Title
Supramolecular Catalyzed Cascade Reduction of Azaarenes Interrogated via Data Science.

Journal
Journal of the American Chemical Society, 146(43)

Authors
Treacy, Sean
Smith, Andrew
Bergman, Robert
et al.

Publication Date
2024-10-30

DOI
10.1021/jacs.4c11482

Peer reviewed

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS

Article

Supramolecular Catalyzed Cascade Reduction of Azaarenes Interrogated via Data Science

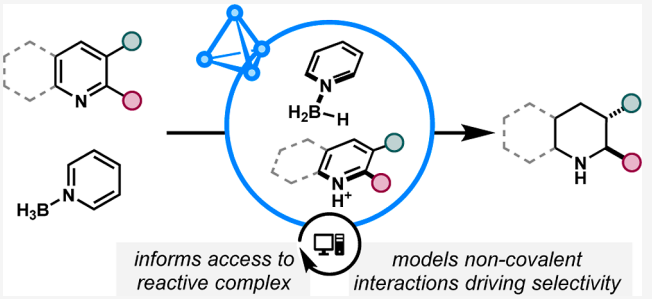
Sean M. Treacy,[§] Andrew L. Smith,[§] Robert G. Bergman, Kenneth N. Raymond, and F. Dean Toste*

Cite This: J. Am. Chem. Soc. 2024, 146, 29792−29800

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Catalysis of multicomponent transformations requires controlled assembly of reactants within the active site. Supramolecular scaffolds possess synthetic microenvironments that enable precise modulation over noncovalent interactions (NCIs) engaged by reactive, encapsulated species. While molecular properties that describe the behavior of single guests in host cavities have been studied extensively, multicomponent transformations remain challenging to design and deploy. Here, simple univariate regression and threshold analyses are employed to model reactivity in a cascade reduction of azaarenes catalyzed by water-soluble metal organic cages. Yield and stereoselectivity models help deduce unknown mechanisms of reactivity by the multicomponent, host−guest complexes. Furthermore, a comprehensive model is established for NCIs driving stereoselectivity in the reported host−guest adducts.



INTRODUCTION

In nature, enzymes leverage an array of carefully tuned noncovalent interactions (NCIs) to decrease activation barriers to reactions along complex biosynthetic pathways.[1,2] To mimic these enzyme binding pockets, chemists have engineered supramolecular assemblies capable of imposing NCIs on encapsulated molecules to enable reactivity not observed in bulk solution.[3−6] We have long studied the host−guest complexation within the distinct microenvironments of anionic metal catecholate covalent organic cages and demonstrated their ability to catalyze various transformations[7,8] (Figure 1A). More recently, we reported numerous synthetic methodologies leveraging reactivity across multiple reaction components simultaneously assembled within the host cavity, furnishing intermediate ternary complexes.[9−11] Our current mechanistic understanding suggests that our reported, host-mediated transformations are not only limited by access into these ternary complexes but also influenced by NCIs along the dynamic, reaction coordinate. As these reactive complexes are highly transient, general chemical features that govern these transformations have remained elusive.[12−15]

To analyze complex reactivity, data science has emerged as an invaluable tool to deconstruct key features that correlate to desired reaction outcomes (Figure 1B). While machine learning-based predictions have rapidly grown in popularity, analysis by simple linear regression models and univariate thresholds is essential to abstract complex reactivity onto a handful of intuitive, chemically tangible principles.[16−18] We posited that these simple data science tools would elucidate the principles that govern access to coencapsulated intermediates and their interactions within active site cavities. Modeling the yield of transformations proceeding through ternary complexes necessarily illuminates chemical features that mediate access into these adducts. Additionally, models conveying any host-mediated chemo-, regio-, or stereoselectivity should provide crucial information regarding the stereoelectronic perturbations of encapsulated intermediates imposed by the cavity microenvironment.

Because previous work from our group has disclosed a host-mediated reduction facilitated by the coencapsulation of both substrate and pyridine−borane, it was anticipated that critical insight could be furnished by the discovery of additional, host-mediated, multicomponent reactivity.[10] Specifically, we hypothesized that encapsulated azaarenes are poised for a protonation−reduction cascade with pyridine−boranes to furnish valuable semihydrogenated products[19,20] (Figure 1C). In this manner, azaarenes would undergo protonation upon encapsulation into the highly anionic, metal catecholate cage, priming the substrate to dearomatizing reduction by coencapsulating pyridine−borane.

Herein, we employ a data science-based analysis, coupled with careful experimental design, to provide quantitative insight into the substrate and host features responsible for the observed, supramolecular catalytic activity. Specifically,

Received: August 20, 2024
Revised: October 8, 2024
Accepted: October 9, 2024
Published: October 21, 2024

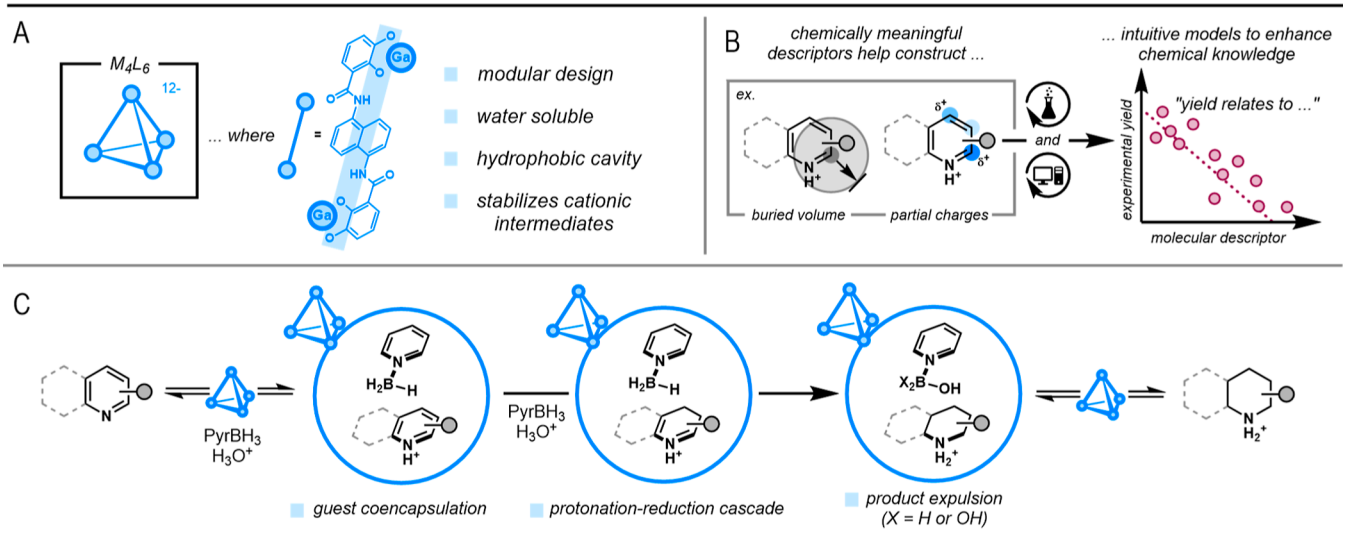


Figure 1. Investigating reactive, ternary complexes requires chemically coherent systems and models. (A) Metal catecholate cages possess synthetically tractable, biomimetic binding pockets. (B) Data science helps generate chemical knowledge based on correlations between chemically meaningful descriptors and experimental observables. (C) Proposed model azaarene reduction cascade to interrogate ternary complex formation and reactivity.

linear regression models based on size exclusion by the host cavity were successfully deployed across systematically modified hosts. These models generalize substrate binding principles of ternary host−guest adducts that undergo the reported reaction. Additionally, transient NCIs responsible for stereoselectivity in the reported reduction cascade could be rationalized through substrate parametrization and calculated NCIs within the binding pocket. Overall, we showcase how a data science-powered workflow precisely quantifies the features of substrates that affect reactivity and stereoselectivity in supramolecular catalysis.

RESULTS

Establishing the Azaarene Reduction Cascade and Mechanism. We first considered whether the reduction cascade was possible with host **I**, a Ga_4L_6 coordination cage (L = *N*,*N*′-bis(2,3-dihydroxybenzoyl)-1,5-diaminonaphthalene). The reduction of model azaarene 2-methylquinoline **1** with pyridine−borane mediated by host **I** furnished the semi-hydrogenated, 1,2,3,4-tetrahydro-2-methylquinoline in 97% yield (entry 1, Figure 2A). The absence of a catalyst or the addition of a strongly encapsulating cationic guest results in low yields of the desired product, which supports that catalytic activation of the quinoline and reductant by host **I** occurs within the cavity (entries 2−4, Figure 2A). Conducting this protocol at lower temperatures affords the reduced product at attenuated yields, and the reductant is required for the desired reactivity (entries 5−6, Figure 2A).

As the reaction necessarily proceeds through a coencapsulated complex, we first systematically evaluated how substrate size influences the reduction. Since host **I** permits encapsulation of guests through aperture dilation,[21] we hypothesized that volume of the most compact conformer for a given substrate may correlate with the observed reactivity. To interrogate this possibility, we calculated the minimum molecular volumes of ten additional mono- and disubstituted quinolines. This was accomplished by first generating conformational ensembles within a 1.8 kcal/mol window using Grimme's CREST package (see the Supporting Information General Calculation and Statistical Considerations section). Then, the molecular volume was computed using a descriptor calculator from the open-source RDKit package, and the minimum value from across the ensemble was extracted. The molecular volumes of the quinolines considered range over 140−210 Å^3 (Figure 2B). The reduction of each of the substrates was then evaluated under the conditions established above, and their yield was recorded. A linear model between yield and minimum molecular volume was anticipated to characterize how host-dependent size exclusion of the reactive complex emerges as a response to guest size. An analogous analysis can be completed using univariate threshold detection, where substrates are binned into active or inactive categories based on their yield, and the molecular volume of these substrates is used to detect where size exclusion formally emerges. Consistent with the hypothesis that guest size relates to reaction yield, a linear correlation was observed where the minimum molecular volume of the substrate captured experimental yield with an R^2 of 0.77 (Figure 2C). The corresponding threshold analysis was similarly robust with an active volume of 172 Å^3 at a reactivity threshold of 30% (MCC = 0.83).

The data science-based analysis supports the hypothesis that quinoline molecular volume has a profound influence on access into the reactive, ternary complex according to molecular volume. Despite the observation of minimal to no reactivity with substrates with minimum molecular volumes > 200 Å^3, encapsulation studies of 2-substituted quinolines **1**−**4** with volumes ranging from 140 to 205 Å^3 show that even the largest members quantitatively encapsulate within the cavity of host **I** under reaction concentrations (see Supporting Information NMR spectra). Thus, in accordance with the above correlations between quinoline volume and reaction yield, the rate of subsequent coencapsulation of the reductant was anticipated to become slower as the volume of the encapsulated quinoline increases. An initial rates kinetic analysis of the reduction was consistent with first-order rate dependence on catalyst and pyridine−borane concentrations, but zero-order dependence on quinoline concentration (see

https://doi.org/10.1021/jacs.4c11482
J. Am. Chem. Soc. 2024, 146, 29792−29800

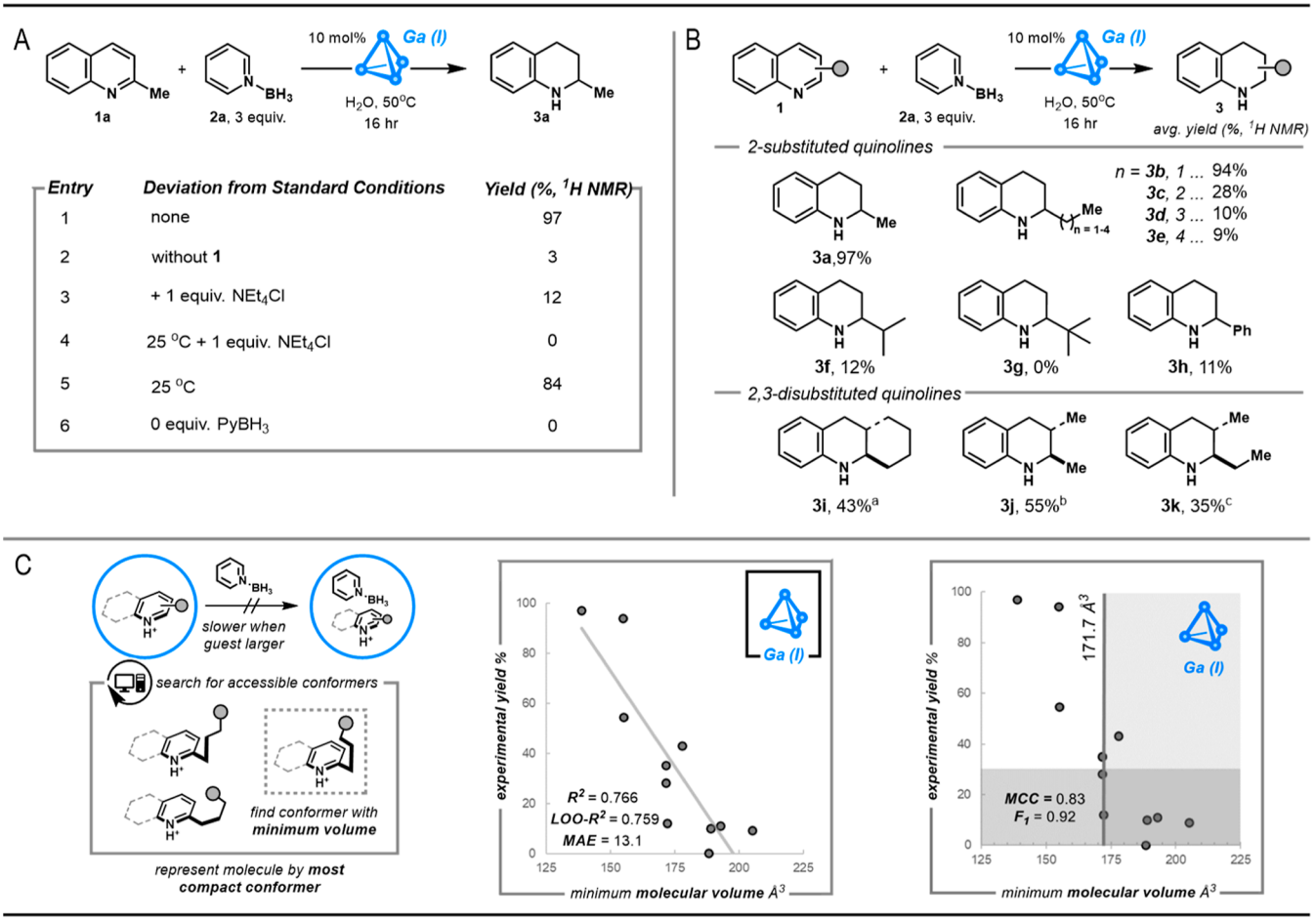


Figure 2. Establishing the host-mediated azaarene reduction cascade reactivity, scope, and mechanism. (A) Reaction discovery and controls. (B) Systematic series of quinolines, see Table S6, [a]4:1 dr, [b]19:1 dr, and [c]9:1 dr. (C) Representation of molecular volume in its most compact conformer linearly correlates (R^2 = 0.77) with yields of supramolecular catalyzed transformation and are well represented by a univariate threshold analysis (yield threshold = 30%, detected molecular volume threshold = 171.7 Å^3, MCC = 0.83).

Figures S1−S4). In agreement with the former quinoline encapsulation studies, these data indicated that the formation of a resting-state complex between the host and the quinoline guest occurs prior to pyridine−borane coencapsulation. To interrogate hydride delivery as the potential rate-determining step in this reaction, kinetic isotope effect experiments were conducted via one-pot competition between $PyBH_3$ and $PyBD_3$, which yielded no significant difference in D versus H incorporation at either position on the product (see Figures S5 and S6). The lack of a kinetic isotope effect in combination with the first-order rate dependence on the reductant is consistent with rate-limiting coencapsulation of pyridine−borane to access the reactive, ternary complex.

Reactivity Principles Conserved across Hosts. Because pyridine−borane coencapsulation is implicated as the turnover-limiting step, the yield of the host-mediated reduction formally relates to how one host−guest complex can accommodate strain along a second encapsulation event. Accelerating this second encapsulation event rationally implies greater yield in the transformation; however, guest exchange rates for these bimolecular events are unknown. Instead, only their unimolecular congeners are precedented.[21] Over the past decade, a modest array of hosts with variable guest exchange rates has been achieved by systematic modifications to the metal vertices of the host and the ligand linker of the metal organic cage[11] (Figure 3A). Tuning the metal vertices of the cage modulates guest-exchange rates but does not change the absolute size of the host cavity.[22] The prototypical Ga (host **I**) vertex can be exchanged to Al (host **II**), affording hosts with slower guest exchang, or, alternatively In (host **III**), affording hosts with more rapid guest exchange, according to the strength of the metal−oxygen bonds of the scaffold. The ligand linker of the assembly can also be substituted as the naphthalene spacer of host **I** can be replaced with the pyrene-furnishing host **IV**.[23] As a result, host **IV** not only exhibits a larger aperture, lowering barriers to guest encapsulation, but also a larger host cavity altogether, enabling an absolute increase to the size of guests tolerated.

With consideration of these principles discovered in unimolecular contexts, reaction yield was leveraged as a platform to measure access into the reactive ternary complexes. Thus, additional data sets for each host assembly **II**−**IV** were generated from the azaarene scope described in Figure 2B and subjected each to linear regression and univariate threshold analysis. Across all hosts, good-to-excellent correlations between yield and minimum molecular volume were observed (up to R^2 = 0.84, Figure 3B). Additionally, excellent univariate threshold detection over minimum molecular volume was observed (up to MCC = 1.0). The In host **III** was expected to provide a more flexible assembly, suitable for more rapid guest

https://doi.org/10.1021/jacs.4c11482
J. Am. Chem. Soc. 2024, 146, 29792−29800

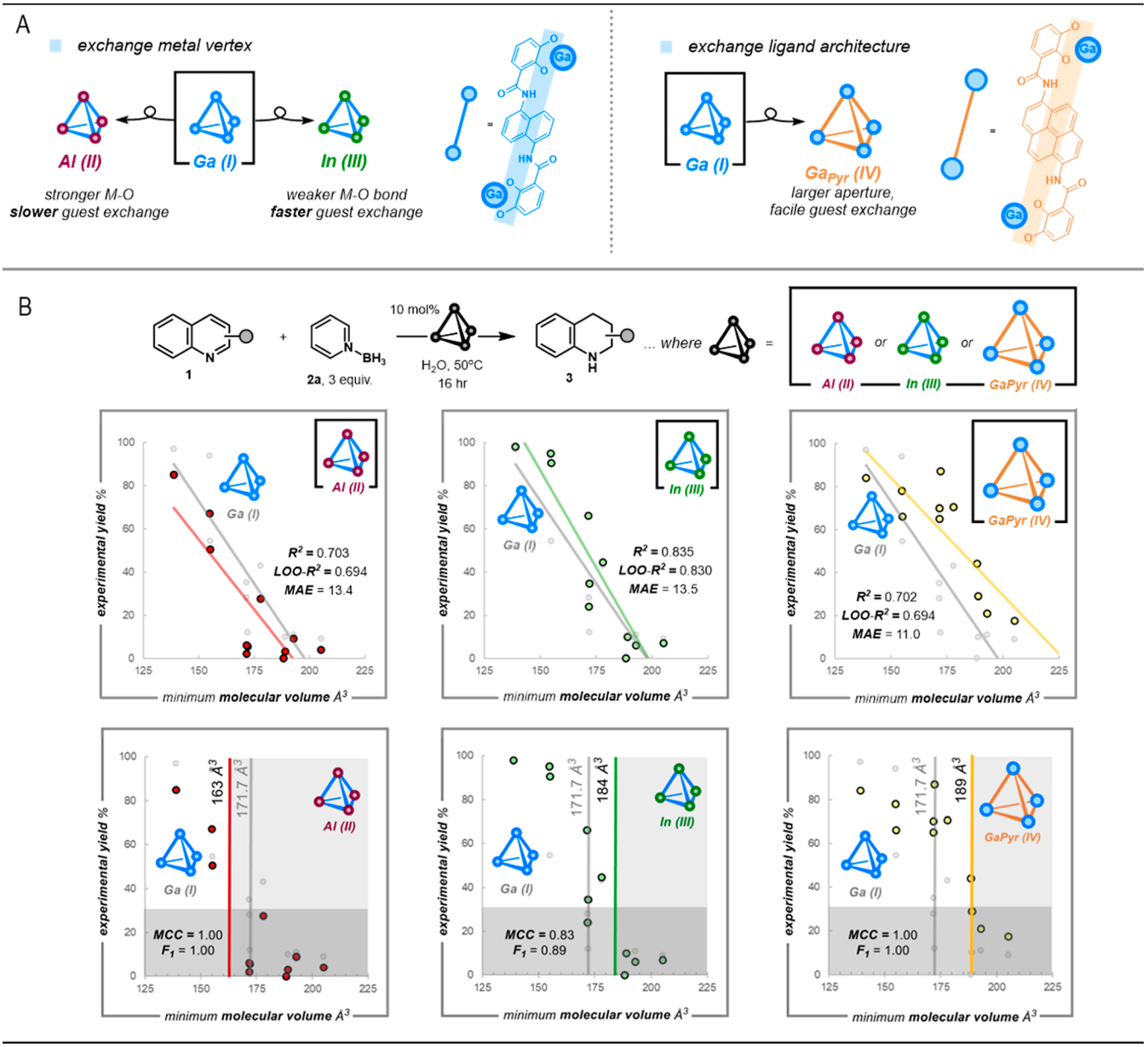


Figure 3. Chemical principles describing reactivity of ternary complexes transfers across distinct hosts assemblies. (A) Exchange of the metal cation of the supramolecular cage increases the flexibility of the cavity, leading to differences in guest exchange rates. Using a pyrene ligand architecture increases the size of the cavity and enables more facile guest exchange. (B) Minimum molecular volume maintains robust linear correlations across the catalyst series (up to R^2 = 0.84), with corresponding excellent univariate threshold analyses (yield threshold = 30%, up to MCC = 1.0), see Tables S6–S9.

exchange. Indeed, the In host **III** shows an upward slant of the yield correlation as well as an increased threshold for reactivity (184 Å^3 versus 172 Å^3, MCC = 0.83) relative to the Ga host **I**, indicating that larger substrates more easily permit coencapsulation with pyridine–borane, thus accessing higher yields, such as product **3k** now formed in 66% yield (vs 35% with host **I**). In contrast, the Al host **II** is expected to have lower exchange rates. The observed correlation and threshold analysis reflect this expectation, showing a pronounced downward slant and depressed threshold for reactivity relative to host **I**, (163 Å^3 versus 172 Å^3, MCC = 1.00), indicating a severe kinetic barrier to coencapsulation of pyridine borane with many quinolines screened. Finally, the Ga host **IV** with a larger pyrene spacer is expected to have not only faster exchange rates due to an increased size of catalyst aperture but also a larger absolute limit to guest encapsulation due to its larger cavity. The observed correlation reflects this by a rightward shift of the yield correlation, now corresponding to higher yields across the 130–210 Å^3 minimum molecular volume series as well as a best-in-class threshold for reactivity (189 Å^3 versus 172 Å^3, MCC = 1.00). Overall, our protocol quantitatively substantiates relationships between guest size and flexibility of the host catalyst for guest coencapsulation, transforming previously established qualitative principles for more complex applications.

Elucidating Stereoelectronic Effects on Host-Catalyzed, Cascade Reduction. To elaborate which features of the intermediate host–guest complex influence the second encapsulation event, the scope of quinolines considered was expanded for the reduction. As the pyrene-based host **IV**

https://doi.org/10.1021/jacs.4c11482
J. Am. Chem. Soc. 2024, 146, 29792–29800

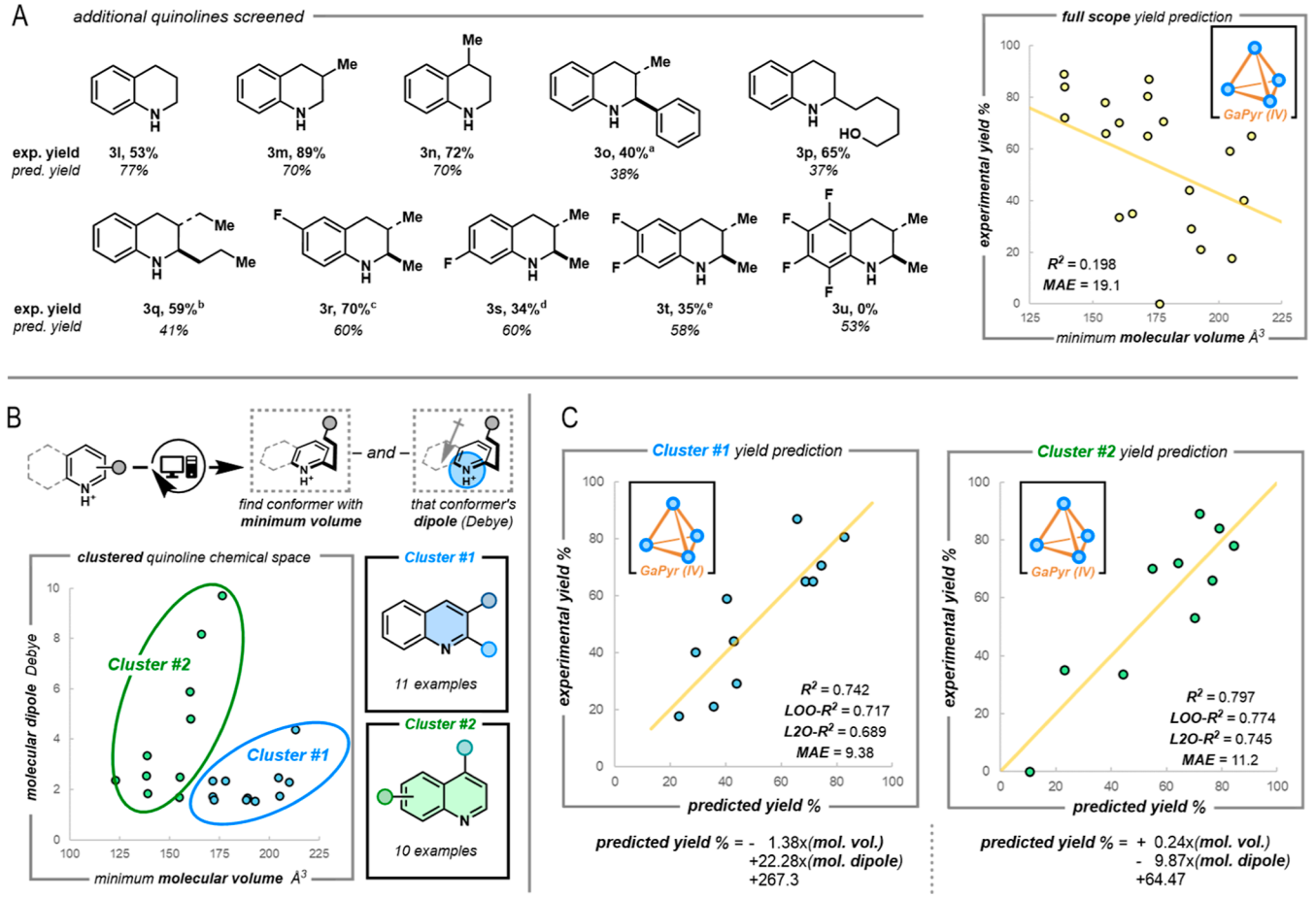


Figure 4. Reactivity of the ternary complex over an expanded substrate scope divergently correlates to substrate stereoelectronic features. (A) Additional quinolines scope and limited prediction capacity of minimum molecular volume on reaction yield, see Table S10; [a]1.7:1, dr, [b]11.2:1 dr, [c]10.5:1 dr, [d]4.6:1 dr, and [e]4.5:1 dr. (B) Unsupervised clustering reveals two classes of quinoline substrates. (C) Cluster specific bivariate linear regression using molecular volume and molecular dipole differentially correlates to reaction yield, up to R^2 = 0.80, see Table S11.

broadly displayed improved yields in the previous data set, we assessed the generality of the yield prediction model based on minimum molecular volume of the quinoline substrate. The expanded scope considered a wider range of substituted quinolines with varying steric bulk and additional functional groups (Figure 4A). After performing the reactions with host **IV** and comparing the outcomes to the predicted yields, the observed minimum molecular volume exhibits little correlation to the experimental yield (R^2 = 0.20, Figure 4A).

To investigate the origin of this result, we hypothesized that the minimum molecular volume of the quinoline guest erroneously assumes that each quinoline maintains a shape and corresponding stereoelectronic character. As such, the expanded substrate scope would be better represented by separate groups of quinolines, where each group shares similar stereoelectronic properties such that their host binding behavior is internally consistent. The substrate scope was categorized by clustering all quinolines considered over a principal component-reduced chemical space built from the minimum molecular volume data set. The minimum molecular volume data set refers to the full array of features extracted from the conformer of each quinoline that has the minimum molecular volume across the ensemble computed (see the Supporting Information Molecular Descriptor Calculation Workflow section). This analysis revealed two clusters, wherein the first cluster was dominated by 2- and 3-substituted quinolines, while the second cluster was best represented by all other substitution patterns (Figure 4B). Within each cluster, yield correlations built from minimum molecular volume remained poor, and univariate threshold analyses lost coherence. Therefore, we screened suitable bivariate correlations and noticed that minimum molecular volume and molecular dipole now rescued yield predictions (up to R^2 = 0.80, Figure 4C). Interrogation of each cluster's yield correlation revealed molecular dipole contributed to each cluster's prediction in the opposite sense. As molecular dipole is related to the conformation of the molecule, we interpret the improved bivariate linear correlations to indicate that molecular dipole differentially corrects each yield prediction based solely on molecular volume to account for apparent shape differences between each cluster (see the Supporting Information Extended Discussion on Figure 4C section). Whereas the previous models explicitly modeled steric effects on tertiary complex inhibition, an expanded chemical space demonstrated that adequate representation of molecular shape with confounding electronic features of the substrates is essential to understand global reactivity trends.

Illuminating Transient Active-Site Interactions. While yield prediction tasks quantified host and guest features influencing access to coencapsulated intermediates, the behavior of the coencapsulated complex itself remained elusive. Here, the diastereoselectivity of the reduction can provide

https://doi.org/10.1021/jacs.4c11482
J. Am. Chem. Soc. 2024, 146, 29792−29800

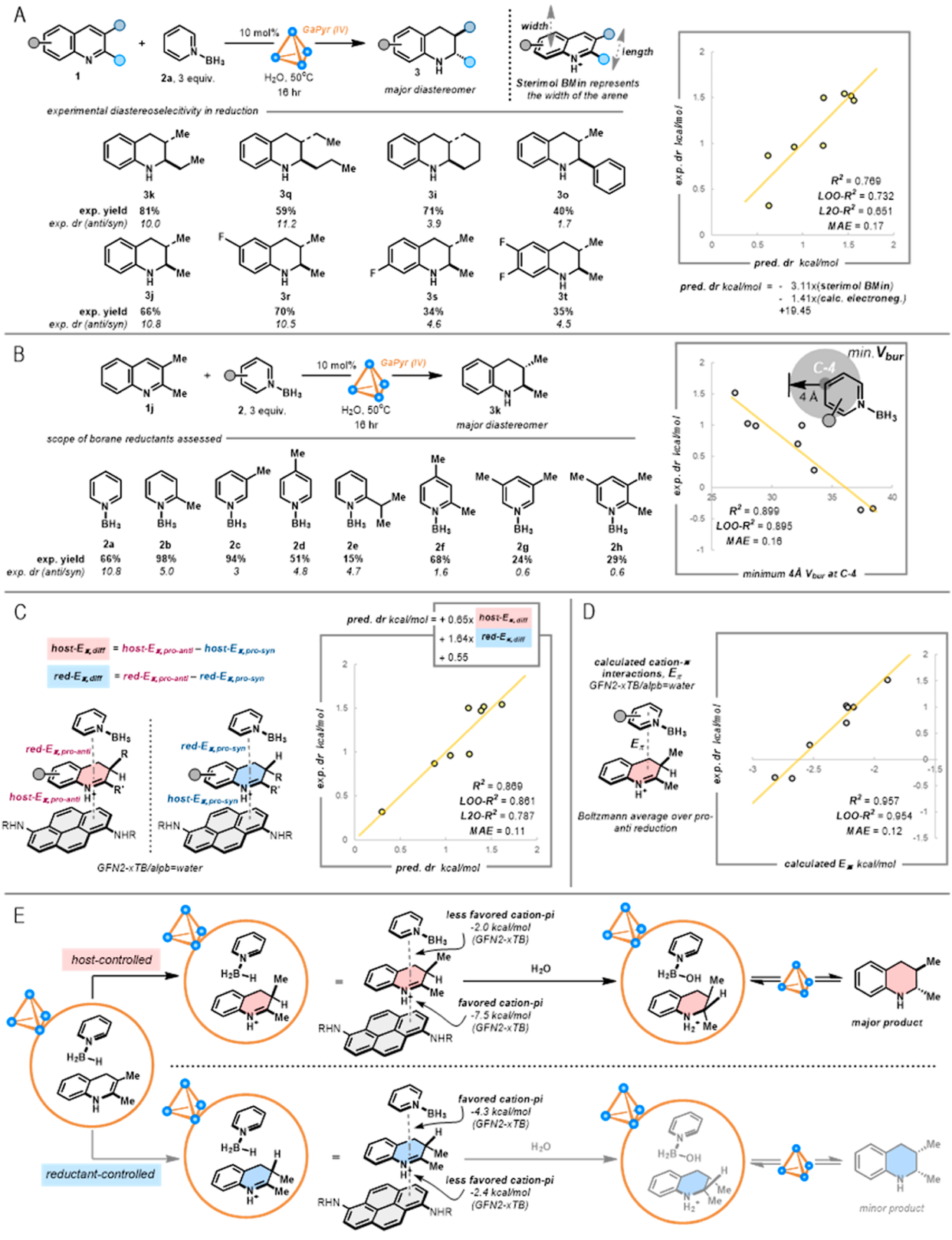


Figure 5. Evaluation of transient, stereoselective interactions across the reactive, ternary complex within the host-mediated azaarene reduction cascade. (A) Sterimol BMin and electronegativity of the quinoline substrate correlate to the diastereoselectivity of the azaarene reduction, R^2 = 0.77, see Table S14. (B) Buried volume at C-4 of pyridine−borane correlates to differences in diastereoselectivity of the **1j** reduction, R^2 = 0.90, see Table S15. (C) NCIs across proposed selectivity determining ternary complex recapitulates observed diastereoselectivity for a series of quinolines with **2a**, R^2 = 0.87, see Table S17. (D) NCIs across proposed selectivity determining ternary complex recapitulates observed diastereoselectivity for a series of pyridine−boranes with **1j**, R^2 = 0.96, see Table S18. (E) Comprehensive model for the NCIs driving stereoselectivity in the reported reactive, ternary host−guest adducts.

https://doi.org/10.1021/jacs.4c11482
J. Am. Chem. Soc. 2024, 146, 29792−29800

valuable information to understand the behavior of transient intermediates within the host cavity. Akin to data science analyses of homogeneous reactions as well as enzymatic transformations,[24−31] we envisioned that we could adapt the data science workflow to agnostically probe for features that correlate to the measured selectivity, where conclusions generated by this data science-driven inquiry can inspire unprecedented models describing outcomes from host-mediated catalysis.

Indeed, various 2,3-disubstituted azaarenes underwent the reduction with high diastereoselectivity for the *anti*-isomer. In this expanded scope, modest-to-good selectivity from 2:1 to 11:1 (anti/syn) was observed for the reduction (Figure 5A). Solution state reduction of **1** also favors the *anti* diastereomer, but only in an attenuated 3:1 selectivity (compared to 10:1 by host **IV**, see Figure S45). When correlations were evaluated between substrate features and the experimental stereoselectivity, a model based on the molecule's width (quantified by sterimol BMin) and electronegativity emerged among the best performing (R^2 = 0.78, LOO-R^2 = 0.73, L2O-R^2 = 0.65. See Table S17). The combination of properties in the correlation could describe how accessible the π-system of the quinoline is both sterically and electronically, but it remains unclear what specific interactions are relevant for the observed reactivity.

To further evaluate how each reactive species contributes to the observed stereoselectivity, a structurally diverse series of different pyridine−boranes were then prepared and screened with quinoline **3j**. Now, a more pronounced range of diastereoselectivity was observed (11:1 to 1:3, anti/syn), where some reductants even favored the formation of the *syn* diastereomer (Figure 5B). Surveyed correlations revealed that the minimum 4 Å buried volume at C-4 of the reductant excellently correlates to the observed selectivity (R^2 = 0.95). Here, stereoselectivity tends to invert as this buried volume increases. In line with the interpretation proposed for the quinoline series, access to favorable NCIs promotes the diastereoselectivity for the *anti*-isomer.

From these data science-based footholds, we then looked to deconvolute the specific NCIs relevant to the observed stereochemical outcomes. In this sense, the diastereoselectivity measured for the quinolines from Figure 5A can be recapitulated with competing NCIs between pro-anti and pro-syn complexes for the intermediate iminium species with the pyrene wall of the host (Figure 5C). These interactions display a strong correlation to the measured diastereoselectivity (R^2 = 0.87). We also revisited the effect of changing the pyridine−borane reductant on NCIs within the coencapsulated complex. Here, we observed that the calculated NCI between the reductant and the iminium intermediate in the pro-anti complex provided a robust correlation (R^2 = 0.96, Figure 5D). This correlation indicates when the interaction between the cationic intermediate and the reductant strengthens; the selectivity in the reduction gradually shifts from the *anti* to the *syn* diastereomer. This suggests that a delicate balance between host−quinoline and quinoline−reductant interactions contributes to the observed diastereoselectivity.

Altogether, we inferred the following model for diastereoinduction of the transient coencapsulated intermediates (Figure 5E). Beginning microscopically from the singly reduced dihydroquinoline, protonation leads to one of two iminium species. Here, we propose that competing NCIs serve as the selectivity-determining criteria. In the pro-anti case, the host wall, approximated with a single pyrene molecule, achieves an energetically favorable, uninhibited interaction, while the pro-syn case is disrupted by steric interactions with the pyrene wall. However, as the reductant achieves stronger interactions with the proposed intermediate, the selectivity gradually inverts to favor the *syn* diastereomer. In this pathway, the reductant can engage in more energetically favorable, sterically unhindered interactions with the reactive species. Through a data science-powered workflow, we have gained valuable mechanistic understanding of the key NCIs that engender stereoselectivity in supramolecular host-catalyzed reactions of coencapsulated intermediates.

CONCLUDING REMARKS

Herein, we showcase how simple linear regressions and univariate threshold analyses reveal complex dynamics in a catalyst structure agnostic workflow. Coupling these tools with data generated from a supramolecular host-catalyzed cascade reduction of azaarenes transforms over a decade of mechanistic investigation of metal organic cages into quantifiable, rational synthetic choices. Our work highlights how supramolecular catalysis uniquely facilitates rational adaptations to host size and flexibility, enabling the design of NCI-rich active sites that productively manipulate reactive, ternary complexes. Critically, our workflow showcases how selection of chemically meaningful features provides interpretable models that synergize and enhance our expert understanding of how to exploit active sites in catalysis.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.4c11482.

Experimental procedure and data for all products, computational procedures, and supplementary discussions (PDF)

Table of descriptors and calculated interactions (XLSX)

AUTHOR INFORMATION

Corresponding Author

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

Authors

Sean M. Treacy − *Department of Chemistry, University of California, Berkeley, California 94720, United States; Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States*

Andrew L. Smith − *Department of Chemistry, University of California, Berkeley, California 94720, United States;* orcid.org/0009-0009-2713-5558

Robert G. Bergman − *Department of Chemistry, University of California, Berkeley, California 94720, United States; Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States;* orcid.org/0000-0002-3105-8366

Kenneth N. Raymond − *Department of Chemistry, University of California, Berkeley, California 94720, United States;*

Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States; orcid.org/0000-0001-6968-9801

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.4c11482

Author Contributions

[§]S.M.T. and A.L.S. contributed equally.

Funding

Director, Office of Science, Office of Basic Energy Sciences, and the Division of Chemical Sciences, Geosciences, and Bioscience of the U.S. Department of Energy at Lawrence Berkeley National Laboratory, Grant DE-AC02-05CH11231 (SMT). National Institutes of Health Kirschstein-NRSA postdoctoral fellowship F32GM153093 (SMT). CCI Center for Computer Assisted Synthesis, CHE-2202693 (ALS).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Dr. Hasan Celik and UC Berkeley's NMR facility in the College of Chemistry (CoC-NMR) for spectroscopic assistance. Instruments in the CoC-NMR are supported in part by the National Institute of Health grant S10OD024998. 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 the National Institute of Health grant S10OD034382. We would like to thank Professor Matthew S. Sigman for helpful discussions.

■ REFERENCES

(1) Williams, D. H.; Stephens, E.; O'Brien, D. P.; Zhou, M. Understanding Noncovalent Interactions: Ligand Binding Energy and Catalytic Efficiency from Ligand-Induced Reductions in Motion within Receptors and Enzymes. *Angew. Chem., Int. Ed.* **2004**, *43* (48), 6596−6616.

(2) Bissantz, C.; Kuhn, B.; Stahl, M. A. A Medicinal Chemist's Guide to Molecular Interactions. *J. Med. Chem.* **2010**, *53* (14), 5061−5084.

(3) Yoshizawa, M.; Tamura, M.; Fujita, M. Diels-Alder in Aqueous Molecular Hosts: Unusual Regioselectivity and Efficient Catalysis. *Science* **2006**, *312* (5771), 251−254.

(4) Kang, J.; Rebek, J. Acceleration of a Diels−Alder Reaction by a Self-Assembled Molecular Capsule. *Nature* **1997**, *385* (6611), 50−52.

(5) Zhang, Q.; Tiefenbacher, K. Terpene Cyclization Catalysed inside a Self-Assembled Cavity. *Nat. Chem.* **2015**, *7* (3), 197−202.

(6) Li, T.-R.; Huck, F.; Piccini, G.; Tiefenbacher, K. Mimicry of the proton wire mechanism of enzymes inside a supramolecular capsule enables *β*-selective O-glycosylations. *Nat. Chem.* **2022**, *14* (9), 985−994.

(7) Kaphan, D. M.; Levin, M. D.; Bergman, R. G.; Raymond, K. N.; Toste, F. D. A Supramolecular Microenvironment Strategy for Transition Metal Catalysis. *Science* **2015**, *350* (6265), 1235−1238.

(8) Hong, C. M.; Bergman, R. G.; Raymond, K. N.; Toste, F. D. Self-Assembled Tetrahedral Hosts as Supramolecular Catalysts. *Acc. Chem. Res.* **2018**, *51* (10), 2447−2455.

(9) Bierschenk, S. M.; Bergman, R. G.; Raymond, K. N.; Toste, F. D. A Nanovessel-Catalyzed Three-Component Aza-Darzens Reaction. *J. Am. Chem. Soc.* **2020**, *142* (2), 733−737.

(10) Morimoto, M.; Cao, W.; Bergman, R. G.; Raymond, K. N.; Toste, F. D. Chemoselective and Site-Selective Reductions Catalyzed by a Supramolecular Host and a Pyridine−Borane Cofactor. *J. Am. Chem. Soc.* **2021**, *143* (4), 2108−2114.

(11) Bierschenk, S. M.; Pan, J. Y.; Settineri, N. S.; Warzok, U.; Bergman, R. G.; Raymond, K. N.; Toste, F. D. Impact of Host Flexibility on Selectivity in a Supramolecular Host-Catalyzed Enantioselective Aza-Darzens Reaction. *J. Am. Chem. Soc.* **2022**, *144* (25), 11425−11433.

(12) Piskorz, T. K.; Martí-Centelles, V.; Young, T. A.; Lusby, P. J.; Duarte, F. Computational Modeling of Supramolecular Metallo-Organic Cages−Challenges and Opportunities. *ACS Catal.* **2022**, *12* (10), 5806−5826.

(13) Rizzuto, F. J.; Von Krbek, L. K. S.; Nitschke, J. R. Strategies for Binding Multiple Guests in Metal−Organic Cages. *Nat. Rev. Chem* **2019**, 3 (4), 204−222.

(14) Parrilla-Gutiérrez, J. M.; Granda, J. M.; Ayme, J.-F.; Bajczyk, M. D.; Wilbraham, L.; Cronin, L. Electron Density-Based GPT for Optimization and Suggestion of Host−Guest Binders. *Nat. Comput. Sci.* **2024**, *4* (3), 200−209.

(15) Norjmaa, G.; Vidossich, P.; Maréchal, J. D.; Ujaque, G. Modeling Kinetics and Thermodynamics of Guest Encapsulation into the $[M_4L_6]^{12-}$ Supramolecular Organometallic Cage. *J. Chem. Inf. Model.* **2021**, *61* (9), 4370−4381.

(16) Medford, A. J.; Kunz, M. R.; Ewing, S. M.; Borders, T.; Fushimi, R. Extracting Knowledge from Data through Catalysis Informatics. *ACS Catal.* **2018**, *8* (8), 7403−7429.

(17) Kitchin, J. R. Machine Learning in Catalysis. *Nat. Catal.* **2018**, *1* (4), 230−232.

(18) Suvarna, M.; Pérez-Ramírez, J. Embracing Data Science in Catalysis Research. *Nat. Catal.* **2024**, *7* (6), 624−635.

(19) Ma, J.; Chen, S.; Bellotti, P.; Guo, R.; Schäfer, F.; Heusler, A.; Zhang, X.; Daniliuc, C.; Brown, M. K.; Houk, K. N.; Glorius, F. Photochemical Intermolecular Dearomative Cycloaddition of Bicyclic Azaarenes with Alkenes. *Science* **2021**, *371* (6536), 1338−1345.

(20) Lückemeier, L.; De Vos, T.; Schlichter, L.; Gutheil, C.; Daniliuc, C. G.; Glorius, F. Chemoselective Heterogeneous Hydrogenation of Sulfur Containing Quinolines under Mild Conditions. *J. Am. Chem. Soc.* **2024**, *146* (9), 5864−5871.

(21) Davis, A. V.; Raymond, K. N. The Big Squeeze: Guest Exchange in an M_4L_6 Supramolecular Host. *J. Am. Chem. Soc.* **2005**, *127* (21), 7912−7919.

(22) Hong, C. M.; Morimoto, M.; Kapustin, E. A.; Alzakhem, N.; Bergman, R. G.; Raymond, K. N.; Toste, F. D. Deconvoluting the Role of Charge in a Supramolecular Catalyst. *J. Am. Chem. Soc.* **2018**, *140* (21), 6591−6595.

(23) Hart-Cooper, W. M.; Zhao, C.; Triano, R. M.; Yaghoubi, P.; Ozores, H. L.; Burford, K. N.; Toste, F. D.; Bergman, R. G.; Raymond, K. N. The Effect of Host Structure on the Selectivity and Mechanism of Supramolecular Catalysis of Prins Cyclizations. *Chem. Sci.* **2015**, *6* (2), 1383−1393.

(24) 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.

(25) 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.

(26) 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.

(27) 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.

(28) 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.

(29) Crawford, J. M.; Kingston, C.; Toste, F. D.; Sigman, M. S. Data Science Meets Physical Organic Chemistry. *Acc. Chem. Res.* **2021**, *54* (16), 3136−3148.

(30) Kelly, S. P.; Shende, V. V.; Flynn, A. R.; Dan, Q.; Ye, Y.; Smith, J. L.; Tsukamoto, S.; Sigman, M. S.; Sherman, D. H. Data Science-Driven Analysis of Substrate-Permissive Diketopiperazine Reverse

Prenyltransferase NotF: Applications in Protein Engineering and Cascade Biocatalytic Synthesis of (−)-Eurotiumin A. *J. Am. Chem. Soc.* **2022**, *144* (42), 19326−19336.
(31) Clements, H. D.; Flynn, A. R.; Nicholls, B. T.; Grosheva, D.; Lefave, S. J.; Merriman, M. T.; Hyster, T. K.; Sigman, M. S. Using Data Science for Mechanistic Insights and Selectivity Predictions in a Non-Natural Biocatalytic Reaction. *J. Am. Chem. Soc.* **2023**, *145* (32), 17656−17664.

https://doi.org/10.1021/jacs.4c11482
J. Am. Chem. Soc. 2024, 146, 29792−29800