

UC Berkeley

UC Berkeley Previously Published Works

Title

Stereoselective Generalizations over Diverse Sets of Chiral Acids Enabled by Buried Volume

Permalink

<https://escholarship.org/uc/item/1v3584mk>

Journal

Journal of the American Chemical Society, 148(2)

ISSN

0002-7863

Authors

Smith, AndrewL

Toste, F Dean

Publication Date

2026-01-21

DOI

10.1021/jacs.5c20342

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Stereoselective generalizations over diverse sets of chiral acids enabled by buried volume

Andrew L. Smith¹, F. Dean Toste^{1,2*}

¹Department of Chemistry, University of California, Berkeley, California 94720, United States

²Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States

asymmetric catalysis, organocatalysis, data science, molecular descriptors

ABSTRACT: Stereoselective catalysis is a paradigm of complexity, navigating arrays of specific, yet weak non-covalent interactions to achieve asymmetric induction. Accordingly, structure-activity relationships modeling stereoselectivity have required bespoke representations, high dimensionality, and/or large datasets to furnish productive trends, hampering their subsequent interpretability and utility. Here, we report active site-based buried volume is a uniquely advantageous descriptor for the construction of models to stereoselectivity by Brønsted acid organocatalysts. Through statistical analyses of 50+ datasets across nearly 200 distinct catalysts, we realize active site-based buried volume is an exceptional representation, accommodating both extensive structural diversity (within datasets) and functional diversity (across datasets). We show the descriptor's value likely relates to its surprising capacity to generalize beyond simple steric interactions and account for diverse stereoelectronic effects consequential to reported stereoselectivity. As such, we anticipate future modeling of stereoselectivity by Brønsted acid are greatly simplified via the concise representation – active site-based buried volume.

Introduction

In chemistry, *structure informs function* is a paradigm invoked to justify the emergence of structure-activity relationships (SARs). In theory, chemical properties are paired with structural changes made, and inferences are drawn to rationalize the emergent patterns with reaction outcomes. To build quantitative relationships, *molecular representations* are used as surrogates for chemical properties (**Fig. 1A**).¹ While many traditional SARs are constructed with *representations* rooted in specific experimental phenomena (ex. Hammett,² Taft,³ Charton⁴), such *representations* display narrow *domains of applicability*⁵ – culminating in critical *discontinuity failures* where certain molecules of interest cannot be described by the model altogether. To address this, modern SARs are typically constructed from calculated representations (ex. sterimol,⁶ NBO partial charge,⁷ etc.) that are intrinsically more generalizable.^{8–10} But, as *molecular representations* become more abstract, their subsequent interpretations become increasingly unintuitive, limiting the production of useful and *general* chemical rationale.^{11,12}

The construction of *intuitive* SARs is essential for understanding complex chemical phenomena. In

synthetic chemistry, stereoselective catalysis is a paradigm of complexity. Here, a chiral catalyst specifically engages in an array of weak diastereoselective interactions to kinetically steer reactive intermediates towards one stereoisomer over another (**Fig. 1B**).^{1,13,14} Notably, Brønsted acid organocatalysts (BAOCs) are one versatile tool popularized over the past two decades, enabling a wide array of applications, including: Brønsted acid/base catalysis,^{15,16} cooperative transition metal catalysis,¹⁷ and counter-anion directed catalysis,¹⁸ among others (**Fig. 1C**). Their functional diversity arises from the pronounced structural modularity of BAOc scaffolds (**Fig. 1D**).^{19,20} To navigate this structural and functional dimensionality, synthetic chemists have made considerable progress engineering stereoselective BAOcs from two criteria – acidity^{15,21} and confinement.^{22,23} These properties are routinely *qualitatively* invoked, but their quantitative *molecular representations* have lacked general coherence and interpretability.

Seminal reports by Cavallo,^{24,25} Denmark,²⁶ List,²⁷ Reid,²⁸ and Sigman²⁹ have demonstrated calculated *representations* of confinement by BAOcs can help to construct SARs for stereoselectivity. However, these models have relied on unintuitive, bespoke

representations for confinement, are complicated by excessive dimensionality (ex. too many other descriptors in the model), or are evaluated on only a very narrow set(s) of reactions. Although confinement reflects a powerful *qualitative* criterion for the construction of stereoselective BAOCs, a precise *molecular representation*, relevant across the breadth of their utility, has yet to be realized.

We herein report a powerful representation for the stereoselectivity by BAOCs – *active site-based* buried volume – emerging from statistical analyses of over 50+ datasets (**Fig. 1E**). Its superior performance was benchmarked against that of many familiar and meaningful molecular descriptors, highlighting its unique ability to correlate to stereoselectivity across many datasets, regardless of the underlying structural and functional diversity of the BAOCs considered. We show that the representation, while traditionally reflecting *confinement* – a steric phenomenon – likely generalizes across an array of stereoelectronic non-covalent interactions. As such, we assert the descriptor is a valuable representation for BAOCs in a broad range of contexts for future modeling endeavors.

Results and Discussion

Active site-based buried volume emerges as an advantageous representation

Stereoselective catalysis relies on an array of weak, but specific, non-covalent interactions (NCIs) between a chiral catalyst and a substrate. Accordingly, a *robust* representation for stereoselective catalysts necessarily (1) accommodates structural diversity of the deployed chiral catalysts, (2) generalizes across the accessible functional manifolds, and (3) mediates mechanistic differences observed from each transformation.

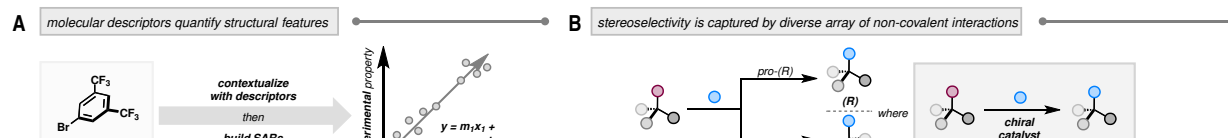
In the case of BAOCs, a range of molecular descriptors have emerged as promising candidates, as reported in SARs by Cavallo,²⁵ Denmark, Reid,²⁸ Sigman,⁹ and our own group.²⁹ Qualitatively, these descriptors aim to reflect *confinement* of the BAOC's active site (*vide supra*) through various geometric and steric representations, including: bond length, angle, sterimol,³⁰ and buried volume²⁴ (**Fig. 2A**). Beyond representations of *confinement*, many reported models also invoke stereoelectronic contributions to stereoselectivity from adjacent, modular substituents. We account for these expectations by considering an

array of *substituent-based* features calculated from the analogous *aryl bromide* as a surrogate (**Fig. 2A**).³¹ Finally, as the *acidity* of BAOCs is often invoked to contextualize the differences across a series of chiral acids, we include calculated pK_a to reflect this expectation (see **Molecular Descriptor Calculation Workflow** for full details; see **dataset_active-site.csv** and **dataset_substituent.csv** for full list of computed descriptors used in this work).³²

Equipped with an array of candidates for general representations of BAOCs, we then collated a set of 50+ reaction datasets from the literature, culminating in nearly 200 unique BAOCs modeled (see **dataset_rxn.xlsx** for full list of curated datasets). To ensure the impact of any representation considered is *clearly* detected, univariate linear regression models trained on each reaction dataset serve as the basis for analysis, avoiding data-intensive, high-dimensional models (ex. neural networks, random forest, multivariate linear regression, etc.) that readily overfit on smaller datasets and become increasingly challenging to interpret (avg. dataset size in *this work* – 11).^{12,33} To compare the overall performance of each representation, we evaluated regression *accuracy* (R^2 , MAE – mean absolute error) and *robustness* (p -value, LOO R^2 – lease-one-out cross-validation) across each dataset.³⁴ For representations spanning multiple atoms, dimensions, and ensembles (ex. partial charge at different atoms, buried volume across different radii), only the most *robust* regression model (largest LOO R^2 across the set) for a given dataset was considered in further analyses (see **dataset_scores.xlsx** for complete statistical analyses).

Across the 50+ reaction datasets analyzed, *active site-based* buried volume uniquely emerged as the most *accurate* and *robust* representation (67% of all models are *statistically meaningful* with >95% confidence, 48% with 99% confidence, avg. R^2 = 0.57, avg. MAE = 0.21; **Fig. 2B-C**). At the strictest confidence shown, the average descriptor only constructed meaningful univariate models for one fifth of the datasets (avg. 22% with >99% confidence). Surprisingly, other active site-based descriptors like sterimol L and BMin, which were expected to convey *confinement* like *active site-based* buried volume, are among the last-ranked descriptors with very sparse utility (avg. 9% with >99% confidence). Instead, the next highest ranked descriptor is *substituent-based* buried volume (39% with >99% confidence; avg. R^2 = 0.56, avg. MAE = 0.22). The descriptor has been

2



deployed to represent *confinement* akin to its *active site-based* analog but apparently cannot correlate to enantioselectivity as generally across the datasets (**Fig. 2B**; see **Detailed Statistical Analyses**; see **Supplementary Fig. 21-74** for a complete list of univariate regression models by *active site-based* buried volume).

Active site-based buried volume accommodates structural heterogeneity

Buried volume captures the largest proportion of *meaningful* univariate models relative to other descriptors, but its position – either *active site-based* or *substituent-based* – influences its utility. The juxtaposition between the two suggests the better-performing *active site-based* representation conveys salient structural changes in a more general manner. It was anticipated that *active site-based* buried volume will advantageously correlate to stereoselectivity when the reaction datasets are constructed from structurally diverse sets of BAOCs, while *substituent-based* buried volume should generate robust correlations when the datasets are dominated by changes to the modular substituent. So, we organized the datasets by their corresponding structural diversity (Fig. 3A; see **Detailed Statistical Analyses** for classification criteria). 27 of the 54 reactions

remaining 27 reactions are classified as *significant diversity*.

Over the *limited diversity* datasets, *active site-based* and *substituent-based* buried volumes both perform well. As these datasets are populated with BAOCs varying primarily at the modular substituent, the *substituent-based* representation more precisely reflects the chemical changes (avg. $R^2 = 0.65$, avg. MAE = 0.19 kcal/mol; Fig. 3B), while the *active site-based* representation performs at a similar caliber (avg. $R^2 = 0.64$, avg. MAE = 0.19 kcal/mol). On the other hand, across *significant diversity* datasets, *active site-based* buried volume outperforms its *substituent-based* analog. The heterogeneity of the datasets requires the descriptors to quantitatively reflect structural changes across the acids. As such, the utility of *substituent-based* buried volume substantially decreases (avg. $R^2 = 0.47$, avg. MAE = 0.25 kcal/mol; Fig. 3C), while *active*

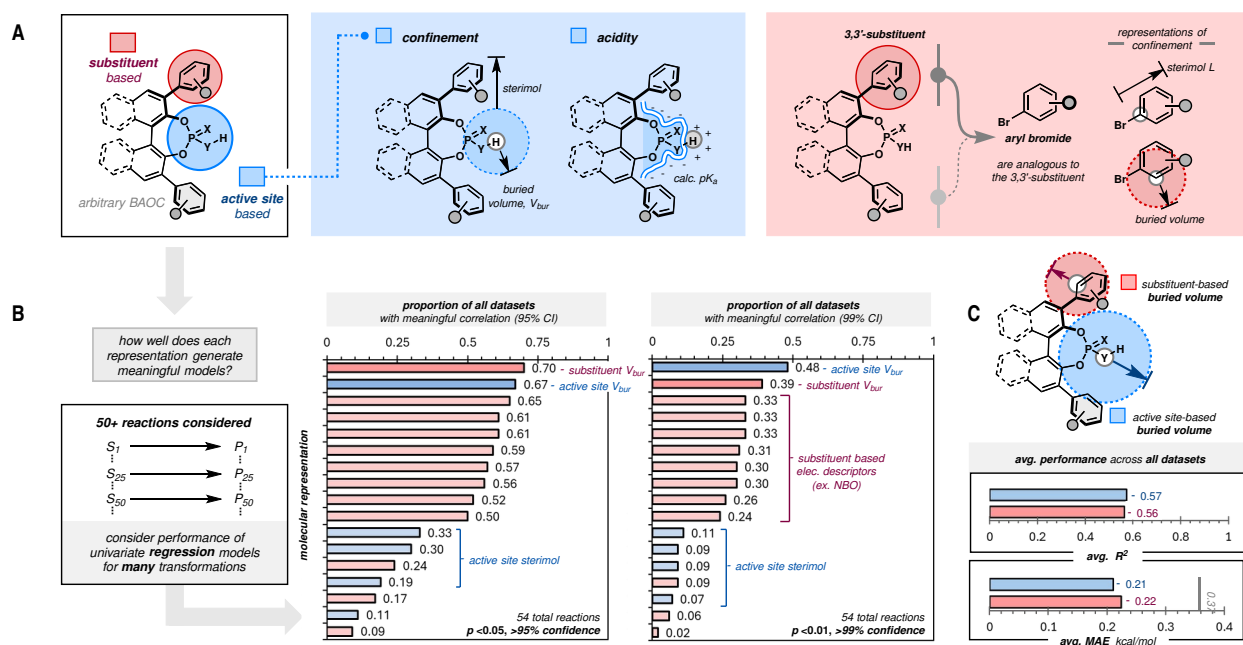


Figure 2. Collation of candidate molecular descriptors and detection of a general representation for stereoselectivity. (a) *Active site-based* representations reflect existing synthetic design criteria (ex. acidity, confinement) without specific attention to underlying scaffold, while *substituent-based* representations are derived from simplified analogs (aryl bromides). (b) Proportion of meaningful univariate regression models constructed by each descriptor at increasingly strict confidence intervals (up to 99% confidence) across the set of 50+ reaction datasets considered in this work. (c) On average, *active site-based* and *substituent-based* buried volumes construct the most *accurate* and *robust* univariate regression models (in comparison, dummy predictor of *mean* for each dataset furnishes an avg. MAE = 0.37 kcal/mol).

considered can be classified as *limited diversity* – essentially, the dataset has one or fewer examples of catalysts differing beyond the modular substituent (ex. different chiral backbones and/or active sites). The

site-based buried volume's is relatively sustained (avg. $R^2 = 0.51$, avg. MAE = 0.23 kcal/mol).

To highlight the impact structural heterogeneity, we showcase the utility of *active site-based* and *substituent-based* buried volume in two reaction datasets. R-X, a seminal dataset reported by Zahrt, et al.²⁶ describes the stereoselective addition of a thiol to an imine (**Fig. 3D**). In this work, we model the average performance of a given BAOC across the 16 combinations of substrates reported by Zahrt, et al. Across the BAOCs considered in the dataset, only two examples are constructed from distinct chiral backbones (13% of the dataset). Accordingly, as most structural changes are concentrated at the modular substituent, *substituent-based* buried volume perfectly recapitulates the reported stereoselectivity ($R^2 = 0.98$, MAE = 0.04 kcal/mol). Even though *active site-based* buried volume cannot reflect the chemical changes with such precision, a useful univariate regression is still observed ($R^2 = 0.70$, MAE = 0.20 kcal/mol). Notably, other representations of *confinement*, such as sterimol, which have been included in multivariate SARs modeling this dataset fail to produce univariate relationships (ex. with *active site-based* sterimol_L $R^2 = 0.09$, MAE = 0.33 kcal/mol).

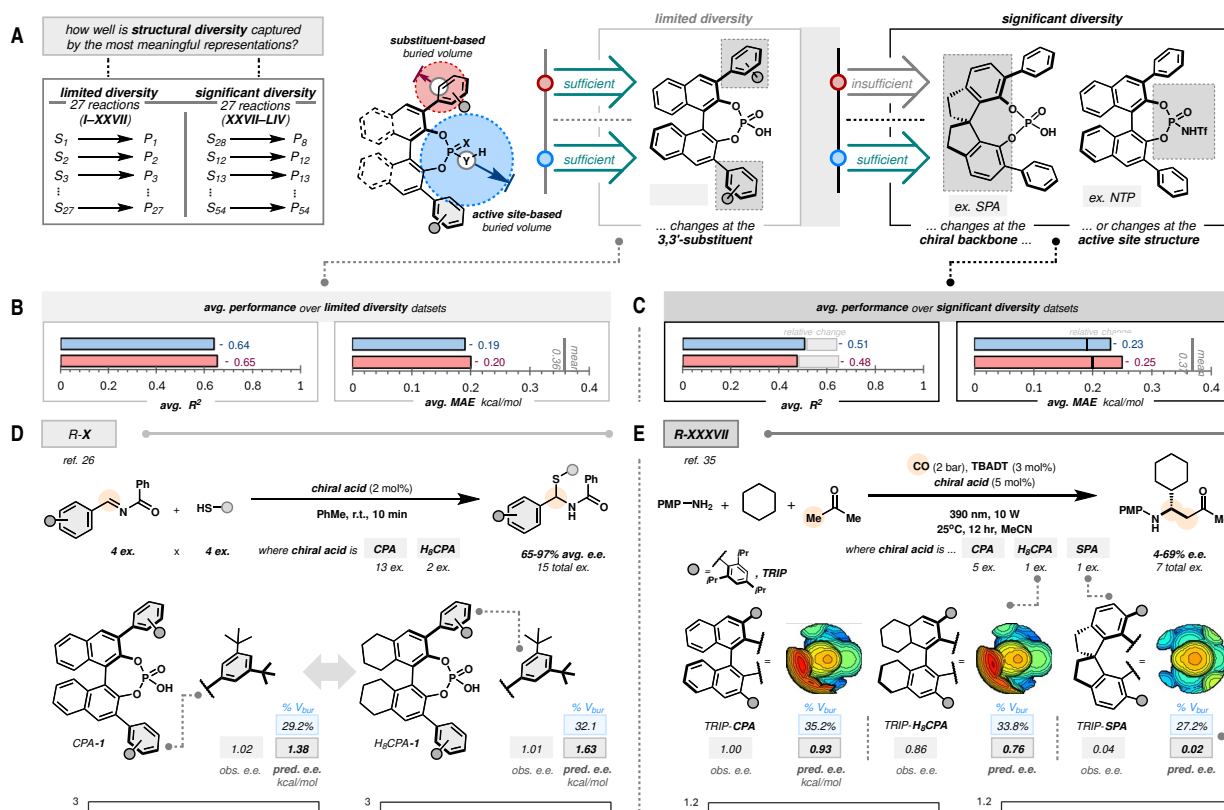
Unlike the homogeneous dataset of R-X, the dataset R-XXXVII³⁵ describes the application of a small, but diverse array of BAOCs in a multicomponent photochemical transformation (**Fig. 3E**). Critically, it was observed in the original report that choice of chiral backbone significantly altered stereoselectivity (ref. TRIP-CPA, TRIP-H₂CPA, TRIP-SPA in **Fig. 3E**). The trend drawn by *active site-based* buried volume seamlessly captures the enantioinduction cooperatively afforded by the chiral backbone and the

modular substituents ($R^2 = 0.74$, MAE = 0.13 kcal/mol). Of course, *substituent-based* buried volume cannot reflect such structural changes over the dataset, so it fails to correlate ($R^2 = 0.15$, MAE = 0.24 kcal/mol; **Supplementary Fig. 7-9**).

Active site-based buried volume retains utility across catalytic manifolds

The success of *active site-based* buried volume across *significant diversity* datasets supports the expectation that the descriptor will convey the collective consequences of each individual structural feature on stereoselection. However, it is unclear whether such performance is also evidence of the descriptor's sustained utility across all BAOC reactivity modes. Traditionally, BAOCs were deployed in acid-catalyzed additions to π -bonds (ex. imine additions).¹⁶ Yet, modern applications of BAOCs have expanded their functional arsenal beyond additions to include stereoselective eliminations and substitutions under a range of conditions (ex. thermal, phase-transfer,³⁶ and photochemical). If *active site-based* buried volume is truly general, then its utility must be demonstrated across the array of reaction manifolds accessible to BAOCs.

To investigate this, we partitioned the 50+ reaction datasets into three distinct classes corresponding to catalytic activity: (1) acid-mediated additions, (2) acid-mediated substitutions and eliminations, (3) and complex, multicomponent paradigms (**Fig. 4A**). Indeed, across each of these categories, *active site-based* buried volume furnishes *accurate* and *robust* regressions (*addition* – avg. $R^2 = 0.56$, *elimination and substitution* – avg. $R^2 = 0.63$, and *complex* – avg. $R^2 =$



0.52; **Fig. 4B**). Other alternative representations of confinement, like sterimol, which have previously reported utility for specific acid-mediated *additions*, fail to accommodate the remaining functional diversity available to the BAOCs (ex. with sterimol_L *addition* – avg. $R^2 = 0.23$, *elimination and substitution* – avg. $R^2 = 0.28$, and *complex* – avg. $R^2 = 0.21$; **Supplementary Fig. 10-13** for full analyses to other representations of *confinement*).

To highlight the consistent utility of *active site-based* buried volume across reaction manifolds, we showcase one reaction from each category enumerated above. While many acid-mediated *additions* rely on basic functional groups (ex. imines) to strongly bind to the chiral acid active site, stereoselective solutions have emerged for substrates lacking such motifs. One early example of this is R-**XV**, a stereoselective aldol-like addition between an enamide and a glyoxal reported by Terada and coworkers (**Fig. 4C**).³⁷ The report realized that traditional stereoselective intuition – *increasing* confinement – does not furnish the desired adduct with selectivity (ref. CPA-2 with 8% *e.e.*). Instead, reducing confinement at the active site was required to achieve the desired alcohol with high selectivity (ref. CPA-3 with 98% *e.e.*). *Active site-based* buried volume reflects Terada's counterintuitive findings, furnishing a useful univariate regression model ($R^2 = 0.81$, MAE = 0.35 kcal/mol).

Beyond acid-mediated *additions*, many stereoselective eliminations (ex. atroposelective condensations) and substitutions mediated by BAOCs have been reported. Here, we highlight R-**XLVII**, an intermolecular substitution of azetidine with thiol nucleophiles (**Fig. 4D**).³⁸ The authors screen a small (7 ex.) but diverse (3 distinct scaffolds) set of BAOCs to realize a solution (up to 91% *e.e.*). *Active site-based* buried volume not only accommodates the structural heterogeneity of the dataset, but also the BAOC-mediated substitution, yielding an excellent linear model ($R^2 = 0.91$, MAE = 0.20 kcal/mol).

Finally, though BAOCs are typically employed as *chiral acids*, their applications have evolved beyond the scope of simple acid/base chemistry to include applications as ligands in asymmetric transition metal catalysis and chiral counterions (ex. asymmetric counterion directed catalysis, ACDC). For years, our group has investigated stereoselective phase-transfer oxidations mediated by chiral phosphates. One

notable example from our group describes a phase-transfer catalyzed Pummerer-type cyclization that had required remote structural changes to enhance the observed stereoselectivity (R-**XIV**, **Fig. 4E**)³⁹. In our original report, a multivariate regression from a collection of sterimol descriptors was required to reflect the impact of the distant structural changes to stereoselectivity (three terms, $R^2 = 0.96$). Here, we show *active site-based* buried volume alone readily recapitulates the relationship (one term, $R^2 = 0.85$, MAE = 0.06, **Fig. 4E**). In comparison, *active site-based* sterimol features are unsuitable for univariate relationships (ex. with sterimol L alone, $R^2 = 0.17$, MAE = 0.13, **Fig. 4E**).

Active site-based buried volume holistically generalizes stereoelectronic effects

Across the 50+ reaction datasets considered herein, *active site-based* buried volume uniquely reflects *confinement* as a quantitative representation for stereoselectivity by BAOCs, regardless of the underlying structure and function. Yet, *confinement* is often considered a *steric* phenomenon, where models for stereoselectivity invoke steric repulsion as the primary destabilizing criterion. More recently, chemists prefer to invoke attractive non-covalent interactions as a primary design criterion for

stereoselectivity. From this perspective, *confinement* qualitatively reflects a propensity for attractive dispersion forces.^{22,40,41} Still, a preponderance of other stereoelectronic non-covalent interactions (NCIs) have proven relevant for stereoselective catalysis (ex. π - π , cation- π , etc., **Fig. 5A**).^{13,14,42,43} Is *active site-based* buried volume capable of reflecting such diverse stereoelectronic NCIs?

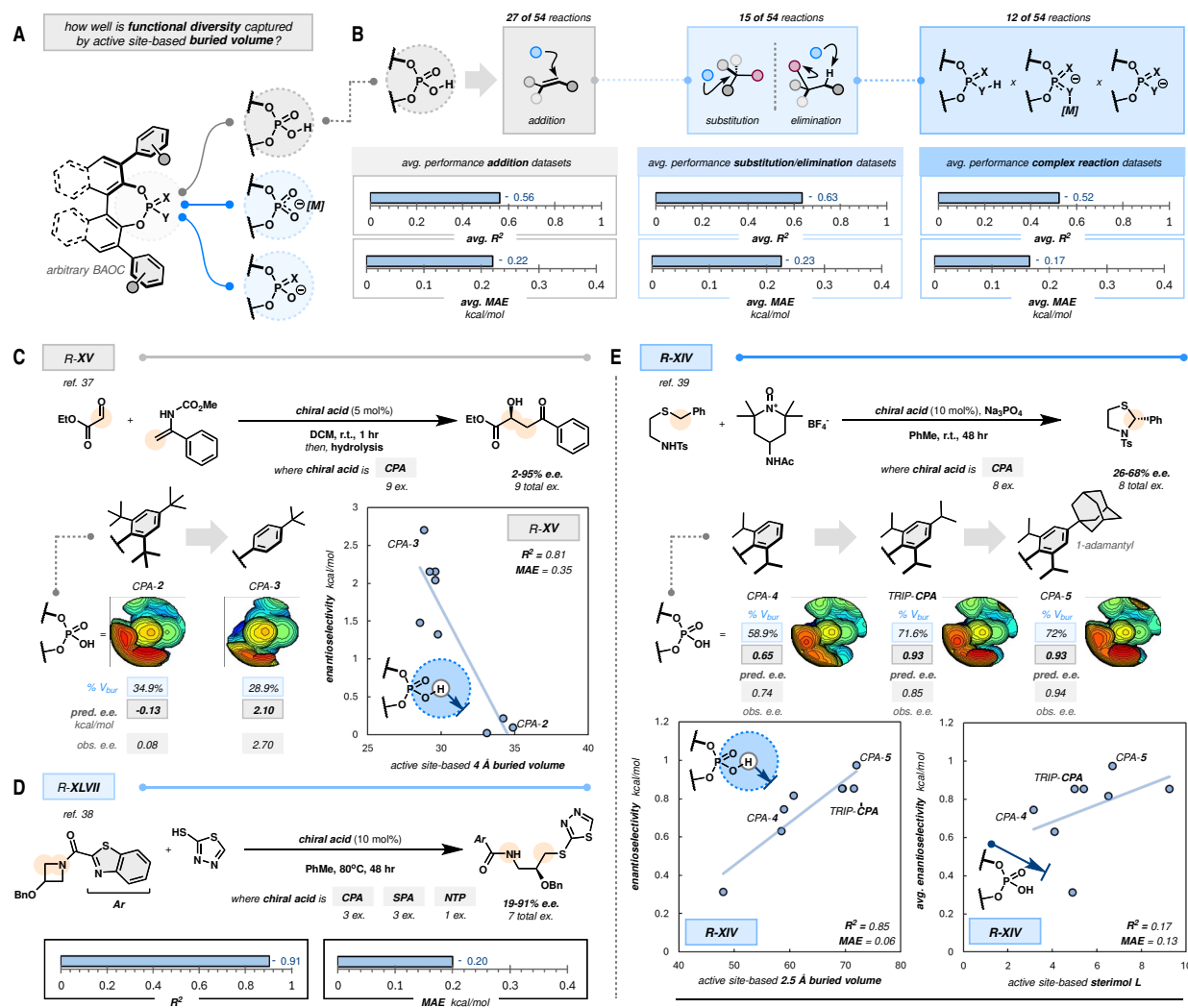


Figure 4. Active site-based buried volume retains utility across BAOC reaction manifolds. (a) BAOC reactivity can be organized into three categories reflecting the role of the acid, where (b) *active site-based* buried volume continues to furnish *accurate* trends across each category. (c) R-XV by Terada, et al.³⁷ highlights how *active site-based* buried volume can reflect counterintuitive relationships for stereoselective additions under confinement. (d) R-XLVII by Wang, et al.³⁸ highlights how *active site-based* buried volume can accommodate substantial structural diversity, even within a stereoselective substitution. (e) R-XIV by Toste, et al.³⁹ highlights how *active site-based* buried volume serves as a privileged representation for stereoselectivity by BAOCs, even within complex, multicomponent transformations like phase-transfer catalysis.

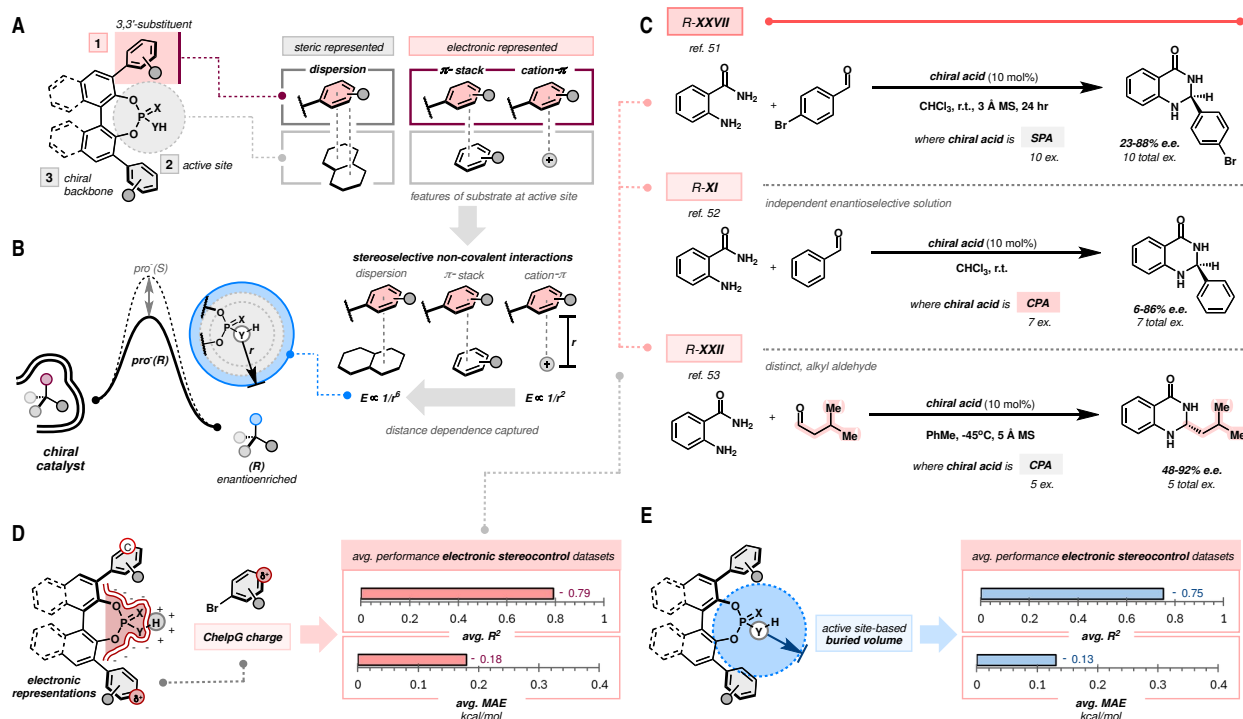


Figure 5. Active site-based buried volume holistically generalizes stereoelectronic effects across reaction datasets considered. (a) Stereoselective NCIs can be categorized as predominantly steric or electronic-mediated. (b) Interaction energies of various stereoselective NCIs relate to distance (denoted r) which is reflected by active site-based buried volume. (c) R-XXVII is a dataset reported by Wheeler, et al.⁵¹ to operate via electronic stereocontrol. R-XI⁵² and R-XXII⁵³ transformations likely operate similarly by analogy to R-XXVII. (d) Electronic representations like substituent-based ChelpG charge furnish excellent trends (averaged across R-XXVII, R-XI, and R-XXII). (e) Active site-based buried volume also furnishes excellent univariate trends (averaged across R-XXVII, R-XI, and R-XXII).

In theory, electronic contributions to NCIs (ex. cation- π or π - π interactions) are defined from the interaction of charges across a distance (ex. ion-dipole scales with $1/r^2$ where r represents distance between charges, van der Waals attraction scales with $1/r^6$, Fig. 5B).^{1,44} It follows that active site-based buried volume, which intimately relates percent steric occupancy across a specific distance (lit. volume), should formally capture such spatial interactions through this relationship. Additionally, it is generally observed that larger π -systems (readily reflected by buried volume) access stronger electrostatic NCIs (ex. cation- π) as a consequence of their greater polarizability.⁴⁴⁻⁴⁷ In fact, our group has previously shown robust, univariate relationships exist between buried volume and the strength of corresponding, stereoselective cation- π NCIs in some putative, stereo-determining reaction complexes.^{48,49} As highlighted in R-XV above, active site-based buried volume reflects stereoselectivity for a phase-transfer catalyst where stereoinduction was attributed to strengthening C-H $\cdots\pi$ interactions in the

major transition state.³⁹ Thus, across the datasets considered, active site-based buried volume likely reflects an array of stereoselective interactions by BAOCs, including those that are principally electronic in nature.

To ensure that such examples are not merely instances of cooperative steric and electronic control, we turned to the literature to identify BAOC-mediated transformations where stereoselectivity was explicitly realized via electronic NCIs. Indeed, elucidated with extensive computational work, Wheeler and coworkers have realized such electronic control was observed in dataset R-XXVII, a stereoselective condensation furnishing cyclic aminals, whereby biased coulombic interactions across the highly polarized active site drive stereoselectivity (Fig. 5C).^{50,51} Notably, this condensation is analogous to two other condensations analyzed herein, R-XI and R-XXII. As expected, substituent-based electronic representations like ChelpG charge furnish

meaningful univariate trends across the collection of transformation operating via electronic control (up to $R^2 = 0.79$, MAE = 0.18 kcal/mol; **Fig. 5D**, see **Supplementary Fig. 15-20** for remaining electronic representations). *Active site-based* buried volume similarly recapitulates these trends ($R^2 = 0.75$, MAE = 0.13 kcal/mol; **Fig. 5E**), supporting the expectation that it can reflect principally electronic, stereoselective NCIs.

Conclusion

We have demonstrated *active site-based* buried volume is a uniquely advantageous descriptor for the construction of simple and holistic relationships to stereoselectivity by BAOCs. These catalysts alter stereoselectivity through an array of diverse NCIs collectively conferred by many distinct structural elements. As such, there is not one universally shared mechanistic attribute specifically exploited by the most selective catalysts. Still, *active site-based* buried volume navigates this stereoselective heterogeneity of each reaction dataset by standardizing spatial interactions, even across structurally diverse BAOCs. Inspired by these findings, we are currently investigating the application of *active site-based* buried volume to inform and guide the construction of stereoselective solutions in BAOc-mediated transformations. More broadly, intentional investigation of general chemical representation for *complex functions* (ex. stereoselectivity) across any chemistry may reveal surprisingly simple and intuitive relationships.

ASSOCIATED CONTENT

Supporting Information.

This material is available free of charge via the Internet at <http://pubs.acs.org>.

General calculation details and statistical analyses. (PDF)
Table of *active site-based* descriptors (CSV)
Table of *substituent-based* descriptors (CSV)
Table of reaction datasets (XLSX)
Table of all regression analyses for each descriptor (XLSX)

AUTHOR INFORMATION

Corresponding Authors

F. Dean Toste - *Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States; Department of Chemistry, University of California, Berkeley, California 94720, United States;* orcid.org/0000-0001-8018-2198; Email: fdtoste@berkeley.edu

Authors

Andrew L. Smith - *Department of Chemistry, University of California, Berkeley, California 94720, United States*

Author Contributions

The manuscript was written through contributions of all authors. / All authors have given approval to the final version of the manuscript.

Funding Sources

CCI Center for Computer Assisted Synthesis, CHE-2202693 (ALS).

ACKNOWLEDGMENT

The authors thank Dr. Kathleen Durkin, Dr. Azhagiya, Singam, Dr. Dave Small, and the Molecular Graphics and Computation Facility (MGCF) at the University of California, Berkeley for calculation and workflow assistance, as supported by National Institute of Health grant S10OD034382. We would like to thank Professor Matthew S. Sigman for helpful discussions.

ABBREVIATIONS

BAOC, Brønsted acid organocatalyst
NCI, non-covalent interaction
SAR, structure-activity relationship

REFERENCES

- (1) Anslyn, E.; Dougherty, D. *Modern Physical Organic Chemistry*; University Science Books, 2006.
- (2) Hammett, L. P. Some Relations between Reaction Rates and Equilibrium Constants. *Chem. Rev.* **1935**, 17 (1), 125–136.
- (3) Taft, R. W. Jr. Linear Steric Energy Relationships. *J. Am. Chem. Soc.* **1953**, 75 (18), 4538–4539.
- (4) Charton, M. Steric Effects. I. Esterification and Acid-Catalyzed Hydrolysis of Esters. *J. Am. Chem. Soc.* **1975**, 97 (6), 1552–1556.
- (5) Milo, A.; Bess, E. N.; Sigman, M. S. Interrogating Selectivity in Catalysis Using Molecular Vibrations. *Nature* **2014**, 507 (7491), 210–214.
- (6) Brethomé, A. V.; Fletcher, S. P.; Paton, R. S. Conformational Effects on Physical-Organic Descriptors: The Case of Sterimol Steric Parameters. *ACS Catal.* **2019**, 9 (3), 2313–2323.
- (7) Weinhold, F. Natural Bond Orbital Analysis: A Critical Overview of Relationships to Alternative Bonding Perspectives. *Journal of Computational Chemistry* **2012**, 33 (30), 2363–2379.
- (8) Williams, W. L.; Zeng, L.; Gensch, T.; Sigman, M. S.; Doyle, A. G.; Anslyn, E. V. The Evolution of Data-Driven Modeling in Organic Chemistry. *ACS Cent. Sci.* **2021**, 7 (10), 1622–1637.

- (9) B. Santiago, C.; Guo, J.-Y.; S. Sigman, M. Predictive and Mechanistic Multivariate Linear Regression Models for Reaction Development. *Chem. Sci.* **2018**, 9 (9), 2398–2412.
- (10) Gallegos, L. C.; Luchini, G.; St. John, P. C.; Kim, S.; Paton, R. S. Importance of Engineered and Learned Molecular Representations in Predicting Organic Reactivity, Selectivity, and Chemical Properties. *Acc. Chem. Res.* **2021**, 54 (4), 827–836.
- (11) Shalit Peleg, H.; Milo, A. Small Data Can Play a Big Role in Chemical Discovery. *Angew. Chem. Int. Ed.* **2023**, 62 (26), e202219070.
- (12) Haas, B. C.; Kalyani, D.; Sigman, M. S. Applying Statistical Modeling Strategies to Sparse Datasets in Synthetic Chemistry. *Science Adv.* **2025**, 11 (1), eadt3013.
- (13) Neel, A. J.; Hilton, M. J.; Sigman, M. S.; Toste, F. D. Exploiting Non-Covalent π Interactions for Catalyst Design. *Nature* **2017**, 543 (7647), 637–646.
- (14) Toste, F. D.; Sigman, M. S.; Miller, S. J. Pursuit of Noncovalent Interactions for Strategic Site-Selective Catalysis. *Acc. Chem. Res.* **2017**, 50 (3), 609–615.
- (15) Akiyama, T. Stronger Brønsted Acids. *Chem. Rev.* **2007**, 107 (12), 5744–5758.
- (16) 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**, 114 (18), 9047–9153.
- (17) Chen, D.-F.; Han, Z.-Y.; Zhou, X.-L.; Gong, L.-Z. Asymmetric Organocatalysis Combined with Metal Catalysis: Concept, Proof of Concept, and Beyond. *Acc. Chem. Res.* **2014**, 47 (8), 2365–2377.
- (18) Mahlau, M.; List, B. Asymmetric Counteranion-Directed Catalysis: Concept, Definition, and Applications. *Angew. Chem. Int. Ed.* **2013**, 52 (2), 518–533.
- (19) Handjaya, J. P.; Patankar, N.; Reid, J. P. The Diversity and Evolution of Chiral Brønsted Acid Structures. *Chem. Eur. J.* **2024**, 30 (46), e202400921.
- (20) Betinol, I. O.; Kuang, Y.; Mulley, B. P.; Reid, J. P. Controlling Stereoselectivity with Noncovalent Interactions in Chiral Phosphoric Acid Organocatalysis. *Chem. Rev.* **2025**, 125 (8), 4184–4286.
- (21) Akiyama, T.; Mori, K. Stronger Brønsted Acids: Recent Progress. *Chem. Rev.* **2015**, 115 (17), 9277–9306.
- (22) Mitschke, B.; Turberg, M.; List, B. Confinement as a Unifying Element in Selective Catalysis. *Chem* **2020**, 6 (10), 2515–2532.
- (23) Schreyer, L.; Properzi, R.; List, B. IDPi Catalysis. *Angew. Chem. Int. Ed.* **2019**, 58 (37), 12761–12777.
- (24) Hillier, A. C.; Sommer, W. J.; Yong, B. S.; Petersen, J. L.; Cavallo, L.; Nolan, S. P. A Combined Experimental and Theoretical Study Examining the Binding of N-Heterocyclic Carbenes (NHC) to the Cp^{*}RuCl (Cp^{*} = H⁵-C⁵Me⁵) Moiety: Insight into Stereoelectronic Differences between Unsaturated and Saturated NHC Ligands. *Organometallics* **2003**, 22 (21), 4322–4326.
- (25) Antinucci, G.; Dereli, B.; Vittoria, A.; Budzelaar, P. H. M.; Cipullo, R.; Goryunov, G. P.; Kulyabin, P. S.; Uborsky, D. V.; Cavallo, L.; Ehm, C.; Voskoboynikov, A. Z.; Busico, V. Selection of Low-Dimensional 3-D Geometric Descriptors for Accurate Enantioselectivity Prediction. *ACS Catal.* **2022**, 12 (12), 6934–6945.
- (26) Zahrt, A. F.; Henle, J. J.; Rose, B. T.; Wang, Y.; Darrow, W. T.; Denmark, S. E. Prediction of Higher-Selectivity Catalysts by Computer-Driven Workflow and Machine Learning. *Science* **2019**, 363 (6424), eaau5631.
- (27) Wakchaure, V. N.; DeSnoo, W.; Laconsay, C. J.; Leutzsch, M.; Tsuji, N.; Tantillo, D. J.; List, B. Catalytic Asymmetric Cationic Shifts of Aliphatic Hydrocarbons. *Nature* **2024**, 625 (7994), 287–292.
- (28) Reid, J. P.; Sigman, M. S. Holistic Prediction of Enantioselectivity in Asymmetric Catalysis. *Nature* **2019**, 571 (7765), 343–348.
- (29) Milo, A.; Neel, A. J.; Toste, F. D.; Sigman, M. S. A Data-Intensive Approach to Mechanistic Elucidation Applied to Chiral Anion Catalysis. *Science* **2015**, 347 (6223), 737–743.
- (30) Verloop, A.; Hoogenstraaten, W.; Tipker, J. Chapter 4 - Development and Application of New Steric Substituent Parameters in Drug Design. In *Drug Design*; Ariëns, E. J., Ed.; Medicinal Chemistry: A Series of Monographs; Academic Press: Amsterdam, 1976; Vol. 11, pp 165–207.
- (31) Kariofillis, S. K.; Jiang, S.; Żurański, A. M.; Gandhi, S. S.; Martinez Alvarado, J. I.; Doyle, A. G. Using Data Science To Guide Aryl Bromide Substrate Scope Analysis in a Ni/Photoredox-Catalyzed Cross-Coupling with Acetals as Alcohol-Derived Radical Sources. *J. Am. Chem. Soc.* **2022**, 144 (2), 1045–1055.
- (32) Peng, B.; Ma, J.; Guo, J.; Gong, Y.; Wang, R.; Zhang, Y.; Zeng, J.; Chen, W.-W.; Ding, K.; Zhao, B. A Powerful Chiral Super Brønsted C–H Acid for Asymmetric Catalysis. *J. Am. Chem. Soc.* **2022**, 144 (7), 2853–2860.
- (33) Dalmau, D.; Sigman, M. S.; Alegre-Requena, J. V. Machine Learning Workflows beyond Linear Models in Low-Data Regimes. *Chem. Sci.* **2025**.
- (34) *Model Selection and Multimodel Inference*; Burnham, K. P., Anderson, D. R., Eds.; Springer: New York, NY, 2004.
- (35) Ding, W.-W.; He, Z.-Y.; Sayed, M.; Zhou, Y.; Han, Z.-Y.; Gong, L.-Z. Enantioselective Synthesis of β - and α -Amino Ketones through Reversible Alkane Carbonylation. *Nat. Synth* **2024**, 3 (4), 507–516.
- (36) Rauniyar, V.; Lackner, A. D.; Hamilton, G. L.; Toste, F. D. Asymmetric Electrophilic Fluorination Using an Anionic Chiral Phase-Transfer Catalyst. *Science* **2011**, 334 (6063), 1681–1684.
- (37) Terada, M.; Soga, K.; Momiyama, N. Enantioselective Activation of Aldehydes by Chiral Phosphoric Acid Catalysts in an Aza-Ene-Type Reaction between Glyoxylate and Enecarbamate. *Angew. Chem. Int. Ed.* **2008**, 47 (22), 4122–4125.
- (38) Wang, Z.; Sheong, F. K.; Sung, H. H. Y.; Williams, I. D.; Lin, Z.; Sun, J. Catalytic Enantioselective Intermolecular Desymmetrization of Azetidines. *J. Am. Chem. Soc.* **2015**, 137 (18), 5895–5898.

- (39) Biswas, S.; Kubota, K.; Orlandi, M.; Turberg, M.; Miles, D. H.; Sigman, M. S.; Toste, F. D. Enantioselective Synthesis of N,S-Acetals by an Oxidative Pummerer-Type Transformation Using Phase-Transfer Catalysis. *Angew. Chem. Int. Ed.* **2018**, *57* (2), 589–593.
- (40) Eschmann, C.; Song, L.; Schreiner, P. R. London Dispersion Interactions Rather than Steric Hindrance Determine the Enantioselectivity of the Corey–Bakshi–Shibata Reduction. *Angew. Chem. Int. Ed.* **2021**, *60* (9), 4823–4832.
- (41) Leinung, W.; Mitschke, B.; Leutzsch, M.; Wakchaure, V. N.; Maji, R.; List, B. From Reagent to Catalyst: Dispersion-Driven Design of a General Asymmetric Transfer Hydrogenation Catalyst. *J. Am. Chem. Soc.* **2025**.
- (42) Knowles, R. R.; Lin, S.; Jacobsen, E. N. Enantioselective Thiourea-Catalyzed Cationic Polycyclizations. *J. Am. Chem. Soc.* **2010**, *132* (14), 5030–5032.
- (43) Wheeler, S. E. Local Nature of Substituent Effects in Stacking Interactions. *J. Am. Chem. Soc.* **2011**, *133* (26), 10262–10274.
- (44) Dougherty, D. A. The Cation- π Interaction in Chemistry and Biology. *Chem. Rev.* **2025**, *125* (5), 2793–2808.
- (45) Orlandi, M.; Coelho, J. A. S.; Hilton, M. J.; Toste, F. D.; Sigman, M. S. Parametrization of Non-Covalent Interactions for Transition State Interrogation Applied to Asymmetric Catalysis. *J. Am. Chem. Soc.* **2017**, *139* (20), 6803–6806.
- (46) Vijay, D.; Sastry, G. N. Exploring the Size Dependence of Cyclic and Acyclic π -Systems on Cation- π Binding. *Phys. Chem. Chem. Phys.* **2008**, *10* (4), 582–590.
- (47) Chen, F.; Chen, Y.; Chang, X.-Y.; He, D.; Yang, Q.; Wang, D. Z.; Xu, C.; Yu, P.; Xing, X. Polarizability Matters in Enantio-Selection. *Nat Commun* **2024**, *15* (1), 3394.
- (48) Treacy, S. M.; Smith, A. L.; Bergman, R. G.; Raymond, K. N.; Toste, F. D. Supramolecular Catalyzed Cascade Reduction of Azaarenes Interrogated via Data Science. *J. Am. Chem. Soc.* **2024**, *146* (43), 29792–29800.
- (49) Heafner, E. D.; Smith, A. L.; Craescu, C. V.; Raymond, K. N.; Bergman, R. G.; Toste, F. D. Probing Enantioinduction in Confined Chiral Spaces through Asymmetric Oxime Reductions. *Chem* **2025**, *11* (4), 102368.
- (50) Maji, R.; Mallojjala, S. C.; Wheeler, S. E. Electrostatic Interactions in Asymmetric Organocatalysis. *Acc. Chem. Res.* **2023**, *56* (14), 1990–2000.
- (51) Laconsay, C. J.; Seguin, T. J.; Wheeler, S. E. Modulating Stereoselectivity through Electrostatic Interactions in a SPINOL-Phosphoric Acid-Catalyzed Synthesis of 2,3-Dihydroquinazolinones. *ACS Catal.* **2020**, *10* (20), 12292–12299.
- (52) Rueping, M.; Antonchick, A. P. Catalytic Asymmetric Aminoallylation of Aldehydes: A Catalytic Enantioselective Aza-Cope Rearrangement. *Angew. Chem. Int. Ed.* **2008**, *47* (52), 10090–10093.
- (53) Cheng, X.; Vellalath, S.; Goddard, R.; List, B. Direct Catalytic Asymmetric Synthesis of Cyclic Aminals from Aldehydes. *J. Am. Chem. Soc.* **2008**, *130* (47), 15786–15787.

