Transferrable Selectivity Profiles Enable Prediction in Synergistic Catalyst Space

Yutao Kuang and Jolene P. Reid*

Department of Chemistry, University of British Columbia, 2036 Main Mall, Vancouver, British Columbia, V6T 1Z1, Canada

* jreid@chem.ubc.ca

Abstract

Organometallic intermediates participate in many multi-catalytic enantioselective transformations directed by a chiral catalyst, but the requirement of optimizing two catalyst components is a significant barrier to widely adopting this approach for chiral molecule synthesis. Algorithms can potentially accelerate the screening process by developing quantitative structure-function relationships from large experimental datasets. However, the chemical data available in this catalyst space is limited. We report a data-driven strategy that effectively translates selectivity relationships trained on enantioselectivity outcomes derived from one catalyst reaction systems where an abundance of data exists, to synergistic catalyst space. We describe three case studies involving different modes of catalysis (Brønsted acid, chiral anion, and secondary amine) that substantiate the prospect of this approach to predict and elucidate selectivity in reactions where more than one catalyst is involved. Ultimately, the success in applying our approach to diverse areas of asymmetric catalysis implies that this general workflow should find broad use in the study and development of new enantioselective, multi-catalytic processes.

Introduction

Small organic molecules provide effective catalysis for a significant number of reactions and in particular, have been essential to advances in the preparation of stereochemically pure compounds.[1-6] Often these transformations involve a single chiral catalyst that activates either substrate (nucleophile or electrophile) and in some cases both.[7-11] However, some transformations are still challenging or completely unobtainable using one catalyst systems. Accordingly, synergistic catalysis (also referred to as co-operative or multi-catalysis) in which more than one catalyst is involved in the activation of substrates is an important technique.[12-14] This approach has been immensely enabling for the discovery of new enantioselective, catalytic processes. Indeed, there are many examples of powerful classes of reactions that were developed upon the reliance of combining an organocatalyst with a metal.[15-22] Unfortunately, implementing this valuable tactic in asymmetric synthesis is often met with the formidable challenge of optimizing two catalyst components (as well as other reaction parameters) to achieve high levels of enantioselectivity.[23,24] While ideal reaction conditions have conventionally been discovered through empiricism, recent applications of data-driven reaction optimization have demonstrated that algorithms can streamline this process. For example, prior efforts from Denmark,[25] Doyle,[26] and Sigman[27] have highlighted how quantum descriptors, regression analysis, and chemical data from distinct reaction classes can be applied to correlate and predict experimental outcomes. Because multi-catalysis have been developed for individual reactions, datasets are difficult to combine meaning the data available to search for underlying chemical relationships in the multi-catalyst domain is scarce, presenting a significant limitation in applying these techniques.

On the basis of our recent efforts to deploy comprehensive multidimensional analysis to develop and leverage general mechanistic models,[28,29] we became interested in investigating if our workflows can be embedded in the optimization and quantitative prediction of reaction systems involving multiple catalysts (Figure 1A). The success of this approach is contingent on the ability to transfer the stereochemical information from one or more reactions facilitated by a single catalyst to another similar process that involves two catalyst systems (Figure 1B). Although multi-catalyst reaction designs share some common mechanistic features with single catalyst systems (e.g. optimal chiral catalyst structure), comparative studies that would reveal reaction specific contacts and general connections have not been performed. Such investigations would be valuable for formalizing mechanistic principles and considering the limits of model generality. As a result, despite the practical appeal of an approach that would preclude the requirement for explicit chemical data on synergistic reaction systems, the applicability domain of such statistical models would be challenging to estimate. We envisioned that the general mechanisms of stereoinduction in asymmetric catalysis should extend to multi-catalyst reaction strategies that focus on combining a chiral catalyst with a reactive intermediate that is generated from a second achiral catalyst species. Because these types of transformations aren't significantly affected by the presence of a second catalyst, they should be particularly amenable to this type of framework that uses data from one or more reactions to predict the result of a similar system.

In contemplating the various types of chiral catalysts and reactions that can participate in multi-catalysis, the effective assessment of this workflow should ideally involve more than one chemical system. Since statistical models describing the nucleophilic additions to iminiums catalyzed by chiral phosphoric acids and phosphates are easily accessible through previous reports these were both identified as suitable case studies for an initial evaluation. The second criteria in selecting a predictive platform is determining a chiral catalyst system that has been widely used in synergistic catalyst space such that significant validation data exists. Consequently, we identified reactions involving secondary amines as a third study. To this end, we develop and deploy multivariate linear regression (MLR) as a transferability method to achieve quantitative predictions and mechanistic analysis in diverse synergistic catalyst space (Figure 1C).

Figure 1. Application of statistical modeling workflow to multi-catalysis. (A) Transferring mechanistic patterns derived from a general reaction class for prediction in synergistic catalyst space. (B) Workflow for MLR analysis and further application in reaction systems involving more than one catalyst. (C) Overview of the study's goals to vet the techniques on three catalyst systems involving distinct catalytic modes of activation.

Results & Discussion

A. Assessment of Previously Reported Statistical Models. Rather than evaluate reactions involving multi-catalysts directly, we pursued a transfer learning strategy wherein we curated enantioselectivity data from various transformations deploying a single catalyst chemotype responsible for stereoinduction. By focusing reaction selection on those operating under a common catalytic mode of activation, reactions can be connected *via* general selectivity features revealed by regression analysis and a predictive model assembled. Since existing statistical models in chiral Brønsted acid[30] and anion catalyst space[29] are available for experimentation we first evaluated their ability to extrapolate to reactions facilitated by two catalysts. It should be noted that these models were applied without alteration to the identified parameters or data sets from their published forms.

In seeking an ambitious and relevant first test, we selected the hydrogenation of imines using molecular hydrogen (Figure 2).[31] Since this process cannot be simply facilitated solely by chiral organic molecules,[32] approaches have focused on two catalyst systems. Beller and co-workers demonstrated that Knölker's complex, a simple achiral iron hydrogenation catalyst, can be used in combination with a chiral Brønsted acid to provide enantioenriched secondary amines. Considering the overlap in structural features of the reaction components we anticipated that a previously generated statistical model constructed of chiral phosphoric acid catalyzed additions of nucleophiles to imines could be deployed to predict the reaction outcomes (Figure 2A).[30] In the previous study, reaction performance was first evaluated using a comprehensive model built from the entire data set constructed of reactions that proceeded through two different pathways, an *E* (+ee) or *Z* (-ee) transition state. While prediction errors were typically larger with this model, its use is imperative to determine the stereochemistry of the final product and the pathway under operation, allowing predictions to be refined with mechanism specific models (*E* or *Z*). While, the comprehensive model does not naturally allow prediction of stereochemistry, the product configuration can be assigned by applying the simple models in Figure 2A. These are based on the amine product generated from an *E* or *Z* TS and catalyzed by the (*S*)-CPA. The steps for ee prediction include: 1) locating the ground state

of the targeted reaction variable by DFT, 2) obtaining the key molecular features necessary for prediction, and 3) submitting these to both mathematical equations. For consistent results, the same level of theory that was used to optimize the organic nucleophiles should be applied to optimize the iron complex (organometallic nucleophile), however, the M06-2X density functional that was originally implemented is not applicable for molecules containing metals.[33] Instead, we employed M06/def2TZVP calculations which are suitable for organometallic systems and when tested against a subset of nine nucleophiles geometry minimizations provide the same value for the key bond angle and Sterimol B5 nucleophile terms when compared to M06-2X/def2TZVP (average deviation calculated to be 0.8° and 0.02 Å, respectively). Confident that this adjustment would not significantly impact the results we optimized the iron complex with this set of computational conditions and collected the necessary Sterimol B5 and angle nucleophile parameters for prediction from this structure. In other words, this computational method comparison suggested that a predictive model built from M06-2X descriptors could be used to predict the impact of hydrogenation reactions given the key nucleophile parameters calculated at the M06 level. Next, we evaluated the twenty reported hydrogenations involving aromatic imines catalyzed by TRIP. Both the catalyst and most of the imines were part of the published training set making the nucleophile (achiral iron complex) the only component not to be explicitly included. Each result was predicted using the comprehensive model, with an average absolute $\Delta\Delta G^{\ddagger}$ error of 0.64 kcal/mol (14 examples within 5% ee) and correctly assigned the absolute stereochemistry as *S*, demonstrating the ability of the model to extrapolate effectively to an organometallic nucleophilic intermediate (Figure 2B). A slightly improved outcome is observed using the *Z*-imine mechanistic model with a 0.48 kcal/mol average error (15 examples within 5% ee).
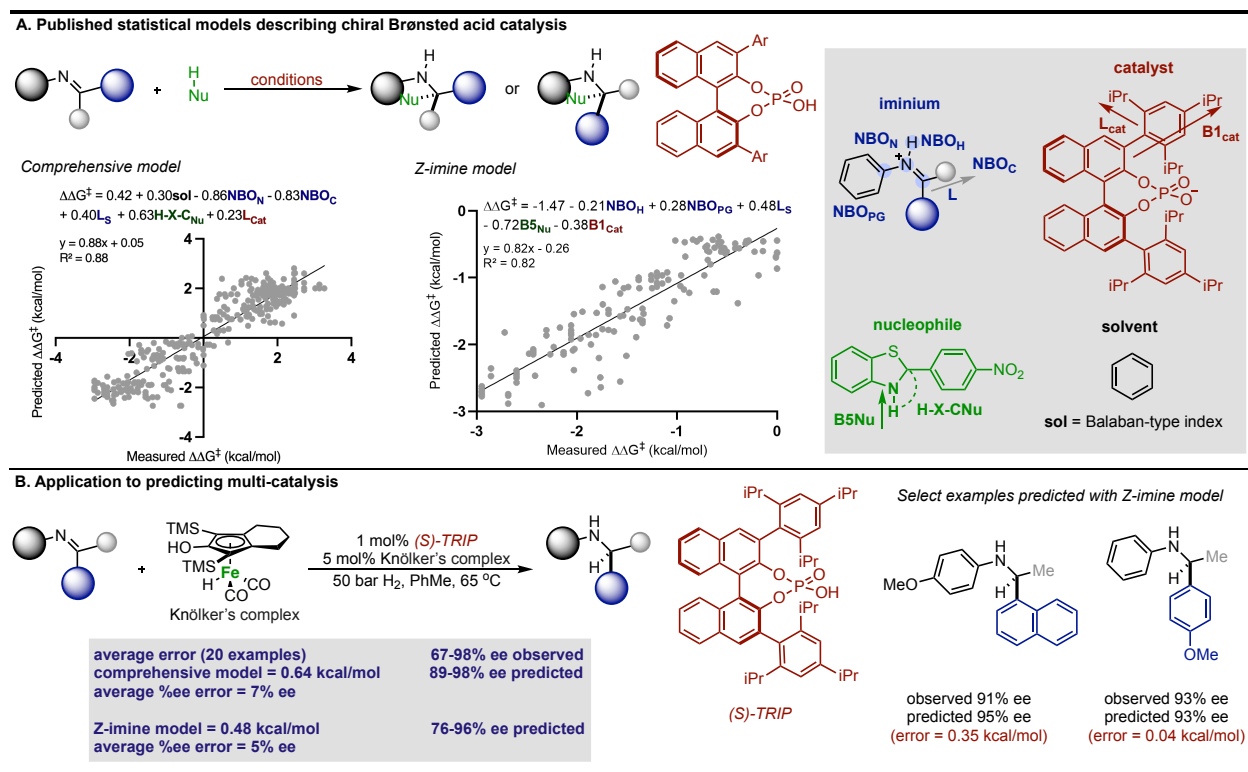
Figure 2. Prediction in chiral phosphoric acid catalyst space. (A) Key information on the published chiral Brønsted acid reaction models. A general amine product shows the stereochemical outcome predicted if the reaction proceeds via the *E* or *Z* TS and catalyzed by an (*S*)-CPA. (B) Extrapolation of previously reported chiral Brønsted acid imine reaction model to the iron catalyzed hydrogenation.

Inspired by the results obtained by analyzing the hydrogenation of imines, we selected to further evaluate the generality in our observations by interrogating chiral phosphate catalyzed reaction systems.[29] In considering this we noted that the addition of naphthols to gold(I) activated allenamides exhibited overlapping transition state features with our previously built statistical model i.e. combines an iminium with a nucleophile in the presence of a chiral phosphate (Figure 3).[34] As with the chiral phosphoric acid study, to deploy the published chiral phosphate model to predict the impact of utilizing an organometallic intermediate as a reaction component, the sensitivity of the previously identified parameters to computational method must be taken into account. Because we expected natural bond orbital (NBO) charges to be sensitive to the computational methods employed, we re-optimized each iminium intermediate at the M06/def2TZVP level to ensure all electrophilic components (organic and organometallic) were treated uniformly. After replacing the iminium parameters (NBO, Sterimol B1 and

polarizability) calculated at M06-2X/def2TZVP with those acquired from M06/def2TZVP optimized structures, the statistical model was re-created in MATLAB using the same enantioselectivity data and identified parameters from the previous publication (Figure 3A). This model could then be deployed to predict the organometallic data set (Figure 3B). However, this is a more challenging scenario, as the structural overlap between the training and the prediction set is slightly reduced. More specifically, the reaction components to be predicted are not explicitly included in the training data but belong to general subclasses of iminiums, naphthols, and chiral phosphates. Next, the key iminium parameters (NBO, Sterimol B1 and polarizability) were collected from the electrophilic gold complex for prediction. Again, accurate predictions were construed with the statistical model with an average absolute $\Delta\Delta G^{\ddagger}$ error of 0.33 kcal/mol (4 examples within 2% ee and 6 examples within 5% ee). The presence of a methyl group had the most detrimental effect on selectivity (4% ee experimental, -11% ee predicted) and excitingly, this could be accurately captured by the model. While superficially surprising that the model can successfully anticipate significant enantioselectivity changes due to minor substrate modifications (i.e. switching a phenyl for a methyl), close examination of key parameters in the model reveals that the lower enantioselectivity for this substrate can be attributed to the more positive NBO (Figure 3C, gold intermediate A).
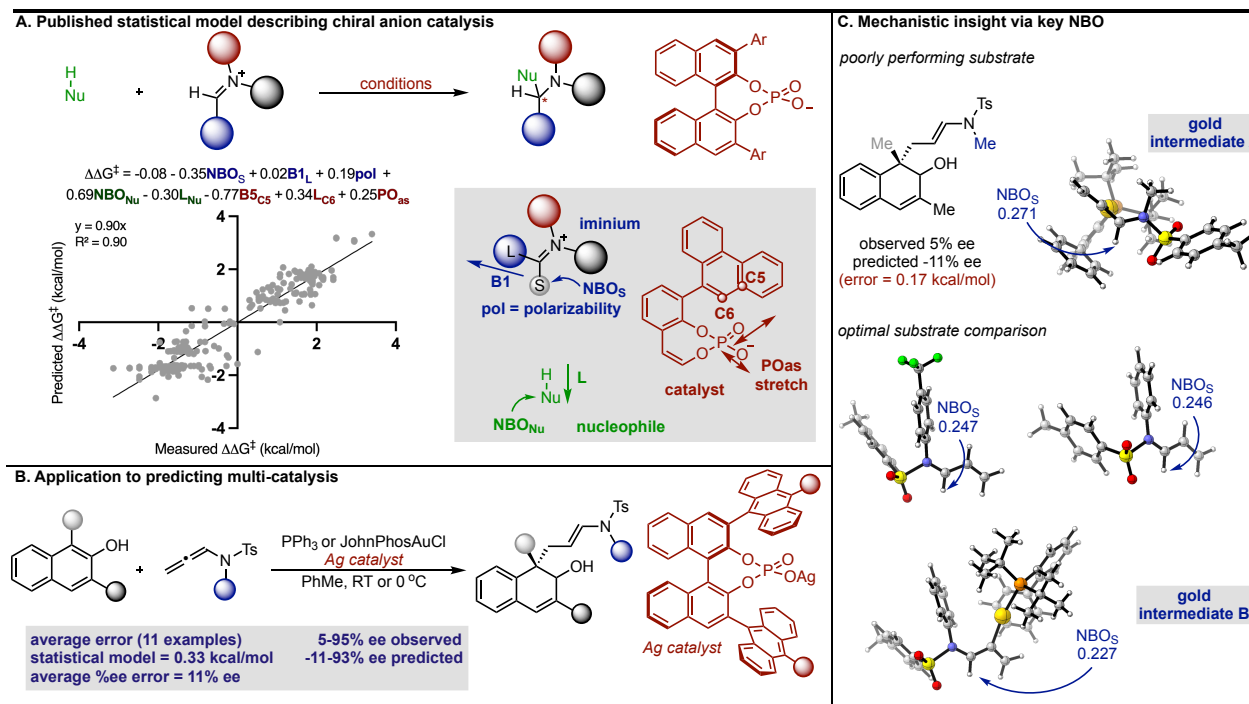
Figure 3. Prediction and mechanistic interrogation in chiral phosphate catalyzed reaction systems. (A) Key information on the published chiral phosphate reaction models. (B) Application of the previously reported chiral phosphate iminium reaction model to the gold catalyzed dearomatization of naphthols. (C) Interpretation of key NBO charges to describe poorly performing substrates.

Furthermore, the ability of the model to accurately reflect the outcomes with different substrates may allow for effective and efficient reaction scope exploration. Perhaps the most powerful analysis of the model is illustrated by comparing the substrate profiles of the one (chiral phosphate only)[35] and two (chiral phosphate combined with gold) catalyst systems. Remarkably, the optimal iminium intermediate was reversed between the two methods. In other words, the lead substrate with chiral phosphate catalysis failed to provide high enantioselectivities under the gold conditions and *vice versa*. Once again, the model clearly explains why certain substrates should be particularly amenable to different protocols. Under chiral phosphate catalysis the NBO$_S$ values are comparable and the difference in polarizability, the second important iminium term, explains the contrast in enantioselectivity. Under gold catalysis, the more negative NBO values associated with gold intermediate B, largely compensate for a slightly lower polarizability term, and the ee is increased (Figure 3C).

B. Secondary Amine Model Development. After evaluating the two published statistical models in chiral phosphoric acid and phosphate catalyzed reaction space, the second stage of this study was directed at evaluating a wider set of synergistic reactions involving secondary amines. To accomplish this, a comprehensive MLR model that relates the features of all of the reaction components to the experimentally obtained enantioselectivity outcomes conveyed as DDG‡ for this catalyst class would be required (see SI for full details).

Despite the potential for extensive catalyst structure modulation, only a limited set of secondary amines have witnessed broad application. This is in contrast to the many other catalyst chemotypes employed in asymmetric synthesis where necessary and extensive optimization efforts have generated considerably sized catalyst libraries. Thus, the most significant challenge in the early stages of implementing our workflow was defining a useful data set containing both high and low enantioselectivities for model construction. Consequently, to supplement our data mining efforts on published data from scientific journals, we explored the use of publicly available Ph.D. theses which typically contain experimental data of high quality (i.e. satisfy the degree requirements) but remain unpublished presumably because the research objective of delivering the product in high enantioselectivity was not met.[36] Throughout this literature evaluation, we strategically avoided two types of reaction examples. First we ignored reactions that showed product racemization to be strongly contributing to the overall enantioselectivity outcome (i.e. time and temperature sensitive).[37,38] In the absence of strongly supporting experimental data, it is only possible to minimize rather than eliminate the influence of such effects in our analysis through the removal of unusual experimental results.[39] Consequently, some proportion of variation between measured and predicted enantioselectivity values will likely be attributable to these factors in addition to experimental and analytical error. Secondly, we did not include examples that combined proline type catalysts with reactants that did not contain strongly electronegative atoms. In these cases, the structure of the reactant would make it difficult to determine if hydrogen bonding was directing its approach and therefore, hard to assign the mode of enantioinduction (in more detail below).

On this basis, to construct a predictive model, an expanded inventory of 452 reactions with varied components was curated from multiple sources (for a list of references see SI). From this survey, we categorized the dataset by general catalyst structure which is a significant factor in determining the mode of enantioinduction (steric blocking, TSA or hydrogen bonding, TSB) wherein imidazolidinone[40] *and* diphenylprolinol ethers[41] are grouped by a +ee value and proline type catalysts, *a* -ee value (Figure 4A). Therefore, the sign of the ee represents one of two transition state (TS) categories, depending on the catalyst involved. This is important in understanding the product enantioselectivities, because reactant addition to the top or bottom face will produce opposite enantiomers. Accordingly, the statistical model will be able to determine whether the reaction proceeds through steric blocking (predicts +ee value) or hydrogen bonding modes of enantioinduction (predicts -ee value) and this information can be used to make predictions about the absolute configurations by using the qualitative pictures TSA or TSB. Furthermore, the TS-guided categorization strategy is useful in producing a well-distributed data set which would be hard to achieve by not taking into account the absolute product stereochemistry. Next, a diverse array of molecular descriptor values were collected from DFT optimized geometries to describe the overlapping structural features of each electrophile, nucleophile, catalyst, solvent, and co-catalyst.[42,43] Because this model is being built with the sole aim of predicting reactions involving an organometallic intermediate, we naturally choose the appropriate computational methods from the beginning. This involved optimizing the reactant at the M06/def2TZVP level and all other components with M06-2X/def2TZVP. The commonality in substrate and catalyst substructure allowed collection of natural bond orbital (NBO) charges and Sterimol values from the conserved portions. However, the nucleophile had minimal structure overlap, thus, polarizability, highest occupied molecular orbital (HOMO), and lowest unoccupied molecular orbital (LUMO) energies, which do not rely on common substructures, were collected to describe this component. Unfortunately, the lack of consistency in the reaction conditions renders the identification of readily comprehensible and extensive parameter sets for the remaining components a challenge. For example, several reactions required Brønsted acid co-catalysts and employed solvent mixtures while many others did not. Guided by the proposed mechanism of catalysis, we postulated that in cases where the acid additive was absent, the proton could originate from another source, a reagent or catalyst, and relevant descriptors could be collected from these components. Because solute-solvent interactions with polar substances will likely dominate over those with non-polar molecules, we collected topological, two-dimensional descriptors from the solvent with the largest dielectric constant (see SI).[44,45]

Prior to model building, the data set was partitioned into 80:20 training:validation sets using MATLAB's deterministic equidistant splitting function. Linear regression algorithms were then applied to the training set (80% of the entire data set that incorporates both +ee and -ee reactions) to identify prospective correlations between the molecular structure of every reaction variable defined by the parameters collected in the previous step of the workflow and the measured enantioselectivity, $DDG^{\ddagger}$ (where $DDG^{\ddagger} = -RTln(e.r.)$ and T is the temperature at which the reaction was performed). Since the training set includes significant diversity in reaction component structure and mechanism, we anticipated that several descriptors would be required to achieve predictive correlations. Using forward stepwise linear regression[46] we determined a model that includes solvent (black), substrate (blue), reactant (green), and catalyst (red) terms distributed over thirteen parameters to be appropriate. Despite the high R^2 value and validation scores, a relatively small number of outliers appeared around 0 kcal/mol on the x-axis. Essentially, these correspond to a few reactions that provided almost racemic mixtures by experiment. Such unique reaction features will not conform to trends revealed by comprehensive MLR models as these operate by linking reactions via general connections i.e. structural effects that apply to the majority of reactions included in the data set .

Previous computational studies show that the enantioselectivity arises from the geometry of the enamine/iminium double bond (*s-trans* or *s-cis*) and the approach of the reactant (top or bottom).[47]
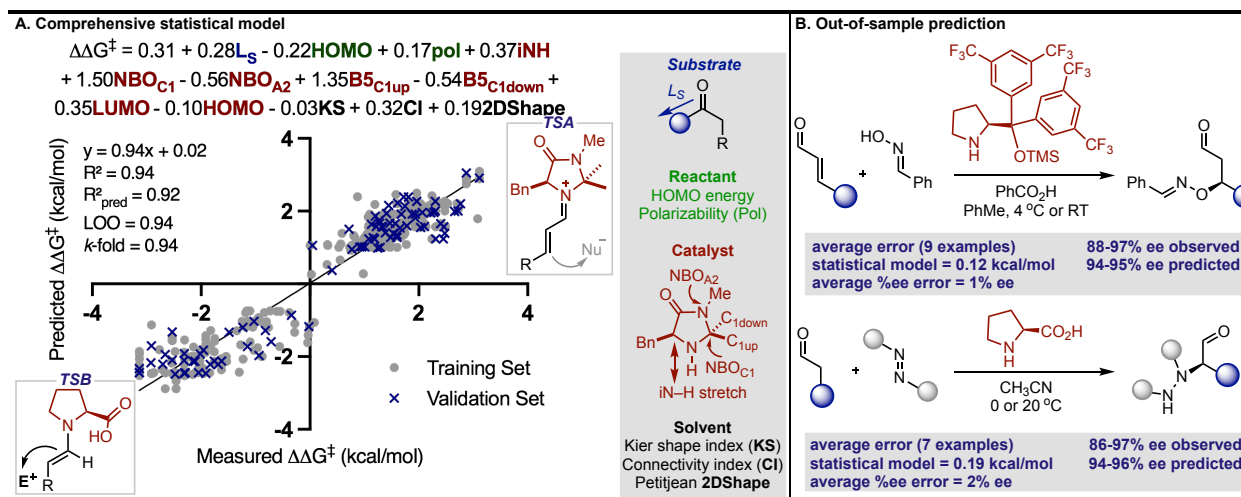
Figure 4. Comprehensive model development and validation. (A) Regression model trained on 361 data points and validated with the remaining 91. (B) Demonstration of mechanistic transferability by predicting enantioselectivity outcomes involving the hydroxylation and amination of aldehydes

Therefore, it is possible that the mathematical model also reveals some of these mechanistic features despite the complex equation. Notably, the catalyst descriptors have the largest coefficients in the normalized equation, demonstrating that stereocontrol is dominated by catalyst architecture for this class of reactions. The presence of $B5_{C1(up)}$ lends to a straightforward analysis by implying that larger substituents at this position makes reactant approach to the double bond from the top less possible. Additionally, large groups at C1 would direct the double bond to occupy the opposite side of the catalyst thereby favoring the formation of the *s-trans* isomer. We interpreted the inclusion of the NBO_{C1} term as a categorical descriptor that essentially highlights that proline type catalysts (typically negative NBO_{C1} and -ee) direct the reactant to the top via a hydrogen bonding interaction whereas steric blocking catalysts (usually positive NBO_{C1} and +ee) that incorporate large alkyl or aromatic groups at this position promote reaction on the opposite face. Importantly, the presence of $B5_{C1(down)}$ with a negative coefficient likely indicates that TS_{major} is also sensitive to the catalyst features. In other words, larger substituents at this position may enhance repulsive interactions between catalyst and reactant in the TS that forms the major product, ultimately favouring the formation of the opposite enantiomer. Indeed, imidazolidinone catalysts that have two large groups at the C1 position generally provide lower levels of enantioselectivity supporting this assertion. The *s-trans/s-cis* isomer ratio also depends on the substrate and having inversely sized groups on either side of the carbonyl will strongly reinforce the preference for the s-trans. This is expressed by the L_S model term and reflected in the lower enantioselectivities obtained for ketones compared to aldehydes. Because ketones are predominantly used in combination with proline (i.e. result in -ee reactions) the associated coefficient with L_S is positive. The role of the reactant is described through polarizability which likely acts as a proxy for chemical size (see SI) and the HOMO energy. The relationship also includes three solvent parameters with relatively small coefficients suggesting most solvents are compatible and the assortment of optimal solvents is a reflection of reaction component solubility.[48,49] Overall and perhaps most intriguingly, the incorporated model terms suggest that catalyst substrate matching is not required explaining the generality in catalyst structure.

Considering the goal of the prediction analysis is to transfer enantioselectivity trends of one catalyst systems to multi-catalysis, the next step in validating the secondary amine model the ability of extrapolating to new reaction types involving a single catalyst (Figure 4B). Thus, for each out-of-sample prediction platform, both the catalyst and substrate are contained in our training set (see SI for additional out-of-sample predictions). It should be noted that the model can only make predictions about the reaction enantioselectivity and not the diastereoselectivity if two chiral centers are created. This second aspect of selectivity arises from the orientation of the reactant substituents relative to those on the enamine/iminium and may be governed by a different set of molecular features. However, diastereoselectivity is typically not the experimental output that requires significant optimization and high levels are usually observed regardless of the reaction conditions. Thus, the prediction of diastereoselectivity is not crucial for reaction development providing the incentive to exclude this output from our regression analysis. As a first assessment, we evaluated the ability to predict nine hydroxylation reactions, involving an oxime and diphenylprolinol ether.[50] This set was predicted accurately, with an average absolute $\Delta\Delta G^{\ddagger}$ error of 0.12 kcal/mol (eight examples predicted within 5% ee). By using the simple reaction model, TSA, presented in Figure 4A the
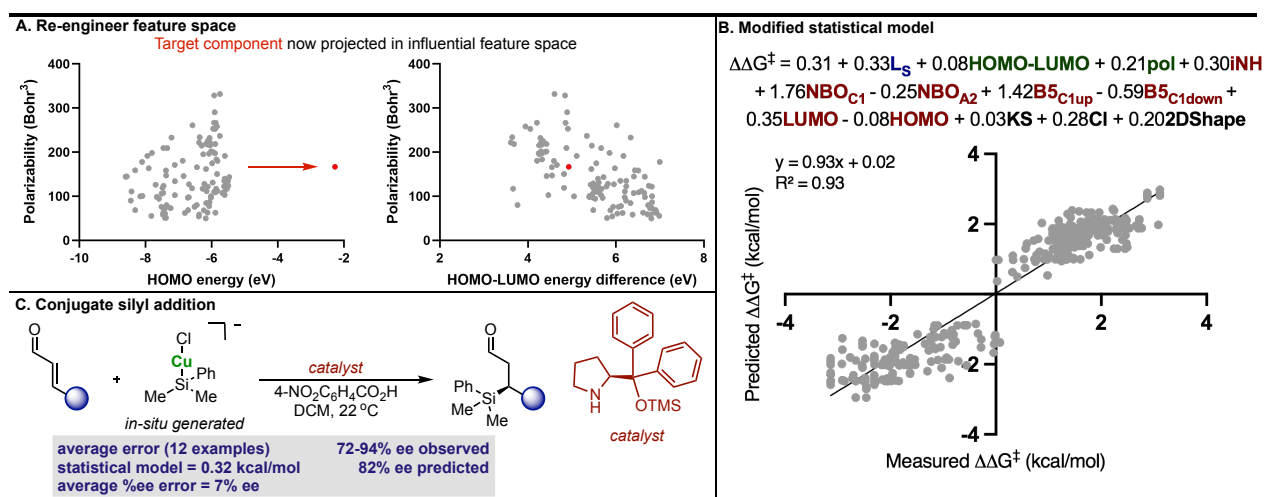
absolute configuration is correctly assigned as R. As a second case study, the model was assessed in the same manner with seven proline catalyzed amination reactions involving an azodicarboxylate.[51] Again accurate predictions were obtained with this statistical model ($\Delta\Delta G^{\ddagger}$ error of 0.19 kcal/mol, six examples predicted within 5% ee) with the qualitative diagram, TSB confidently determining the stereochemical outcome to be R.

C. Application to synergistic catalysis. With our secondary amine statistical model thoroughly validated, we next sought to test its performance in the prediction of synergistic reaction systems involving secondary amines (Figures 5 and 6).
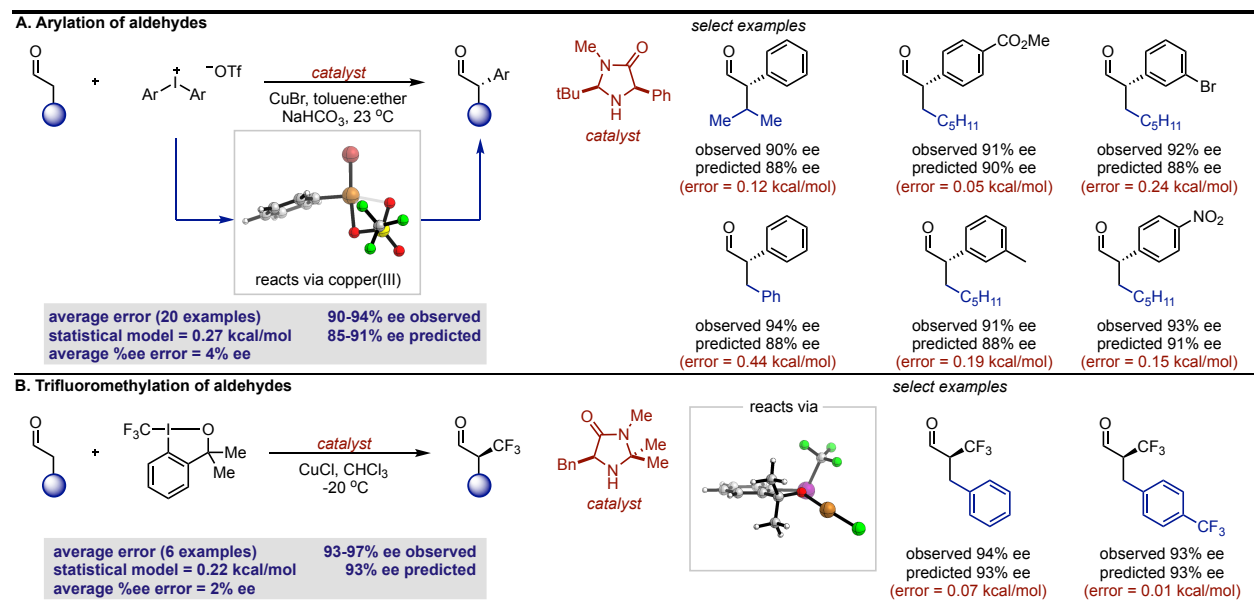
Figure 5. Applying the secondary amine regression model to predict reaction outcomes involving copper catalyzed intermediates. (A) Assessing prediction capabilities with the arylation of aldehydes. (B) Effective prediction of trifluoromethylation reactions.

Because copper can both generate electrophilic and nucleophilic reactive intermediates complementing the reactivity profile of enamide and iminium catalysis, there are a number of examples in the chemical literature where this type of merger is employed. We focused on the arylation and trifluoromethylation of aldehydes reported by MacMillan as representative systems involving enamide.[52-53] The predictions obtained from the model are shown in Figure 5 alongside the experimental results, and satisfyingly, the agreement was generally excellent. More specifically, the first case utilizes an aryl copper (III) species and a catalyst not included in the training set. With a novel catalyst/reactant pairing, an average $\Delta\Delta G^{\ddagger}$ error of 0.27 kcal/mol over twenty examples was determined (Figure 5A). As exemplified in the second case, copper can also be employed as a Lewis acid to increase the reactivity of the electrophilic trifluoromethylation agent and in the presence of an imidazolidinone we predicted six reactions with an average $\Delta\Delta G^{\ddagger}$ error of 0.22 kcal/mol (Figure 5B). Because the model only incorporates a single substrate parameter that essentially classifies if an aldehyde or ketone was employed, the model correctly predicts that large changes in aldehyde structure leads to small changes in the observed $\Delta\Delta G^{\ddagger}$ for both cases. As before, TSA displayed in Figure 4A can be applied to correctly assign the stereochemistry as S for both examples. To test this approach on reactions proceeding via iminium intermediates, the copper catalyzed silyl addition was probed.[54] On collecting the key reactant HOMO parameters required for prediction we detected that the values were significantly different from those included in the training set. Based on the premise that we can predict reactions most similar to that of the training set, we hypothesized that proximity in the chemical space representation provided by mapping HOMO against polarizability (i.e. influential feature space) would correspond to accuracy in out-of-sample prediction. Ultimately, our plot shown in Figure 6A suggested extrapolation to this reaction component would lead to large errors in predicting the enantioselectivity. This prompted us to search for an alternative descriptor that would capture the reactant in influential feature space. Since the HOMO-LUMO energy gap is correlative to the original parameter we generated property maps including this descriptor. Intriguingly, these indicated that the organometallic intermediate is now projected in the same feature space as the training set (Figure 6A). Next, we manually altered the model by replacing the HOMO energy term for the HOMO-LUMO difference and predicted the enantioselectivity outcomes (Figure 6B). Each result was predicted using the modified model, with an average absolute $\Delta\Delta G^{\ddagger}$ error of 0.32 kcal/mol (ten examples within 10% ee) and correctly assigned the absolute stereochemistry as *S* (Figure 6C). This result is compelling in that we could rationally re-engineer the influential features to generalize the statistical model across diverse reaction space.

A. Re-engineer feature space

Target component now projected in influential feature space

B. Modified statistical model

$$\Delta\Delta G^{\ddagger} = 0.31 + 0.33\mathbf{L_S} + 0.08\mathbf{HOMO\text{-}LUMO} + 0.21\mathbf{pol} + 0.30\mathbf{iNH}$$
$$+ 1.76\mathbf{NBO_{C1}} - 0.25\mathbf{NBO_{A2}} + 1.42\mathbf{B5_{C1up}} - 0.59\mathbf{B5_{C1down}} +$$
$$0.35\mathbf{LUMO} - 0.08\mathbf{HOMO} + 0.03\mathbf{KS} + 0.28\mathbf{Cl} + 0.20\mathbf{2DShape}$$

$y = 0.93x + 0.02$
$R^2 = 0.93$

C. Conjugate silyl addition

in-situ generated

catalyst
4-NO$_2$C$_6$H$_4$CO$_2$H
DCM, 22 °C

catalyst

average error (12 examples)   72-94% ee observed
statistical model = 0.32 kcal/mol   82% ee predicted
average %ee error = 7% ee

Figure 6. Applying the secondary amine regression model to predict reaction outcomes involving copper catalyzed intermediates. (A) Assessing prediction capabilities with the arylation of aldehydes. (B) Effective prediction of trifluoromethylation reactions.

Conclusion

Here, we describe three case studies involving different modes of catalysis that demonstrate the benefits of utilizing MLR as a transferability tool to predict and elucidate enantioselectivity outcomes in reactions where more than one catalyst is involved. Specifically, our strategy focused on revealing general mechanistic models produced through extensive data mining and advanced parameter sets. Because the selectivity discriminants were consistent across a reaction range, the resulting correlation could be leveraged for the translation of experimental observations derived from reactions utilizing a single catalyst to another similar process that involves two catalyst systems. In general, we expect this transferability workflow to be valuable in data limiting situations, for example where practitioners have incomplete data sets either early in an optimization campaign or the complexity of reaction conditions makes it difficult to explore reaction component space completely. Consequently, our findings should be broadly applicable and beneficial for the prediction and interrogation of other catalytic systems widely applied in asymmetric synthesis.

ACKNOWLEDGEMENT

REFERENCES

(1) Knowles, R. R.; Jacobsen E. N. Attractive Noncovalent Interactions in Asymmetric Catalysis: Links between Enzymes and Small Molecule Catalysts. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 20678–20685.

(2) Yoon, T. P.; Jacobsen, E. N. Privileged Chiral Catalysts. *Science* **2003**, *299*, 1691–1693.

(3) Berkessel, A.; Groger, H. *Asymmetric Organocatalysis: From Biomimetic Concepts to Applications in Asymmetric Synthesis*; Wiley-VCH, 2005.

(4) Akiyama, T. Stronger Brønsted acids. *Chem. Rev*. **2007**, *107*, 5744–5758.

(5) Doyle, A. G.; Jacobsen, E. N. Small-Molecule H-Bond Donors in Asymmetric Catalysis. *Chem. Rev.* **2007**, *107*, 5713−5743.

(6) Flanigan, D. M.; Romanov-Michalidis, F.; White, N. A.; Rovis, T. Organocatalytic Reactions Enabled by N-Heterocyclic Carbenes. *Chem. Rev.* **2015**, *115*, 9307−9387.

(7) Walsh, P. J.; Kozlowski, M. C. *Fundamentals of Asymmetric Catalysis*; University Science Books, 2008.

(8) Yamamoto, H. *Lewis Acids in Organic Synthesis*; Wiley-VCH, 2000.

(9) Phipps, R. J.; Hamilton, G. L.; Toste, F. D. The Progression of Chiral Anions from Concepts to Applications in Asymmetric Catalysis. *Nat. Chem.* **2012**, *4*, 603−614.

(10) Houk, K. N.; Cheong, P. H. Y. Computational prediction of small-molecule catalysts. *Nature* **2008**, *455*, 309–313.

(11) Reid, J. P.; Simón, L.; Goodman, J. M. A Practical Guide for Predicting the Stereochemistry of Bifunctional Phosphoric Acid Catalyzed Reactions of Imines. *Acc. Chem. Res.* **2016**, *49*, 1029−1041.

(12) Allen, A. E.; MacMillan, D. W. C. Synergistic catalysis: A powerful synthetic strategy for new reaction development. *Chem. Sci.* **2012**, *3*, 633−658.

(13) Kim, U. Bin; Jung, D. J.; Jeon, H. J.; Rathwell, K.; Lee, S. G. Synergistic Dual Transition Metal Catalysis. *Chem. Rev.* **2020**, *120*, 13382−13433.

(14) Martínez, S.; Veth, L.; Lainer, B.; Dydio, P. Challenges and Opportunities in Multicatalysis. *ACS Catal.* **2021**, *11*, 3891−3915.

(15) Shao, Z.; Zhang, H. Combining transition metal catalysis and organocatalysis: a broad new concept for catalysis. *Chem. Soc. Rev.* **2009**, *38*, 2745−2755.

(16) Du, Z.; Shao, Z. Combining transition metal catalysis and organocatalysis - an update. *Chem. Soc. Rev.* **2013**, *42*, 1337−1378.

(17) Bradley G. Jellerichs; Jong-Rock Kong, A.; Krische, M. J. Catalytic Enone Cycloallylation via Concomitant Activation of Latent Nucleophilic and Electrophilic Partners: Merging Organic and Transition Metal Catalysis. *J. Am. Chem. Soc.* **2003**, *125*, 7758−7759.

(18) Schwarz, K. J.; Amos, J. L.; Klein, J. C.; Do, D. T.; Snaddon, T. N. Uniting C1-Ammonium Enolates and Transition Metal Electrophiles via Cooperative Catalysis: The Direct Asymmetric α-Allylation of Aryl Acetic Acid Esters. *J. Am. Chem. Soc.* **2016**, *138*, 5214−5217.

(19) Jiang, X.; Beiger, J. J.; Hartwig, J. F. Stereodivergent Allylic Substitutions with Aryl Acetic Acid Esters by Synergistic Iridium and Lewis Base Catalysis. *J. Am. Chem. Soc.* **2017**, *139*, 87−90.

(20) Afewerki, S.; Cordova, A. Combinations of Aminocatalysts and Metal Catalysts: A Powerful Cooperative Approach in Selective Organic Synthesis. *Chem. Rev.* **2016**, *116*, 13512−13570.

(21) Rueping, M.; Koenigs, R. M.; Atodiresei, I. Unifying Metal and Brønsted Acid Catalysis –Concepts, Mechanisms, and Classifications. *Chem. - Eur. J.* **2010**, *16*, 9350−9365.

(22) Wang, M. H.; Scheidt, K. A. Cooperative Catalysis and Activation with N-Heterocyclic Carbenes. *Angew. Chem., Int. Ed.* **2016**, *55*, 14912−14922.

(23) Robbins, D. W.; Hartwig, J. F. A Simple, Multidimensional Approach to High-Throughput Discovery of Catalytic Reactions. **2011**, *333*, 1423–1427.

(24) McNally, A.; Prier, C. K.; MacMillan, D. W. C. Discovery of an alpha-C-H Arylation Reaction Using the Strategy of Accelerated Serendipity. *Science* **2011**, *334*, 1114–1117.

(25) Rinehart, N. I.; Zahrt, A. F.; Henle, J. J.; Denmark, S. E. Dreams, False Starts, Dead Ends, and Redemption: A Chronicle of the Evolution of a Chemoinformatic Workflow for the Optimization of Enantioselective Catalysts. *Acc. Chem. Res.* **2021**, *54*, 2041–2054.

(26) Żurański, A. M.; Martinez Alvarado, J. I.; Shields, B. J.; Doyle, A. G. Predicting Reaction Yields via Supervised Learning. *Acc. Chem. Res.* **2021**, *54*, 1856–1865.

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

(28) Shoja, A.; Reid, J. P. Computational Insights into Privileged Stereocontrolling Interactions Involving Chiral Phosphates and Iminium Intermediates. *J. Am. Chem. Soc.* **2021**, *143*, 7209−7215.

(29) Shoja, A.; Zhai, J.; Reid, J. P. Comprehensive Stereochemical Models for Selectivity Prediction in Diverse Chiral Phosphate-Catalyzed Reaction Space. *ACS Catal.* **2021**, *11*, 11897−11905.

(30) Reid, J. P.; Sigman, M. S. Holistic prediction of enantioselectivity in asymmetric catalysis. *Nature* **2019**, *572*, 343−348.

(31) Zhou, S.; Fleischer, S.; Junge, K.; Beller, M. Cooperative Transition-Metal and Chiral Brønsted Acid Catalysis: Enantioselective Hydrogenation of Imines to Form Amines. *Angew. Chem., Int. Ed.* **2011**, *50*, 5120−5124.

(32) Chase, P. A.; Welch, G. C.; Jurca, T.; Stephan, D. W. Metal-Free Catalytic Hydrogenation. *Angew. Chem., Int. Ed.* **2007**, *46*, 8050−8053.

(33) Zhao, Y.; Truhlar, D. G. Density Functionals with Broad Applicability in Chemistry. *Acc. Chem. Res.* **2008**, *41*, 157−167.

(34) Pedrazzani, R.; An, J.; Monari, M.; Bandini, M. New Chiral BINOL-Based Phosphates for Enantioselective [Au(I)]-Catalyzed Dearomatization of β-Naphthols with Allenamides. *Eur. J. Org. Chem.* **2021**, *11*, 1732−1736.

(35) Yang, B.; Zhai, X.; Feng, S.; Hu, D.; Deng, Y.; Shao, Z. Organocatalyzed Intermolecular Asymmetric Allylic Dearomatization of Both α- and β-Naphthols. *Org. Lett.* **2019**, *21*, 330−334.

(36) Andrews, D. M.; Broad, L. M.; Edwards, P. J.; Fox, D. N.; Gallagher, T.; Garland, S .L.; Sweeney, J. B. The creation and characterisation of a national compound collection: the royal society of chemistry pilot. *Chem. Sci.* **2016**, *7*, 3869–3878.

(37) Halland, N.; Braunton, A.; Bachmann, S.; Marigo, M.; Jørgensen, K. A. Direct Organocatalytic Asymmetric α-Chlorination of Aldehydes. *J. Am. Chem. Soc.* **2004**, *126*, 4790−4791.

(38) Marigo, M.; Fielenbach, D.; Braunton, A.; Kjærsgaard, A.; Jørgensen, K. A. Enantioselective Formation of Stereogenic Carbon-Fluorine Centers by a Simple Catalytic Method. *Angew. Chem., Int. Ed.* **2005**, *44*, 3703−3706.

(39) Burés, J.; Armstrong, A.; Blackmond, D. G. Explaining Anomalies in Enamine Catalysis: "Downstream Species" as a New Paradigm for Stereocontrol. *Acc. Chem. Res.* **2016**, *49*, 214–222.

(40) Lelais, G.; MacMillan, D. W. C. Modern Strategies in Organic Catalysis: The Advent and Development of Iminium Activation. *Aldrichimica Acta* **2016**, *39*, 78–88.

(41) Palomo, C.; Mielgo, A. Diarylprolinol ethers: Expanding the potential of enamine/iminium-ion catalysis. *Angew. Chem., Int. Ed.* **2006**, *45*, 7876–7880.

(42) Santiago, C. B.; Guo, J. Y.; Sigman, M. S. Predictive and mechanistic multivariate linear regression models for reaction development. *Chem. Sci.* **2018**, *9*, 2398–2412.

(43) 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*, 827−836.

(44) Amar, Y.; Schweidtmann, A. M.; Deutsch, P.; Cao, L.; Lapkin, A. Machine learning and molecular descriptors enable rational solvent selection in asymmetric catalysis. *Chem. Sci.* **2019**, *10*, 6697−6706.

(45) Metsanen, T. T.; Lexa, K. W.; Santiago, C. B.; Chung, C. K.;¨ Xu, Y.; Liu, Z.; Humphrey, G. R.; Ruck, R. T.; Sherer, E. C.; Sigman, M. S. Combining Traditional 2D and Modern Physical Organic-derived Descriptors to Predict Enhanced Enantioselectivity for the Key AzaMichael Conjugate Addition in the Synthesis of Prevymis (Letermovir). *Chem. Sci.* **2018**, *9*, 6922−6928.

(46) Harrell, F. E. Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis; Springer-Verlag, 2001.

(47) Cheong, P. H. Y.; Legault, C. Y.; Um, J. M.; Çelebi-Ölçüm, N.; Houk, K. N. Quantum Mechanical Investigations of Organocatalysis: Mechanisms, Reactivities, and Selectivities. *Chem. Rev.* **2011**, *111*, 5042−5137.

(48) Burés, J.; Dingwall, P.; Armstrong, A.; Blackmond, D. G. Rationalization of an Unusual Solvent-Induced Inversion of Enantiomeric Excess in Organocatalytic Selenylation of Aldehydes. *Angew. Chem., Int. Ed.* **2014**, *53*, 8700–8704.

(49) Hutchinson, G.; Alamillo-Ferrer, C.; Burés, J. Mechanistically Guided Design of an Efficient and Enantioselective Aminocatalytic α-Chlorination of Aldehydes. *J. Am. Chem. Soc.* **2021**, *143*, 6805−6809.

(50) Bertelsen, S.; Diner, P.; Johansen, R. L.; Jørgensen, K. A. Asymmetric Organocatalytic β-Hydroxylation of α, β-Unsaturated Aldehydes. *J. Am. Chem. Soc.* **2007**, *129*, 1536–1537.

(51) List, B. Direct Catalytic Asymmetric α-Amination of Aldehydes. *J. Am. Chem. Soc.* **2002**, *124*, 5656−5657.

(52) Allen, A. E.; MacMillan, D. W. C. Enantioselective α-Arylation of Aldehydes via the Productive Merger of Iodonium Salts and Organocatalysis. *J. Am. Chem. Soc.* **2011**, *133*, 4260−4263.

(53) Allen, A. E.; MacMillan, D. W. C. The Productive Merger of Iodonium Salts and Organocatalysis: A Non-photolytic Approach to the Enantioselective α-Trifluoromethylation of Aldehydes. *J. Am. Chem. Soc.* **2011**, *132*, 4986–4987.

(54) Ibrahem, I.; Santoro, S.; Himo, F.; Córdova, A. Enantioselective Conjugate Silyl Additions to α, β-Unsaturated Aldehydes Catalyzed by Combination of Transition Metal and Chiral Amine Catalysts. *Adv. Synth. Catal.* **2011**, *353*, 245−252.