support of this work. The support and resources from the Center for High Performance Computing at the University of Utah are gratefully acknowledged. A. J. N. gratefully acknowledges an Amgen Fellowship in Organic Chemistry for funding and Jörg Hehn for early contributions to this work.

References and Notes

1. Collins KD, Gensch T, Glorius F. Contemporary screening approaches to reaction discovery and development. *Nat Chem*. 2014; 6:859–871. [PubMed: 25242480]
2. Reetz MT. New methods for the high-throughput screening of enantioselective catalysts and biocatalysts. *Angew Chem Int Ed Engl*. 2002; 41:1335–1338. [PubMed: 19750756]
3. Robbins DW, Hartwig JF. A simple, multidimensional approach to high-throughput discovery of catalytic reactions. *Science*. 2011; 333:1423–1427. [PubMed: 21903809]
4. McNally A, Prier CK, MacMillan DWC. Discovery of an α -Amino C-H Arylation Reaction Using the Strategy of Accelerated Serendipity. *Science*. 2011; 334:1114–1117. [PubMed: 22116882]
5. Schmink JR, Bellomo A, Berritt S. Scientist-Led High-Throughput Experimentation (HTE) and Its Utility in Academia and Industry. *Aldrichimica Acta*. 2013; 46:71–80.
6. Friedfeld MR, et al. Cobalt Precursors for High-Throughput Discovery of Base Metal Asymmetric Alkene Hydrogenation Catalysts. *Science*. 2013; 342:1076–1080. [PubMed: 24288328]
7. Heider PL, et al. Development of a Multi-Step Synthesis and Workup Sequence for an Integrated, Continuous Manufacturing Process of a Pharmaceutical. *Org Process Res Dev*. 2014; 18:402–409.
8. Metola P, Nichols SM, Kahr B, Anslyn EV. Well plate circular dichroism reader for the rapid determination of enantiomeric excess. *Chem Sci*. 2014; 5:4278–4282. [PubMed: 25386332]
9. Santanilla AB, Regalado EL, Pereira T, Shevlin M. Nanomole-scale high-throughput chemistry for the synthesis of complex molecules. *Science*. 2015
10. Curtin DY. Stereochemical Control of Organic Reactions Differences in Behaviour of Diastereoisomers. *Rec Chem Prog*. 1954; 15:110–128.
11. Halpern J. Mechanism and stereoselectivity of asymmetric hydrogenation. *Science*. 1982; 217:401–407. [PubMed: 17782965]
12. Anslyn, EV.; Dougherty, DA. *Modern Physical Organic Chemistry*. University-Science Books; 2006.
13. Knowles RR, Jacobsen EN. Attractive noncovalent interactions in asymmetric catalysis: Links between enzymes and small molecule catalysts. *Proc Nat Acad Sci*. 2010; 107:20678–20685. [PubMed: 20956302]
14. Harper KC, Sigman MS. Predicting and optimizing asymmetric catalyst performance using the principles of experimental design and steric parameters. *Proc Nat Acad Sci*. 2011; 108:2179–2183. [PubMed: 21262844]

15. Neel AJ, Hehn JP, Tripet PF, Toste FD. Asymmetric Cross-Dehydrogenative Coupling Enabled by the Design and Application of Chiral Triazole-Containing Phosphoric Acids. *J Am Chem Soc.* 2013; 135:14044–14047. [PubMed: 24025163]
16. Bobbitt, JM.; Brückner, C.; Merboudh, N. *Organic Reactions*. John Wiley & Sons, Inc; 2004.
17. Li CJ. Cross-Dehydrogenative Coupling (CDC): Exploring C–C Bond Formations beyond Functional Group Transformations. *Acc Chem Res.* 2009; 42:335–344. [PubMed: 19220064]
18. Yeung CS, Dong VM. Catalytic Dehydrogenative Cross-Coupling: Forming Carbon–Carbon Bonds by Oxidizing Two Carbon–Hydrogen Bonds. *Chem Rev.* 2011; 111:1215–1292. [PubMed: 21391561]
19. García Mancheño O, Stopka T. TEMPO Derivatives as Alternative Mild Oxidants in Carbon–Carbon Coupling Reactions. *Synthesis.* 2013; 45:1602–1611.
20. Akiyama T. Stronger Brønsted Acids. *Chem Rev.* 2007; 107:5744–5758. [PubMed: 17983247]
21. Terada M. Chiral Phosphoric Acids as Versatile Catalysts for Enantioselective Transformations. *Synthesis.* 2010; 2010:1929–1982.
22. Parmar D, Sugiono E, Raja S, Rueping M. Complete Field Guide to Asymmetric BINOL-Phosphate Derived Brønsted Acid and Metal Catalysis: History and Classification by Mode of Activation; Brønsted Acidity, Hydrogen Bonding, Ion Pairing, and Metal Phosphates. *Chem Rev.* 2014
23. Silver, N. *The signal and the noise: why so many predictions fail--but some don't*. Penguin Press; New York: 2012.
24. Verloop A, Tipker J. A comparative study of new steric parameters in drug design. *Pharmacochem Libr.* 1977; 2:63–81.
25. Milo A, Bess EN, Sigman MS. Interrogating selectivity in catalysis using molecular vibrations. *Nature.* 2014; 507:210–214. [PubMed: 24622199]
26. Harper KC, Bess EN, Sigman MS. Multidimensional steric parameters in the analysis of asymmetric catalytic reactions. *Nat Chem.* 2012; 4:366–374. [PubMed: 22522256]
27. Lima, M. *Visual complexity. Mapping patterns of information*. Princeton Architectural Press; New York: 2013.
28. Hunter CA, Sanders JKM. The nature of π - π interactions. *J Am Chem Soc.* 1990; 112:5525–5534.
29. Cozzi F, Cinquini M, Annunziata R, Dwyer T, Siegel JS. Polar/ π interactions between stacked aryls in 1,8-diarylnaphthalenes. *J Am Chem Soc.* 1992; 114:5729–5733.
30. Martinez CR, Iverson BL. Rethinking the term “ π -stacking”. *Chem Sci.* 2012; 3:2191.
31. Krenke EH, Houk KN. Aromatic Interactions as Control Elements in Stereoselective Organic Reactions. *Acc Chem Res.* 2013; 46:979–989. [PubMed: 22827883]
32. Riley KE, Hobza P. On the Importance and Origin of Aromatic Interactions in Chemistry and Biodisciplines. *Acc Chem Res.* 2013; 46:927–936. [PubMed: 22872015]
33. Sherrill CD. Energy Component Analysis of π Interactions. *Acc Chem Res.* 2013; 46:1020–1028. [PubMed: 23020662]
34. Parrish RM, Sherrill CD. Quantum-Mechanical Evaluation of π - π versus Substituent- π Interactions in π Stacking: Direct Evidence for the Wheeler–Houk Picture. *J Am Chem Soc.* 2014:17386–17389. [PubMed: 25423285]
35. Hohenstein EG, Sherrill CD. Effects of Heteroatoms on Aromatic π - π Interactions: Benzene–Pyridine and Pyridine Dimer. *J Phys Chem A.* 2009; 113:878–886. [PubMed: 19132847]
36. Wheeler SE. Understanding Substituent Effects in Noncovalent Interactions Involving Aromatic Rings. *Acc Chem Res.* 2013; 46:1029–1038. [PubMed: 22725832]
37. Shchegolev BF, McKee ML, Zhuravlev AV, Savvateeva-Popova EV. Analysis of intermolecular interaction energy inputs in benzene-imidazole and imidazole-imidazole systems in parallel displaced and T-configuration. *Biophys.* 2013; 58:355–360.
38. Wheeler SE, Houk KN. Substituent Effects in the Benzene Dimer are Due to Direct Interactions of the Substituents with the Unsubstituted Benzene. *J Am Chem Soc.* 2008; 130:10854–10855. [PubMed: 18652453]
39. O'Hagan D. Understanding organofluorine chemistry. An introduction to the C–F bond. *Chem Soc Rev.* 2008; 37:308. [PubMed: 18197347]

40. Physical Organic Chemistry for the 21st Century. Pure Appl Chem. 1997; 69:1–76.
41. Frisch, MJ., et al. Gaussian, Inc; Wallingford, CT: 2009.
42. Rein K, et al. Chiral dipole-stabilized anions: experiment and theory in benzylic and allylic systems. Stereoselective deprotonations, pyramidal inversions, and stereoselective alkylations of lithiated (tetrahydroisoquinolyl) oxazolines. J Am Chem Soc. 1989; 111:2211–2217.
43. Goddard-Borger ED, Stick RV. An Efficient, Inexpensive, and Shelf-Stable Diazotransfer Reagent: Imidazole-1-sulfonyl Azide Hydrochloride. Org Lett. 2007; 9:3797–3800. [PubMed: 17713918]
44. Wang YM, Wu J, Hoong C, Rauniyar V, Toste FD. Enantioselective Halocyclization Using Reagents Tailored for Chiral Anion Phase-Transfer Catalysis. J Am Chem Soc. 2012; 134:12928–12931. [PubMed: 22830953]
45. Quinn, JA. Norgwyn Montgomery Software Inc; North Wales, PA:
46. MATLAB. The MathWorks Inc; Natick, MA:
47. Rauniyar V, Lackner AD, Hamilton GL, Toste FD. Asymmetric Electrophilic Fluorination Using an Anionic Chiral Phase-Transfer Catalyst. Science. 2011; 334:1681–1684. [PubMed: 22194571]
48. Merrick JP, Moran D, Radom L. An Evaluation of Harmonic Vibrational Frequency Scale Factors. J Phys Chem A. 2007; 111:11683–11700. [PubMed: 17948971]
49. Goodman S. A Dirty Dozen: Twelve P-Value Misconceptions. Seminars in Hematology. 2008; 45:135–140. [PubMed: 18582619]
50. Zhao Y, Truhlar DG. 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. 2007; 120:215–241.
51. Schäfer A, Horn H, Ahlrichs R. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. J Chem Phys. 1992; 97:2571.
52. 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. 1994; 100:5829.
53. Valero R, Gomes JRB, Truhlar DG, 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. 2008; 129:124710. [PubMed: 19045051]

Fig. 1. System under study: asymmetric C-N dehydrogenative coupling

(A) BINOL-based phosphoric acid scaffold, enantioselective cross dehydrogenative coupling reaction scheme, and nitrogen-deletion experiment. (B) Proposed mechanism involving a chiral phosphate-substrate ion pair. (C) Enantiomeric excess (ee) values obtained using various substrate/catalyst combinations. (D) Library design and parameter identification strategy.

Fig. 2. Kinetic isotope effect studies and mechanistic implications

(A) Considerations regarding the origin of enantioselectivity. (B) KIEs of deuterated enantiomerically enriched substrate **2a-d₁** with (*R*) and (*S*) PA catalysts: adamantyl-substituted triazolyl (**1e**) and pyrazoyl (*pyr*-**1e**), and TCYP (**6**). (C) Revised mechanism of enantioselectivity determination.

Fig. 3. Impact of heterocyclic catalyst substituent on enantioselectivity

(A) Predicted and measured enantioselectivities for various substrates with adamantyl-substituted triazolyl (**1e**), pyrazoyl (**pyr-1e**), and imidazolyl (**imid-1e**) PAs. Values identified with a^a are external validations. (B) Mathematical correlation of normalized catalyst and substrate vibrational parameters to enantioselectivity. (C) Predicted vs. measured $33G^\ddagger$ plot. (D) Illustration of the structural features implicated by the identified parameters.

Fig. 4. Graphical representation of catalyst structure-selectivity trends as a function of substrate. Enantioselectivity trends for every catalyst against all substrates (each trend line represents a catalyst and each x axis tick-line a substrate).

Fig. 5. Graphical representation of substrate structure-selectivity trends as a function of catalyst. Enantioselectivity trends for every substrate against all catalysts (each trend line represents a substrate and each x axis tick-line a catalyst).

Fig. 6. Validation of mechanistic hypotheses through directed catalyst design

(A) Normalized equation for the prediction of enantioselectivity using parameters that describe catalyst and substrate structural features. (B) Predicted vs. measured $33G^\ddagger$ plot. (C) Hypothesis-driven external predictions and their comparison to other relevant catalysts.