ACS Catalysis

pubs.acs.org/acscatalysis Research Article

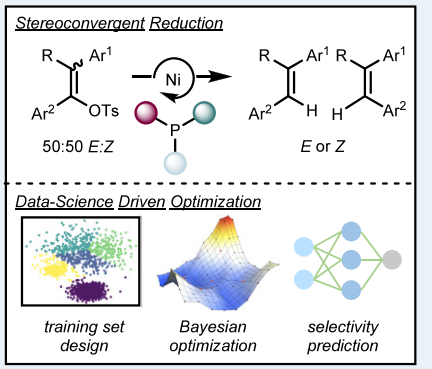
Data Science Guided Multiobjective Optimization of a Stereoconvergent Nickel-Catalyzed Reduction of Enol Tosylates to Access Trisubstituted Alkenes

Natalie P. Romer, Daniel S. Min, Jason Y. Wang, Richard C. Walroth, Kyle A. Mack, Lauren E. Sirois, Francis Gosselin, Daniel Zell,* Abigail G. Doyle,* and Matthew S. Sigman*

Cite This: ACS Catal. 2024, 14, 4699–4708 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Herein we report a method for a stereoconvergent synthesis of trisubstituted alkenes in two steps from simple ketone starting materials. The key step is a nickel-catalyzed reduction of the corresponding enol tosylates that predominantly relies on a monophosphine ligand to direct the stereoconvergent formation of either the *E*- or *Z*-trisubstituted alkene products. Reaction optimization was accomplished using a data science workflow including monophosphine training set design, statistical modeling, and multiobjective Bayesian optimization. The optimization campaign significantly improved access to both the *E*- and *Z*-trisubstituted products in up to ∼90:10 diastereoselectivity and >90% yield. After identifying superior ligands using training set design, only 25 reactions were required for each objective (*E*- and *Z*-isomer formation) to converge on improved reaction parameters from a search space of ∼30,000 potential conditions using the EDBO+ platform. Additionally, a hierarchical machine learning model was developed to predict the stereoselectivity of untested monophosphine ligands to achieve a validation mean absolute error (MAE) of 7.1% selectivity (0.21 kcal/mol). Ultimately, we present a synergistic data science workflow leveraging the integration of training set design, statistical modeling, and Bayesian optimization, thereby expanding access to stereodefined trisubstituted alkenes.



KEYWORDS: *alkenes, asymmetric catalysis, Bayesian optimization, statistical modeling, stereoconvergent synthesis*

Trisubstituted alkenes find diverse applications as valuable starting materials and intermediates in the synthesis of drug substances and polymeric materials.[1] Consequently, significant effort has been devoted to developing efficient and selective methods for their synthesis. While a myriad of existing methods to synthesize trisubstituted alkenes have been reported, many suffer from narrow substrate scope and significant challenges associated with stereocontrol and regioselectivity. The most popular strategies include alkyne functionalization,[2] cross-metathesis,[3] and C−H functionalization.[4] Among the more robust methods to overcome stereocontrol issues are stereoretentive metal-catalyzed cross-couplings using stereodefined starting materials; however, this approach requires access to diastereomerically pure alkene substrates (Figure 1A).[5] Conversely, many metal-catalyzed examples are stereoinvertive whereby the stereoselectivity is governed by thermodynamics. In stereoinvertive examples, isomerization of the *Z*-alkene is facilitated by the catalyst complex to generate the *E*-alkene.[6] An attractive, and potentially superior alternative would be the stereoconvergent cross-coupling of diastereomeric mixtures of alkene starting materials that can react and interconvert to primarily access one isomeric product. We recently applied such a stereoconvergent approach to the synthesis of tetrasubstituted alkenes via nickel-catalyzed cross-coupling of enol tosylates with various organometallic coupling partners such as Suzuki-Miyaura boronic esters (Figure 1B).[8] Our work stands apart from thermodynamic control, as both *E*- and *Z*-alkene products can be accessed selectively depending on the choice of monophosphine ligand. Consequently, the reaction can employ diastereomeric mixtures of enol tosylate substrates, which further simplifies their preparation from readily available ketone precursors. Data science tools were applied throughout our optimization efforts, and this analysis uncovered a significant influence of ligand structure on both the yield and the diastereoselectivity of the reaction (Figure 1B). We rationalized these ligand effects based on experimental evidence and computational models supporting discrete nickel-mediated isomerization events.[7]

Received: January 29, 2024
Revised: February 26, 2024
Accepted: February 27, 2024

https://doi.org/10.1021/acscatal.4c00650
ACS Catal. 2024, 14, 4699−4708

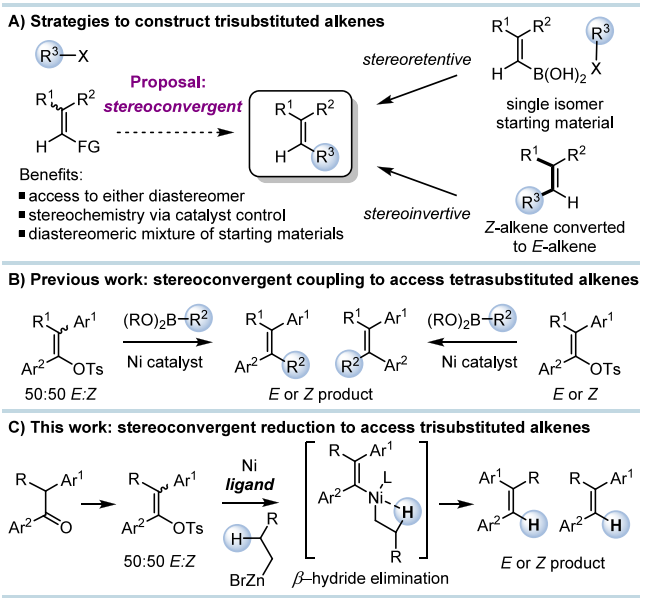


Figure 1. (A) Synthesis of trisubstituted alkenes. (B) Previous synthesis of tetrasubstituted alkenes via stereoconvergent cross-coupling. (C) Ligand-directed reduction of diastereomeric mixtures of enol tosylates to form stereodefined trisubstituted alkenes.

We subsequently endeavored to leverage this underutilized alkene isomerization manifold toward the synthesis of trisubstituted alkenes from diastereomeric mixtures of enol tosylates by using alkylzinc coupling partners to afford a net reduction of the carbon-tosylate bond (Figure 1C). Indeed, when attempting to perform sp^2-sp^3 Negishi or alkylmagnesium couplings between the aforementioned enol tosylates and alkylzinc partners containing β-hydrogen atoms, we observed reduction to trisubstituted alkenes as a significant side reaction. This presumably results from β-hydride elimination in competition with reductive elimination of the transmetalated alkyl-nickel intermediate (*vide infra*). Unfortunately, the translation of our previous reaction conditions did not provide acceptable yields or selectivity for the desired trisubstituted alkene products. We viewed this as an opportunity to develop a new transformation and implement a multifaceted machine learning workflow for reaction optimization (Figure 2A). Specifically, this reaction type requires the optimization of multiple objectives, a rapidly evolving aim in the application of machine learning in organic chemistry.[9] Ultimately, the data would be used to predict the diastereoselectivity *in silico*.

First, we planned to apply our recently reported computational database of calculated monophosphine descriptors to define an unbiased training set of ligands.[10] The training set was selected through dimensionality reduction of the descriptors and clustering to identify diverse monophosphine ligands.[11] This was followed by analysis of the experimental outcomes using statistical models to determine the most important ligand features. These features were then used to define search spaces for the recently disclosed multiobjective Bayesian optimization (EDBO+) algorithm to optimize yield and stereoselectivity simultaneously (Figure 2A).[12]

This approach showcases the synergy between training set design to identify high performing ligands, statistical modeling to determine important descriptors, and Bayesian optimization to accelerate the exploration of catalyst and reaction condition

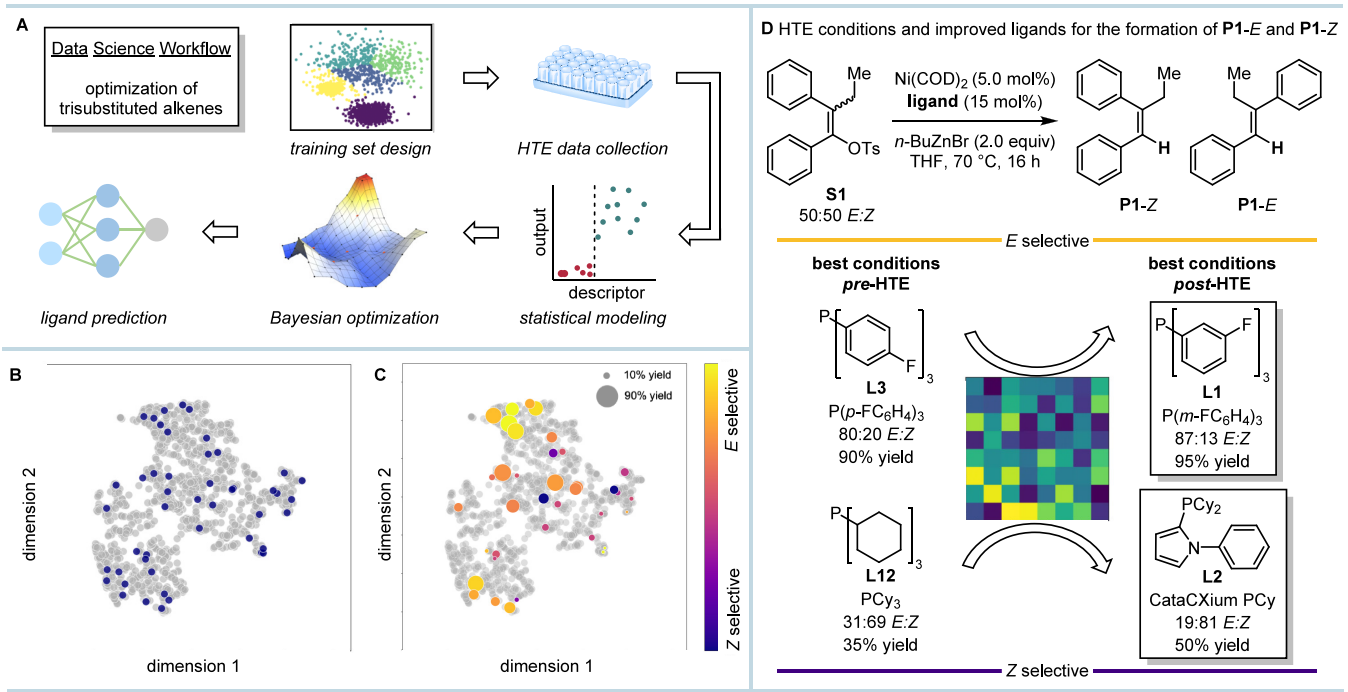


Figure 2. (A) Data science optimization workflow. (B) Monophosphine training set (blue) plotted against commercially available ligands (gray) visualized with uniform manifold approximation and projection (UMAP). Original clusters were generated using principal component analysis and *k*-means clustering, visualized here with UMAP for enhanced interpretability. (C) HTE results of monophosphine training set with color indicating selectivity and data point size representing yield, visualized with UMAP. (D) Reaction conditions used in the HTE screening campaign with enol tosylates **S1** and alkene products **P1**-*Z* and **P1**-*E*. Improved ligands are highlighted for **P1**-*E* and **P1**-*Z* product formation relative to the initial "hits" (assay yields determined by HPLC analysis).

https://doi.org/10.1021/acscatal.4c00650
ACS Catal. 2024, 14, 4699−4708

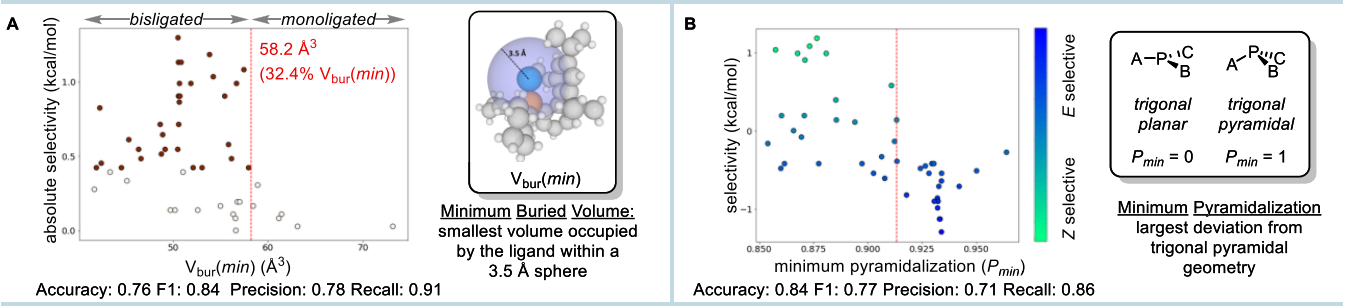


Figure 3. (A) Selectivity classification using the absolute value of $\Delta\Delta G^{\ddagger}$ to normalize *E*- and *Z*-selectivity, plotted against minimum buried volume (V_{bur}(min)). Classification threshold (red dotted line) at 58.2 Å^3 V_{bur}(min) (or 32.4% V_{bur}(min)), and 0.4 kcal/mol $\Delta\Delta G^{\ddagger}$. (B) Classification with minimum pyramidalization (P_{min}). Classification threshold at 0.91 units, P_{min}.

pairings, especially when multiple optimization objectives are present. Notably, the training set of phosphine ligands facilitated the exploration of broad regions of chemical space, allowing us to pre-train the EDBO+ algorithm and expedite the optimization process. Ultimately, we were able to identify robust reaction conditions with optimized ligands to selectively access either **P1**-*E* or **P1**-*Z* trisubstituted alkene product from a 50:50 diastereomeric mixture of enol tosylates **S1** (Figure 2D). As a final step, the experiments from the optimization campaign were used as a training set in a low-data machine learning model to predict *E*- vs *Z*-selectivity for previously unexplored ligands.

RESULTS AND DISCUSSION

Training Set Design and Ligand Classification. Training set design is increasingly recognized as an essential tool for successful optimization campaigns.[13] Given the previous structure−function relationship observed between phosphine ligand identity and diastereoselectivity in the synthesis of tetrasubstituted alkenes, we initiated our optimization by selecting a structurally diverse training set of monophosphine ligands.[14] This strategy facilitates the selection of a relatively broad sampling of ligand structures with the hypothesis that diverse ligand structures will produce a range of reaction outcomes.[15] A data distribution with both low and high selectivity is critical in the construction of robust statistical models. Furthermore, training set screening facilitates the discovery of structurally distinct, active ligand scaffolds. To construct the chemical space of the monophosphine ligands, Principal Component Analysis (PCA) was deployed as an unsupervised dimensionality reduction technique using the DFT-computed molecular descriptors from the *kraken* monophosphine descriptor database (Figure 2B).[10]

This step was followed by training set selection using *k*-means clustering to identify similar groups of phosphines based on their molecular descriptors. From clustering analysis, 47 ligand clusters were identified and one ligand per cluster was selected for screening via high throughput experimentation (HTE, Figure 2B). The reactions were performed in duplicate with one ligand-free control to assess reproducibility and control for experimental error. This workflow produced a standardized data set primed for analysis via statistical modeling.

In general, most ligands under these conditions promoted the formation of the *E*-alkene isomer (Figure 2C, yellow). In particular, the *E*-alkene product (**P1**-*E*) was preferentially produced, with triaryl and predominantly electron-deficient phosphines being the most effective ligands (for example, **L1**, Figure 2D). In contrast, the *Z*-alkene product (**P1**-*Z*) is accessed via several ligands featuring a dicyclohexyl substitution pattern with a third, sterically encumbered phosphorus substituent (as illustrated by **L2**, Figure 2D). Using a statistically diverse training set of monophosphine ligands allowed identification of ligands with divergent reactivity that enhance the formation of either *E*- or *Z*-alkene product. In particular, both the yield and selectivity of **P1**-*Z* were greatly improved with **L2** (CataCXium PCy), a nonintuitive leap from our initial hit of tricyclohexylphosphine (**L12**, Figure 2D).

Given these anecdotal structural differences, we sought to further rationalize and quantify the physicochemical ligand properties responsible for selectivity in this reaction. We initially evaluated a binary classification algorithm as it offers a simple means of categorizing the selectivity based on a single molecular descriptor value.[16] We first focused on identifying which ligand features lead to high selectivity, irrespective of the isomer formed. A single node decision tree was applied to bin the absolute value of selectivity ($\Delta\Delta G^{\ddagger}$) to a mechanistically relevant *kraken* descriptor (Figure 3A, code available on GitHub).[10] The decision tree selection procedure iteratively computes accuracy and F1 score for the data set with each molecular descriptor in the *kraken* library, penalizing false-negatives with a 10:1 weighting. This weighting ensures all true positive data points are correctly classified so no active ligands are overlooked. This analysis resulted in a model using the minimum buried volume (V_{bur}(min)) molecular descriptor, which describes the smallest volume a ligand can adopt within a 3.5 Å sphere centered on the metal. We have previously found V_{bur}(min) to be a surrogate for the ligation state in Ni- and Pd-catalyzed reactions.[17] The results indicated that monoligated (L_1) Ni-species associated with large V_{bur}(min) values (>58.2 Å^3, or >32% in the original publication) lead to poor selectivity for either isomer (Figure 3A). This process allows the curation of the data set to focus on ligands presumed to be bisligated (L_2) and remove data associated with nonselective ligands. To further explore the differentiating factors between **P1**-*E* and **P1**-*Z* product formation, we next performed a single-node decision tree on the active L_2 ligands, now differentiating between *E*- and *Z*-measured selectivity.

A correlation to the minimum pyramidalization (P_{min}) divided two groups of ligands that form stereochemically distinguished products (Figure 3B). Pyramidalization measures

https://doi.org/10.1021/acscatal.4c00650
ACS Catal. 2024, 14, 4699−4708

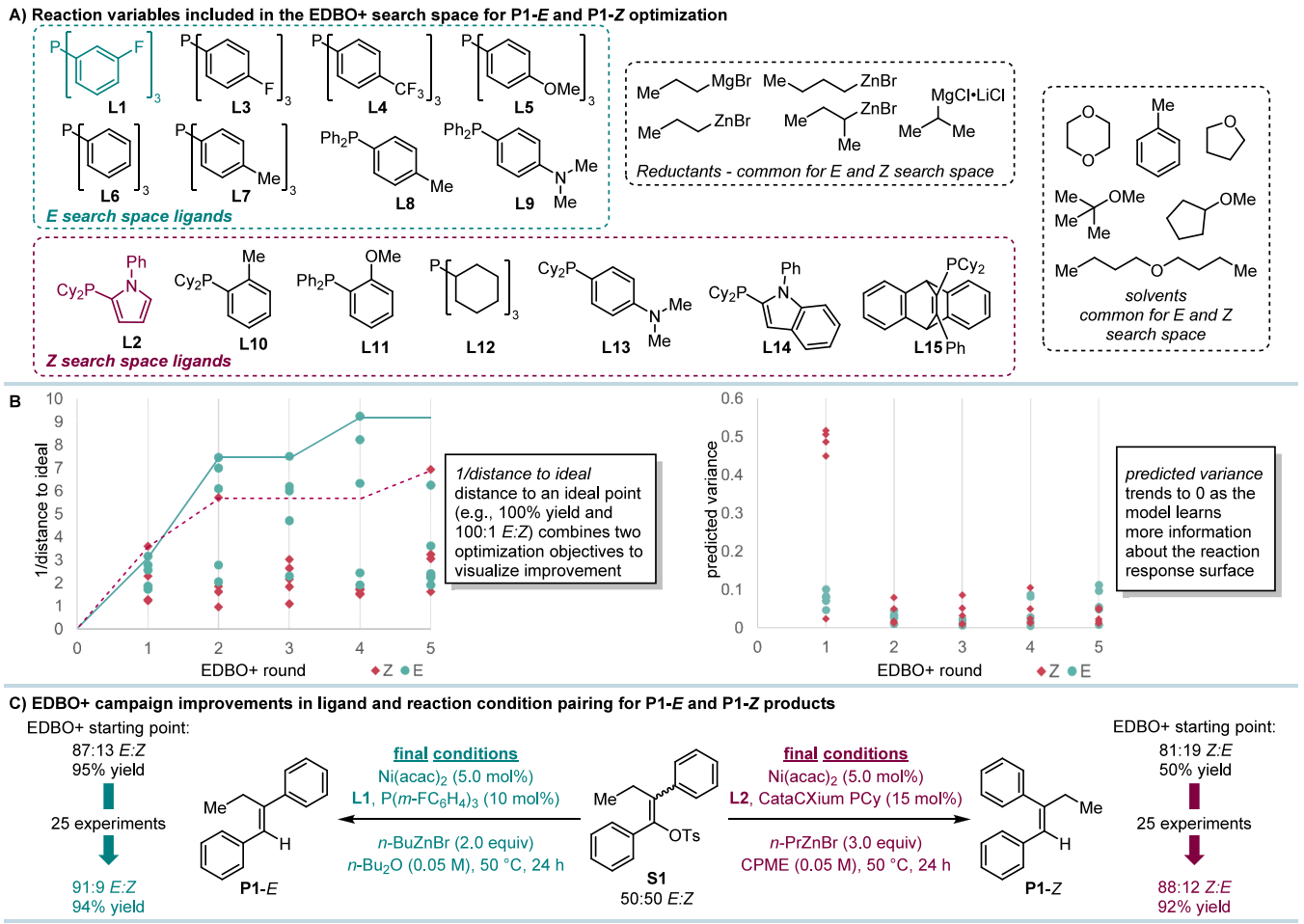


Figure 4. (A) Search space components for both *E*- and *Z*-selective EDBO+ campaigns. Reductants, solvents, concentration, and temperature were consistent between both search spaces, and different ligands were selected to promote the formation of *E*- vs. *Z*-products. (B) EDBO+ round-over-round improvement visualized by inverse distance to ideal (distance of each data point to 100% yield and 100:1 selectivity) and decrease in predicted variance. Yield and selectivity data were obtained from quantitative ^{1}H NMR with 1,3,5-trimethoxybenzene as the spectroscopic standard (see SI for details). (C) Final conditions and improvement in *E*- and *Z*-alkene formation from EDBO+ campaign.

the deviation from trigonal planar geometry influenced by the repulsive and attractive forces of the substituents on phosphorus.[18] The "minimum" of pyramidalization refers to the lowest value across a conformational ensemble for each ligand.[10] Trigonal pyramidal ligand geometries are associated with the formation of the *E*-alkene isomer (P_{min}>0.91). Conversely, more planar ligands perturbed from the ideal trigonal pyramidal geometry promote the formation of *Z*-alkenes (P_{min}<0.91). Both classification models support the conclusion that steric interactions of the ligand with the substrate substantially influence the *E*/*Z* isomerization during the catalytic cycle (*vide infra*).[19] While a single molecular descriptor does not explain the selectivity observed for every ligand, both the P_{min} and V_{bur}(min) thresholds point to an ideal range of steric bulk that favors the formation of the *E*- or *Z*-trisubstituted alkene product.

This initial optimization campaign identified several ligands capable of achieving promising yields and selectivity for both alkene isomers (Figure 2D). However, it was unclear if we had determined global maxima for each diastereoselective regime, especially considering this evaluation focused primarily on ligand structure. When considering the dimensionality of the other reaction variables, Bayesian optimization was chosen as an efficient statistical approach to reduce the required number of experiments, while still considering all optimization variables.[20]

Multiobjective Bayesian Optimization (EDBO+). Integrating training set design, statistical modeling, and Bayesian optimization (BO) offers significant advantages over design of experiments (DOE) or intuition-guided optimization methods. The key strength of BO lies in its ability to relate prediction and uncertainty.[21] Unlike DOE, Bayesian optimization incorporates information about the uncertainty of predictions, enabling the prioritization and weighting of proposed experiments based on the likelihood of success.[22] Furthermore, BO can incorporate many variables, including discrete variables such as reagents and ligands, that can dramatically increase the dimensionality of the search space; therefore, BO reduces the experimental workload. In BO, input data is modeled using a surrogate function, which approximates the overall reaction surface and predicts uncertainty across different regions of that surface. Subsequently, an acquisition function combines the modeled reaction surface and associated uncertainty to select experiments that balance exploration and exploitation.[23] Exploratory experiments aim to uncover regions of the reaction surface with large predicted uncertainty, while exploitative experiments focus on sampling areas near the predicted local maximum. This interplay between exploration and exploitation

https://doi.org/10.1021/acscatal.4c00650
ACS Catal. 2024, 14, 4699−4708

allows for rapid identification of improved reaction conditions. Statistical sampling of the reaction surface, instead of sampling random or extreme points, is highly efficient in capturing the nonlinear relationships between reaction components. The link between predicted uncertainty and global maximum provides valuable information to streamline the optimization process.

The curation of the search space and careful selection of variables play a crucial role in successful optimization campaigns. Prior knowledge of reaction sensitivities significantly aids in determining which variables to include or exclude. In this study, we selected *focused* search spaces to optimize for the formation of **P1**-*E* and **P1**-*Z* products independently. We constructed search spaces to uncover the nuanced interdependence of different reaction components, exploiting the knowledge gleaned from our diverse training set of phosphines and additional control reactions. Several of the top-performing phosphine ligands from the training set were selected to include in the initial search space due to the ligands' strong influence on selectivity (Figure 4A). The included phosphine molecular descriptors for the search space were identified using multivariate linear regression modeling to find the most highly correlated descriptors. These descriptors were further curated based on chemical expertise (see SI for details). Other critical variables such as solvent, reductant, temperature, reaction concentration, and catalyst loading were incorporated in both search spaces for a total of 33,600 possible conditions in the *E*-search space and 29,400 combinations in the *Z*-search space (see SI for details).[24] The EDBO+ model was then initialized with Gaussian Process Regression (GPR) and expected hypervolume improvement (qEHVI) as surrogate and acquisition functions, respectively.[12] The training set data was used to pre-train these models, and suggested five experiments for each isomer campaign round (*E* and *Z*). The number of experiments was set as a user defined hyperparameter to balance experimental throughput and the model training rate. Yield and selectivity data were acquired via quantitative ^{1}H NMR with 1,3,5-trimethoxybenzene as the spectroscopic standard. The EDBO+ model was then updated with the new round of data and the next set of experiments were suggested. This cycle continued until the predicted uncertainty and predicted improvement converged over five rounds (Figure 4B).

After concluding the optimization, selectivity for the formation of **P1**-*E* product had improved from 87:13 to 91:9 *E*:*Z* using **L1** (+0.5 kcal/mol) with a high yield of 94% (Figure 4C). Though the formation of **P1**-*E* maintained a high yield throughout the EDBO+ rounds, the best reaction conditions were obtained from BO round four (Figure 4B). The optimization campaign for the **P1**-*Z* alkene was initiated at a more challenging starting point, initially using PCy_3 (**L12**, 69:31 *Z*:*E* and 35% yield) as the ligand prior to the HTE campaign (Figure 2D). Two ligands identified in the HTE campaign performed particularly well, **L2** and **L10** (Figure 5). Both contain the conserved PCy_2(Ar) structural scaffold common in *Z*-selective ligands. The CataCXium PCy ligand (**L2**)[25] offers robust conversion, albeit with slightly reduced diastereoselectivity relative to the PCy_2(*o*-tolyl) ligand (**L10**, up to 91:9 *Z*:*E*). Ultimately, we chose to prioritize **L2** due to its suitable selectivity and robust yield of the *Z*-alkene. Through all our optimization efforts, the *Z*-alkene formation metrics improved by 2.5-fold in yield (92%) and from 69:31 to 88:12 *Z*:*E* selectivity (+0.7 kcal/mol) after the HTE campaign and 25 EDBO+ guided experiments.

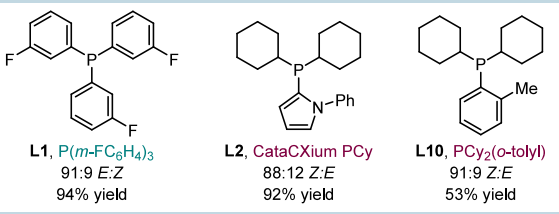
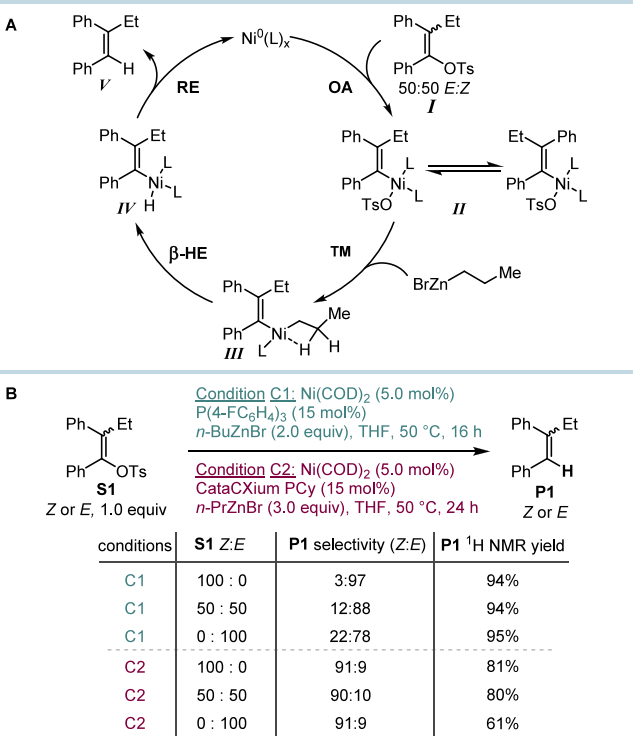


Figure 5. Top-performing *E*- and *Z*-selective ligands.

Overall, we determined that the ligand was the most crucial factor in directing selectivity; however, solvent, reductant, and temperature all play a nuanced role. Ultimately, we performed only 50 reactions after the HTE campaign to optimize these two divergent stereochemical goals. In our experience, this is a relatively modest number of experiments in the context of multiple optimization objectives with >30,000 possible combinations.

Mechanistic Considerations. Based on our previous efforts and experimental results, a catalytic cycle is proposed for this transformation (Figure 6A). First, oxidative addition



conditions	S1 *Z*:*E*	P1 selectivity (*Z*:*E*)	P1 ^{1}H NMR yield
C1	100 : 0	3:97	94%
C1	50 : 50	12:88	94%
C1	0 : 100	22:78	95%
C2	100 : 0	91:9	81%
C2	50 : 50	90:10	80%
C2	0 : 100	91:9	61%

Figure 6. (A) Proposed catalytic cycle. (B) Curtin-Hammett studies to assess the influence of starting material geometry on selectivity. Selectivity and yield determined by ^{1}H NMR spectroscopic analysis.

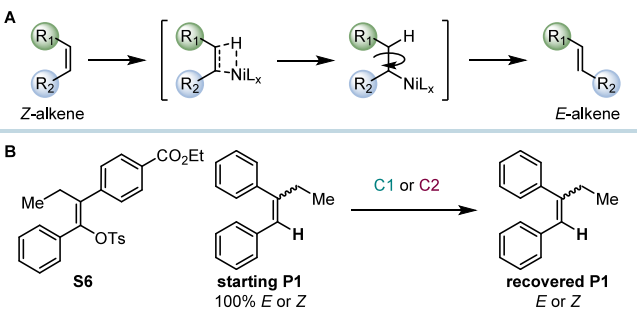
into the enol tosylate **I** (*E* or *Z*) produces the vinyl-nickel species **II**, followed by facile isomerization to interconvert *E*- and *Z*-isomers of complex **II**. During the investigation of the stereoconvergent nickel-catalyzed Suzuki-Miyaura reaction, we determined that extensive isomerization of the oxidative addition complex occurs, with selectivity linked to the ligand identity.[8]

To identify if the isomerization process obeys the Curtin-Hammett principle, we initiated the reaction with different diastereomeric ratios of enol tosylate substrates **S1** (Figure

https://doi.org/10.1021/acscatal.4c00650
ACS Catal. 2024, 14, 4699−4708

6B). If the reaction is under Curtin-Hammett control, then by definition, isomerization is faster than the subsequent stereodetermining step, and starting material geometry should have no effect on the resulting selectivity.[26] We observe that only the *Z*-selective reaction (conditions C2) obeys the Curtin-Hammett principle, whereby the diastereoselectivity for product **P1** does not depend on the geometry of the starting material **S1** (Figure 6B, C2 conditions). This is consistent with our previous results, indicating the C1 conditions for selective *E*-alkene formation may promote a "kinetic quench scenario" in which the barrier to isomerization is competitive with that of the stereodetermining step.[27] This allows for enhanced selectivity when using the corresponding **S1**-*Z* starting material, favoring stereoretentive formation of the **P1**-*E* alkene product. Following oxidative addition, we propose that transmetalation of **II** with the alkyl zinc reagent gives rise to Ni-alkyl species **III**. This pendant Ni−alkyl can then undergo *β*-hydride elimination through a four-coordinate transition state to yield complex **IV**.

Several examples[6] of Ni-mediated alkene isomerizations take advantage of the steric driving force to convert *Z*-alkenes to the less strained *E*-alkenes (Figure 7A). These stereoinvertive



entry	S6	starting P1 E:Z	ligand	recovered P1 E:Z	yield
1	1.0 equiv	100:0	L2	100:0	>99%
2	1.0 equiv	0:100	L1	0:100	>99%

Figure 7. (A) Ni−H mediated isomerization. (B) Crossover experiment: yield and selectivity determined by HPLC and ^{1}H NMR spectroscopic analyses. **S6** produces corresponding reduced products **P6**, see SI for details. C1 conditions: Ni(acac)$_2$ (5.0 mol%), **L1** (10 mol%), *n*-BuZnBr (2.0 equiv), *n*-Bu$_2$O (0.05 M), 50 °C, 24 h. C2 conditions: Ni(acac)$_2$ (5.0 mol%), **L2** (15 mol%), *n*-PrZnBr (3.0 equiv), CPME (0.05 M), 50 °C, 24 h.

mechanisms involve exogenous Ni-hydride (Ni−H) addition across the alkene. To test for this pathway and assess the potential impact of off-cycle Ni−H species in our reaction system, we designed a crossover experiment that was performed by subjecting a single isomer of each product (**P1**-*E* or **P1**-*Z*) to the reaction conditions and monitoring any subsequent isomerization (Figure 7B). A substituted enol tosylate substrate (Figure 7B, **S6**) was also included to generate the relevant oxidative addition and transmetalation catalytic intermediates and replicate any potential off-cycle species.[28] Complete recovery of both products **P1** was observed with no erosion of diastereomeric purity, indicating that an off-cycle Ni−H-mediated isomerization of **P1** is unlikely under these reaction conditions. The results of the crossover study further suggest that the trisubstituted alkene products are stable to the reaction conditions, and thus reductive elimination is likely irreversible. Our prior investigation of the Suzuki-Miyaura reaction mechanism indicated a complex isomerization landscape, with multiple factors influencing selectivity.[8]

Reaction Scope. The reaction scope was assessed with several structural perturbations to the initial model system **S1**/**P1** (Figure 8A). Overall, the reactions to form *E*-alkenes with

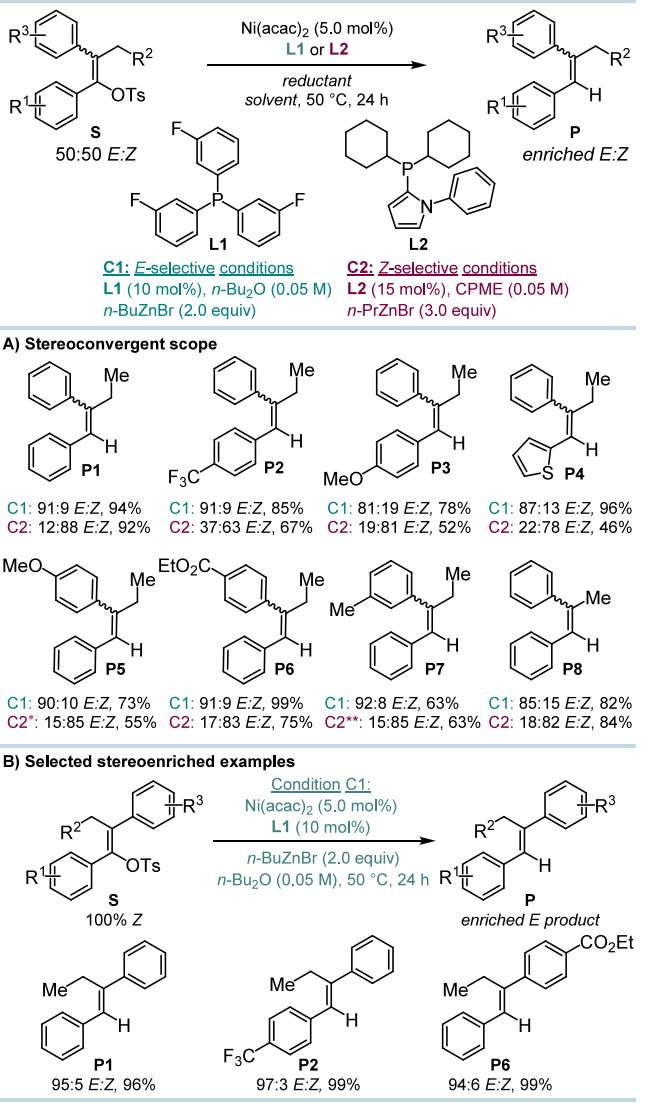


Figure 8. (A) Stereoconvergent scope examples from 50:50 *E*:*Z* enol tosylate. Selectivity was determined from ^{1}H NMR analysis of the crude reaction mixture. Isolated yields were reported for mixtures of isomers, isolated selectivity is reported in the SI. ***P5**-*Z*: isolated selectivity reported. ****P7**-*Z*: 10 mol% Ni(acac)$_2$ and 30 mol% **L2**. (B) Stereoenriched scope starting from *Z*-enol tosylates under C1 conditions. Selectivity was determined by crude ^{1}H NMR analysis.

P(*m*-FC$_6$H$_4$)$_3$ (**L1**) as the ligand are robust to substrate modifications, including electron donating and withdrawing groups and modification of the aromatic rings, with yields of typically >80% and diastereoselectivities up to 92:8 *E*:*Z* (Figure 8A, conditions C1). Not surprisingly, the reactions to form sterically congested *Z*-alkenes were more challenging (Figure 8A, conditions C2). We observed enriched mixtures of *Z*-products for both electron-rich and -poor aromatic rings, albeit with reduced selectivity in some cases. The product *E*-isomer can be enriched by leveraging the kinetic quench

https://doi.org/10.1021/acscatal.4c00650
ACS Catal. 2024, 14, 4699−4708

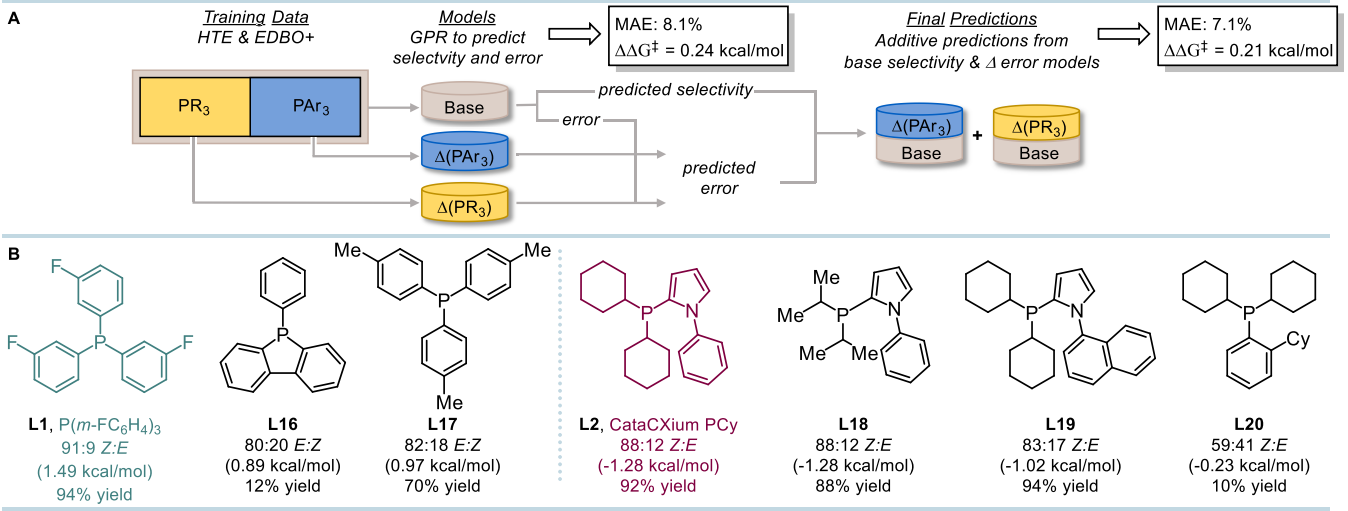


Figure 9. (A) Model development workflow for extrapolation ligand predictions. (B) Predicted *E*-selective ligands (**L16**–**L17**) and two of the three predicted *Z*-selective ligands (**L18**–**L20**) perform similarly to top-performing ligands (**L1** and **L2**) from the optimization campaign.

scenario of the Curtin-Hammett principle and using stereodefined starting materials (Figure 8B). When diastereomerically pure *Z*-enol tosylates are employed as substrates under conditions C1, the diastereoselectivities for the formation of the *E*-products are enhanced, as depicted in Figure 6B. Overall, the reaction scope demonstrates that a variety of conjugated trisubstituted alkenes can be synthesized diastereoselectively with this method. The diastereomeric mixtures of starting materials can be prepared in two steps from readily available ketones, obviating the need for the challenging preparation of stereodefined starting materials[5c,29] and the proper choice of stereoretentive or stereoinvertive cross-coupling methods (Figure 1).

Selectivity Predictions Using Machine Learning. By taking advantage of a rich data pool of 242 data points from the HTE and EDBO+ campaigns, we developed a model to predict the selectivity for untested ligands in the *kraken* library. Initially, the training set of monophosphine ligands included only commercially available ligands, as our goals involved profiling ligands in the lab. This initial assumption narrowed our search for the best monophosphine ligands to 300 commercially available structures.[30] While we were able to identify two highly active and selective ligands for the formation of both the *E*- and *Z*-alkenes from this pool, we became curious if we had overlooked fruitful regions of chemical space, particularly in the more challenging *Z*-selective regime. To leverage the existing data collected from the HTE campaign, we first curated it by removing results with discrepancies of greater than 15% yield or 5% selectivity between duplicate runs (see SI for details). Several machine learning algorithms were assessed, and Gaussian Process Regression (GPR) resulted in the lowest mean absolute error (MAE) with respect to product selectivity (base model, Figure 9A). Due to the small data set size (for machine learning algorithms) and the extensive influence of ligand choice on selectivity, the base model's predictive power was assessed using leave-one-ligand-out cross-validation, where all reactions with a specific ligand were removed from the training set and used in the validation set. This process was repeated for each ligand in the training set, producing an MAE of 8.1% ($\Delta\Delta G^{\ddagger}$ = 0.24 kcal/mol). Seeking to improve the model further, we implemented a hierarchical learning algorithm inspired by the work of Hong and co-workers.[31] In this regime, the data set is divided into multiple subsets and a set of additional models are used to predict the error of each subset produced by the base model. The predicted errors are then summed with the selectivity predictions from the base model to afford a composite prediction with higher accuracy.

Specifically, two delta models were trained on the partitioned data set, triaryl monophosphines (ΔPAr_3) and non-triaryl monophosphines (ΔPR_3). This allowed a more fine-tuned prediction that captures the structural deviations in each partition. Additionally, dividing the data set into PAr_3 and PR_3 was chemically intuitive as it roughly partitioned ligands that induce *E*- vs. *Z*-alkene formation, respectively.

GPR was used for both the base model and the delta models with identical hyperparameters. For each of the models, a greedy sequential feature elimination with the identical GPR model was used to select the features for the corresponding data sets, yielding three distinct sets of features. This workflow improved the accuracy to a MAE of 7.1% ($\Delta\Delta G^{\ddagger}$ = 0.21 kcal/mol Figure 9A). Next, we evaluated the final predictions to decide which ligands to evaluate experimentally. We included our top-performing ligands in this prediction to validate model accuracy further. For each ligand, the composite model also predicted the best reaction conditions from the search space for a total of 5.9 million combinations. Since the *E*-alkene forming reaction had achieved both a robust yield and >90:10 *E*:*Z* selectivity, we selected two commercially available ligands to validate the ΔPAr_3 model (Figure 9B, **L16**, **L17**). These ligands were predicted to be interpolations, with selectivity below our top-performing ligand **L1.** Despite the structural and electronic differentiation of **L16** and **L17**, we were able to accurately predict the selectivity of these structures, thus validating our ΔPAr_3 model. Next, we chose three highly ranked ligands for assessing the formation of the *Z*-alkene product, as this reaction had proven more challenging. Besides our optimized ligands **L2** and **L10**, the composite model ranked several other intriguing ligand motifs with a high probability of success. The predicted ligands were selected by triaging the 20 highest selectivity predictions, including commercially unavailable structures. First, ligands tested in the training set were removed, followed by synthetically inaccessible structures and incompatible functional groups (see

https://doi.org/10.1021/acscatal.4c00650
ACS Catal. 2024, 14, 4699−4708

SI for details). Intuitively, these predicted ligands are structurally similar to our top-performing scaffolds, adding confidence that our predictions would perform well in the reaction.

We were delighted to see ligands bearing a pyrrole scaffold maintained robust conversion and delivered high selectivity. P(*i*-Pr)$_2$(*N*-phenyl pyrrole) (**L18**, Figure 9B) produced a competitive yield and selectivity to our best ligand, CataCXium PCy (**L2**). Additionally, **L19** maintained good conversion of starting material, albeit at reduced selectivity. The final predicted ligand had low conversion, indicating it is incapable of promoting catalysis (**L20**, Figure 9B). This outcome may be the result of our specific focus on predicting selectivity, the more challenging reaction metric to achieve in this particular reaction. In summary, the pyrrole scaffold is uniquely suited in achieving high conversion to the *Z*-alkene.

This workflow enabled selectivity prediction for ∼1200 monophosphine ligands, resulting in a robust model with multiple successful projections. The development of the composite model highlights several steps the modern data chemist can take to refine model predictions in complex reaction optimization campaigns. Furthermore, this strategy validates the concept that diverse training sets are critical to a broad exploration of chemical space. The composite model was trained on the HTE and EDBO+ data, allowing for a diverse sampling across the *kraken* library and resulting in accurate selectivity predictions for a wide range of ligand scaffolds.

CONCLUSIONS

The catalytic, stereoconvergent synthesis of trisubstituted alkenes presented herein demonstrates the synergy of training set design, statistical modeling, and Bayesian optimization in addressing the challenges associated with reaction development. Data collected from the training set was critical to selecting ideal catalyst complexes to incorporate into the EDBO+ search space. We were further able to inform and streamline our Bayesian optimization campaign to focus the search space and direct optimization to the most fruitful reaction conditions based on an initial set of 96 high-throughput experiments.[32] This strategy additionally minimized the required number of experimental sampling rounds, as we could pretrain the EDBO+ surrogate function with existing HTE data and thus allow for rapid identification of the global maxima across a well-established reaction surface. Where traditional one-factor-at-a-time or design of experiments (DOE) optimization consider select ranges of only a few variables (e.g., solvent, temperature, ligand), EDBO+ allowed us to profile a combinatorial reaction space of >30,000 reaction conditions and arrive at improved conditions for *E*- and *Z*-alkene formation within just 25 experiments each. Using the data in hand, we were able to construct a composite model to predict the selectivity of all ligands in the *kraken* library. These predictions were validated on selected ligands and included identification of a highly active ligand **L18**, which promoted the formation of the challenging *Z* isomer with high diastereoselectivity.

Active regions of chemical space can be quickly identified by leveraging a statistically sound training set and integrating statistical modeling and optimization algorithms. The EDBO+ platform specifically allowed for the simultaneous optimization of these two high-dimensionality reaction campaigns to discover efficient reaction conditions and improved ligands for the stereoselective synthesis of trisubstituted alkenes.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscatal.4c00650.

The Supporting Information contains all experimental procedures and resulting data for the HTE and Bayesian optimization campaigns, mechanism experiments, and scope. NMR and HRMS characterization data is provided for all compounds. Computational details from the EDBO+ campaign, statistical modeling, and the ligand prediction model are provided, and the relevant code is linked to our GitHub repository (PDF)

AUTHOR INFORMATION

Corresponding Authors

Daniel Zell − *Department of Synthetic Molecule Process Chemistry, Genentech Inc., South San Francisco, California 94080, United States;* orcid.org/0000-0002-2241-6301; Email: zell.daniel@gene.com

Abigail G. Doyle − *Department of Chemistry, University of California, Los Angeles, California 90095, United States;* orcid.org/0000-0002-6641-0833; Email: agdoyle@chem.ucla.edu

Matthew S. Sigman − *Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, United States;* orcid.org/0000-0002-5746-8830; Email: matt.sigman@utah.edu

Authors

Natalie P. Romer − *Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, United States*

Daniel S. Min − *Department of Chemistry, University of California, Los Angeles, California 90095, United States*

Jason Y. Wang − *Department of Chemistry, University of California, Los Angeles, California 90095, United States;* orcid.org/0000-0001-5826-2554

Richard C. Walroth − *Department of Synthetic Molecule Process Chemistry, Genentech Inc., South San Francisco, California 94080, United States*

Kyle A. Mack − *Department of Synthetic Molecule Process Chemistry, Genentech Inc., South San Francisco, California 94080, United States*

Lauren E. Sirois − *Department of Synthetic Molecule Process Chemistry, Genentech Inc., South San Francisco, California 94080, United States;* orcid.org/0000-0002-1948-3749

Francis Gosselin − *Department of Synthetic Molecule Process Chemistry, Genentech Inc., South San Francisco, California 94080, United States;* orcid.org/0000-0001-9812-4180

Complete contact information is available at: https://pubs.acs.org/10.1021/acscatal.4c00650

Author Contributions

All authors have given approval to the final version of the manuscript.

Funding

Researchers in the Sigman and Doyle laboratories acknowledge financial support from the NSF under the CCI Center for Computer Assisted Synthesis (CHE−2202693).

Notes

The authors declare no competing financial interest.

The authors have cited additional references within the Supporting Information. Code and supporting files can also be found at https://github.com/doyle-lab-ucla/tosylate-redution-ML This material is available free of charge via the Internet at http://pubs.acs.org.

■ ACKNOWLEDGMENTS

The authors acknowledge Colin Masui (Genentech Inc.) for HTE experiments, Dr. Christopher M. Crittenden (Genentech Inc.) for HRMS analysis, and Quentinn Pearce and the Proteomics and Mass Spectrometry Cores at the University of Utah. We thank Maria Ruiz-Gonzalez (Genentech Inc.), Crystal Ye (Genentech Inc.), and Mengling Wong (Genentech Inc.) for purification, Karissa Cruz (Genentech Inc.) for analytical method development, and Paul Oblad (University of Utah) for NMR discussions.

■ ABBREVIATIONS

MAE:Mean Absolute Error; HTE:High-Throughput Experimentation; UMAP:Uniform Manifold Approximation and Projection; BO:Bayesian Optimization; EDBO+:Experimental Design via Bayesian Optimization, multi-objective; GPR:Gaussian Process Regression.

■ REFERENCES

(1) (a) He, Q.; Pu, M. P.; Jiang, Z.; Wang, H.; Feng, X.; Liu, X. Asymmetric Epoxidation of Alkenes Catalyzed by a Cobalt Complex. *J. Am. Chem. Soc.* **2023**, *145* (28), 15611−15618. (b) Kolb, H. C.; VanNieuwenhze, M. S.; Sharpless, K. B. Catalytic Asymmetric Dihydroxylation. *Chem. Rev.* **1994**, *94* (8), 2483−2547. (c) Kumar, G.; Bhattacharya, D.; Mistry, P.; Chatterjee, I. In-Catalyzed Transfer Hydrogenation and Regioselective Hydrogen-Deuterium Addition to the Olefins. *J. Org. Chem.* **2023**, *88* (11), 6987−6994. (d) Negishi, E.; Huang, Z.; Wang, G.; Mohan, S.; Wang, C.; Hattori, H. Recent advances in efficient and selective synthesis of di-, tri-, and tetrasubstituted alkenes via Pd-catalyzed alkenylation-carbonyl olefination synergy. *Acc. Chem. Res.* **2008**, *41* (11), 1474−1485. (e) Mei, J.; Hong, Y.; Lam, J. W.; Qin, A.; Tang, Y.; Tang, B. Z. Aggregation-induced emission: the whole is more brilliant than the parts. *Adv. Mater.* **2014**, *26* (31), 5429−5479. (f) Liu, H.; Xiong, L. H.; Kwok, R. T. K.; He, X.; Lam, J. W. Y.; Tang, B. Z. AIE Bioconjugates for Biomedical Applications. *Adv. Opt. Mater.* **2020**, *8* 14, 2000162. DOI: 10.1002/adom.202000162.

(2) (a) Kutateladze, D. A.; Mai, B. K.; Dong, Y.; Zhang, Y.; Liu, P.; Buchwald, S. L. Stereoselective Synthesis of Trisubstituted Alkenes via Copper Hydride-Catalyzed Alkyne Hydroalkylation. *J. Am. Chem. Soc.* **2023**, *145* (32), 17557−17563. (b) Till, N. A.; Smith, R. T.; MacMillan, D. W. C. Decarboxylative Hydroalkylation of Alkynes. *J. Am. Chem. Soc.* **2018**, *140* (17), 5701−5705. (c) Janson, P. G.; Ghoneim, I.; Ilchenko, N. O.; Szabo, K. J. Electrophilic trifluoromethylation by copper-catalyzed addition of CF3-transfer reagents to alkenes and alkynes. *Org. Lett.* **2012**, *14* (11), 2882−2885. (d) Xu, T.; Cheung, C. W.; Hu, X. Iron-catalyzed 1,2-addition of perfluoroalkyl iodides to alkynes and alkenes. *Angew. Chem., Int. Ed. Engl.* **2014**, *53* (19), 4910−4914. (e) Hazra, A.; Kephart, J. A.; Velian, A.; Lalic, G. Hydroalkylation of Alkynes: Functionalization of the Alkenyl Copper Intermediate through Single Electron Transfer Chemistry. *J. Am. Chem. Soc.* **2021**, *143* (21), 7903−7908.

(3) Nguyen, T. T.; Koh, M. J.; Mann, T. J.; Schrock, R. R.; Hoveyda, A. H. Synthesis of *E*- and *Z*-trisubstituted alkenes by catalytic cross-metathesis. *Nature* **2017**, *552* (7685), 347−354.

(4) (a) Ilies, L.; Asako, S.; Nakamura, E. Iron-catalyzed stereospecific activation of olefinic C-H bonds with Grignard reagent for synthesis of substituted olefins. *J. Am. Chem. Soc.* **2011**, *133* (20), 7672−7675. (b) Nakashima, Y.; Matsumoto, J.; Nishikata, T. Iron-Catalyzed Stereoconvergent Tertiary Alkylation of (*E*)- and (*Z*)-Mixed Internal Olefins with Functionalized Tertiary Alkyl Halides. *ACS Catal.* **2021**, *11* (18), 11526−11531.

(5) (a) Teh, W. P.; Michael, F. E. Palladium-Catalyzed Cross-Coupling of N-Sulfonylaziridines and Alkenylboronic Acids: Stereospecific Synthesis of Homoallylic Amines with Di- and Trisubstituted Alkenes. *Org. Lett.* **2017**, *19* (7), 1738−1740. (b) Cheung, C. W.; Hu, X. Stereoselective Synthesis of Trisubstituted Alkenes through Sequential Iron-Catalyzed Reductive anti-Carbozincation of Terminal Alkynes and Base-Metal-Catalyzed Negishi Cross-Coupling. *Chemistry* **2015**, *21* (50), 18439−18444. (c) Li, B. X.; Le, D. N.; Mack, K. A.; McClory, A.; Lim, N. K.; Cravillion, T.; Savage, S.; Han, C.; Collum, D. B.; Zhang, H.; et al. Highly Stereoselective Synthesis of Tetrasubstituted Acyclic All-Carbon Olefins via Enol Tosylation and Suzuki-Miyaura Coupling. *J. Am. Chem. Soc.* **2017**, *139* (31), 10777−10783.

(6) (a) Ho, G. M.; Sommer, H.; Marek, I. Highly E-Selective, Stereoconvergent Nickel-Catalyzed Suzuki-Miyaura Cross-Coupling of Alkenyl Ethers. *Org. Lett.* **2019**, *21* (8), 2913−2917. (b) Shimasaki, T.; Konno, Y.; Tobisu, M.; Chatani, N. Nickel-catalyzed cross-coupling reaction of alkenyl methyl ethers with aryl boronic esters. *Org. Lett.* **2009**, *11* (21), 4890−4892.

(7) (a) Allen, S. R.; Beevor, R. G.; Green, M.; Norman, N. C.; Orpen, A. G.; Williams, I. D. *Reactions of coordinated ligands* Part 33, Mononuclear η2-vinyl complexes: synthesis, structure, and reactivity. *J. Chem. Soc., Dalton Trans.* **1985**, 435−450. (b) Tanke, R. S.; Crabtree, R. H. Unusual activity and selectivity in alkyne hydrosilylation with an iridium catalyst stabilized by an oxygen-donor ligand. *J. Am. Chem. Soc.* **1990**, *112* (22), 7984−7989.

(8) Zell, D.; Kingston, C.; Jermaks, J.; Smith, S. R.; Seeger, N.; Wassmer, J.; Sirois, L. E.; Han, C.; Zhang, H.; Sigman, M. S.; et al. Stereoconvergent and -divergent Synthesis of Tetrasubstituted Alkenes by Nickel-Catalyzed Cross-Couplings. *J. Am. Chem. Soc.* **2021**, *143* (45), 19078−19090.

(9) (a) Kershaw, O. J.; Clayton, A. D.; Manson, J. A.; Barthelme, A.; Pavey, J.; Peach, P.; Mustakis, J.; Howard, R. M.; Chamberlain, T. W.; Warren, N. J.et al.Machine learning directed multi-objective optimization of mixed variable chemical systems. *J. Chem. Eng.* **2023**, *451*. DOI: 138443. (b) Ottoboni, S.; Wareham, B.; Vassileiou, A.; Robertson, M.; Brown, C. J.; Johnston, B.; Price, C. J. A Novel Integrated Workflow for Isolation Solvent Selection Using Prediction and Modeling. *Org. Process Res. Dev.* **2021**, *25* (5), 1143−1159.

(10) Gensch, T.; Dos Passos Gomes, G.; Friederich, P.; Peters, E.; Gaudin, T.; Pollice, R.; Jorner, K.; Nigam, A.; Lindner-D'Addario, M.; Sigman, M. S.; et al. A Comprehensive Discovery Platform for Organophosphorus Ligands for Catalysis. *J. Am. Chem. Soc.* **2022**, *144* (3), 1205−1217. See the Supporting Information for additional molecular descriptor information

(11) Bisphosphines were assessed and found to promote the formation of *E*-alkene, however, no bisphosphines promoted the formation of *Z*-alkene. Thus, monophosphines were prioritized for optimization due to the ability to access both trisubstituted product isomers.

(12) Torres, J. A. G.; Lau, S. H.; Anchuri, P.; Stevens, J. M.; Tabora, J. E.; Li, J.; Borovika, A.; Adams, R. P.; Doyle, A. G. A Multi-Objective Active Learning Platform and Web App for Reaction Optimization. *J. Am. Chem. Soc.* **2022**, *144* (43), 19999−20007.

(13) (a) Strieth-Kalthoff, F.; Sandfort, F.; Segler, M. H. S.; Glorius, F. Machine learning the ropes: principles, applications and directions in synthetic chemistry. *Chem. Soc. Rev.* **2020**, *49* (17), 6154−6168. (b) Pollice, R.; Dos Passos Gomes, G.; Aldeghi, M.; Hickman, R. J.; Krenn, M.; Lavigne, C.; Lindner-D'Addario, M.; Nigam, A.; Ser, C. T.; Yao, Z.; et al. Data-Driven Strategies for Accelerated Materials Design. *Acc. Chem. Res.* **2021**, *54* (4), 849−860.

(14) Gensch, T.; Smith, S. R.; Colacot, T. J.; Timsina, Y. N.; Xu, G.; Glasspoole, B. W.; Sigman, M. S. Design and Application of a Screening Set for Monophosphine Ligands in Cross-Coupling. *ACS Catal.* **2022**, *12* (13), 7773−7780.

https://doi.org/10.1021/acscatal.4c00650
ACS Catal. 2024, 14, 4699−4708

(15) (a) 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), No. eaau5631. (b) Hueffel, J. A.; Sperger, T.; Funes-Ardoiz, I.; Ward, J. S.; Rissanen, K.; Schoenebeck, F. Accelerated dinuclear palladium catalyst identification through unsupervised machine learning. *Science* **2021**, *374* (6571), 1134−1140. (c) Karl, T. M.; Bouayad-Gervais, S.; Hueffel, J. A.; Sperger, T.; Wellig, S.; Kaldas, S. J.; Dabranskaya, U.; Ward, J. S.; Rissanen, K.; Tizzard, G. J.; et al. Machine Learning-Guided Development of Trialkylphosphine Ni(I) Dimers and Applications in Site-Selective Catalysis. *J. Am. Chem. Soc.* **2023**, *145* (28), 15414−15424. (d) Liles, J. P.; Rouget-Virbel, C.; Wahlman, J. L. H.; Rahimoff, R.; Crawford, J. M.; Medlin, A.; O'Connor, V.; Li, J.; Roytman, V. A.; Toste, F. D.; et al. Data Science Enables the Development of a New Class of Chiral Phosphoric Acid Catalysts. *Chem.* **2023**, *9* (6), 1518−1537. (e) Bess, E. N.; Bischoff, A. J.; Sigman, M. S. Designer substrate library for quantitative, predictive modeling of reaction performance. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111* (41), 14698−14703. (f) Kariofillis, S. K.; Jiang, S.; Zuranski, 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. (g) Kutchukian, P. S.; Dropinski, J. F.; Dykstra, K. D.; Li, B.; DiRocco, D. A.; Streckfuss, E. C.; Campeau, L. C.; Cernak, T.; Vachal, P.; Davies, I. W.; et al. Chemistry informer libraries: a chemo-informatics enabled approach to evaluate and advance synthetic methods. *Chem. Sci.* **2016**, *7* (4), 2604−2613.

(16) (a) Ayres, L. B.; Gomez, F. J. V.; Linton, J. R.; Silva, M. F.; Garcia, C. D. Taking the leap between analytical chemistry and artificial intelligence: A tutorial review. *Anal. Chim. Acta* **2021**, *1161*, No. 338403. (b) Myles, A. J.; Feudale, R. N.; Liu, Y.; Woody, N. A.; Brown, S. D. An introduction to decision tree modeling. *J. Chemom.* **2004**, *18* (6), 275−285.

(17) Newman-Stonebraker, S. H.; Smith, S. R.; Borowski, J. E.; Peters, E.; Gensch, T.; Johnson, H. C.; Sigman, M. S.; Doyle, A. G. Univariate classification of phosphine ligation state and reactivity in cross-coupling catalysis. *Science* **2021**, *374* (6565), 301−308.

(18) Radhakrishnan, T. P.; Agranat, I. Measures of pyramidalization. *Struct. Chem.* **1991**, *2*, 107−115.

(19) Correlations to electronic molecular descriptors were less robust, with the best correlation to the NBO bond occupancy of the P-Ni bond. Accuracy 62%.

(20) Shahriari, B.; Swersky, K.; Wang, Z.; Adams, R. P.; de Freitas, N. Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proceedings of the IEEE* **2016**, *104* (1), 148−175.

(21) (a) Lee, R. Statistical Design of Experiments for Screening and Optimization. *Chem. Ing. Technol.* **2019**, *91* (3), 191−200. (b) Mateos, C.; Nieves-Remacha, M. J.; Rincón, J. A. Automated platforms for reaction self-optimization in flow. *React. Chem. Eng.* **2019**, *4* (9), 1536−1544.

(22) Wu, J.; Poloczek, M.; Wilson, A. G.; Frazier, P. Bayesian Optimization with Gradients. *Adv. Neural Inf. Processing Syst.* **2017**, *30*5268−5279

(23) Brochu, E.; Cora, V. M.; de Freitas, N. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. *arXiv preprint arXiv:1012.2599* **2010** DOI: 10.48550/arXiv.1012.2599

(24) We determined that Ni precatalyst had little impact on the yield or selectivity. $Ni(acac)_2$, $Ni(COD)_2$, and (1,2-dimethoxyethane) nickel dibromide were tested and found to be equivalent. $Ni(acac)_2$ was selected for the EDBO+ campaign due to its enhanced shelf life over $Ni(COD)_2$.

(25) (a) Zapf, A.; Jackstell, R.; Rataboul, F.; Riermeier, T.; Monsees, A.; Fuhrmann, C.; Shaikh, N.; Dingerdissen, U.; Beller, M. Practical synthesis of new and highly efficient ligands for the Suzuki reaction of aryl chlorides. *Chem. Commun. (Camb)* **2004**, *1*, 38−39. (b) Rataboul, F.; Zapf, A.; Jackstell, R.; Harkal, S.; Riermeier, T.; Monsees, A.; Dingerdissen, U.; Beller, M. New ligands for a general palladium-catalyzed amination of aryl and heteroaryl chlorides. *Chemistry* **2004**, *10* (12), 2983−2990.

(26) Seeman, J. I. The Curtin-Hammett principle and the Winstein-Holness equation: new definition and recent extensions to classical concepts. *J. Chem. Educ.* **1986**, *63*, 42−48.

(27) (a) Eliel, E. L.; Wilen, S. H. *Stereochemistry of Organic Compounds*; John Wiley & Sons, 1994. (b) Meyers, A. I.; Seefeld, M. A.; Lefker, B. A.; Blake, J. F. Origin of Stereochemistry in Simple Pyrrolidinone Enolate Alkylations. *J. Am. Chem. Soc.* **1997**, *119*, 4565−4566. (c) Chakraborty, S.; Saha, C. The Curtin-Hammett Principle. *Resonance* **2016**, *21*, 151−171.

(28) Conversion of **S6** was observed to **P6**-*E* and **P6**-*Z* indicating the catalytic cycle was indeed active in the background to generate any off-cycle intermediates.

(29) Tanabe, Y.; Nakatsuji, H.; Ashida, Y.; Sato, Y.; Honda, A. A General and Robust Method for the Preparation of (*E*)- and (*Z*)-Stereodefined Fully Substituted Enol Tosylates: Promising Cross-Coupling Partners. *Synthesis* **2016**, *48* (23), 4072−4080.

(30) Determined as available via the MilliporeSigma and/or Strem Chemicals online catalogs (06/2021).

(31) Xu, L. C.; Zhang, S. Q.; Li, X.; Tang, M. J.; Xie, P. P.; Hong, X. Towards Data-Driven Design of Asymmetric Hydrogenation of Olefins: Database and Hierarchical Learning. *Angew. Chem., Int. Ed. Engl.* **2021**, *60* (42), 22804−22811.

(32) High-throughput experiments were collected in duplicate to gauge experimental variability and enhance accuracy.

https://doi.org/10.1021/acscatal.4c00650
ACS Catal. 2024, 14, 4699−4708