Chemoinformatic Catalyst Selection Methods for the Optimization of Copper-Bis(oxazoline) Mediated, Asymmetric, Vinylogous Mukaiyama Aldol Reactions

Casey L. Olen[1], Andrew F. Zahrt[1], Sean W. Reilly[2], Danielle Schultz[2], Khateeta Emerson[2], David Candito[2], Neil A. Strotman[2]*, Scott E. Denmark[1]*

[1]Department of Chemistry, University of Illinois at Urbana-Champaign, 600 S. Mathews Ave, Urbana, IL 61801, USA
[2]Merck Research Laboratories, Merck & Co., Inc., 126 East Lincoln Ave Rahway NJ 07065, USA

Address Correspondence to:
Professor Scott E. Denmark
245 Roger Adams Laboratory, Box 18
Department of Chemistry
University of Illinois
600 S. Mathews Ave.
Urbana, IL 61801

tel: (217) 333-0066
FAX: (217) 333-3984
e-mail: sdenmark@illinois.edu

Abstract: A catalyst selection method for the optimization of an asymmetric, vinylogous Mukaiyama aldol reaction is described. A large library of commercially available and synthetically accessible copper-bis(oxazoline) catalysts was constructed *in silico*. Conformer-dependent, grid-based descriptors were calculated for each catalyst, defining a chemical feature space suitable for machine learning. Selection of a diverse subset of catalyst space produced an initial training set of 26 novel bis(oxazoline) ligands which were synthesized and tested for stereoselectivity in the copper-catalyzed, vinylogous Mukaiyama aldol reaction for five substrate combinations. One ligand in the training set provided 88% average enantiomeric excess, exceeding the performance of catalysts identified through an initial optimization campaign. Supervised and semi-supervised catalyst selection methods, including quantitative structure-selectivity relationship modelling, nearest neighbors analysis, and a focused analogue clustering strategy, were employed to identify an additional 12 novel bis(oxazoline) ligands. The selected ligands outperformed the initial training set hit in four out of five product classes, and in some cases demonstrated excellent enantiocontrol exceeding 95% ee. The effectiveness of the unsupervised training set selection process is discussed, and the expediency of the nearest neighbor and focused analogue approaches are contrasted with the supervised quantitative structure-selectivity relationship modelling approach.

Introduction

The development of asymmetric catalysis using chiral, small molecules represents one of the most important advances in chemical synthesis over the past four decades as reflected in the 2001 and 2021 Nobel Prizes in Chemistry.[1] The most common strategy to identify and optimize a catalyst for an asymmetric, chemical transformation has traditionally involved a combination of expert knowledge, human intuition, and serendipity. Although this approach has yielded many successes, it is still often inefficient and plagued by uncertainty when selectivity plateaus are encountered. Recent applications of informatics-driven approaches to catalyst identification and optimization have opened new avenues for algorithmically-guided selection of optimal, chiral, small-molecule catalysts.[2] Coupled with technological advances in high throughput experimentation (HTE), chemoinformatic methods can aid an experimentalist in identifying highly selective catalysts for challenging transformations that may not have been discoverable by the traditional approach.

Ever-increasing pressures to accelerate drug development and shorten timelines have resulted in rapid uptake of multiwell plate-based HTE techniques in the pharmaceutical industry over the past decade.[3] HTE approaches have seen significant development for the optimization of chemocatalytic transformations, with Suzuki-Miyaura[4] and Buchwald-Hartwig[5] cross-coupling methods emerging as the most popular.[3a,3b] Asymmetric, chemocatalytic transformations mediated by chiral, small molecules are less explored. This area is particularly challenging for the discovery of new chiral catalysts beyond the optimization of other reaction variables (*e.g.* solvent, temperature, additives). The current trend to incorporate three-dimensional molecular structure in drug design[6] and the push for greener, more atom economical asymmetric processes[7] continue to motivate HTE-driven optimization and discovery of chiral, small-molecule catalysts.

Soluble, transition-metal complexes derived from chiral ligand scaffolds encompass a diverse chemical space affecting tunable, catalytic stereoinduction well-suited for exploration by HTE approaches.[8] Enantioselective hydrogenation has emerged as the primary area of industrial focus in this respect,[3b] and several reports demonstrate the success of current HTE design principles in discovering new, optimal catalysts. Chirik *et al*. have elegantly demonstrated the HTE-driven development of cobalt-bisphosphine catalysts for enantioselective alkene

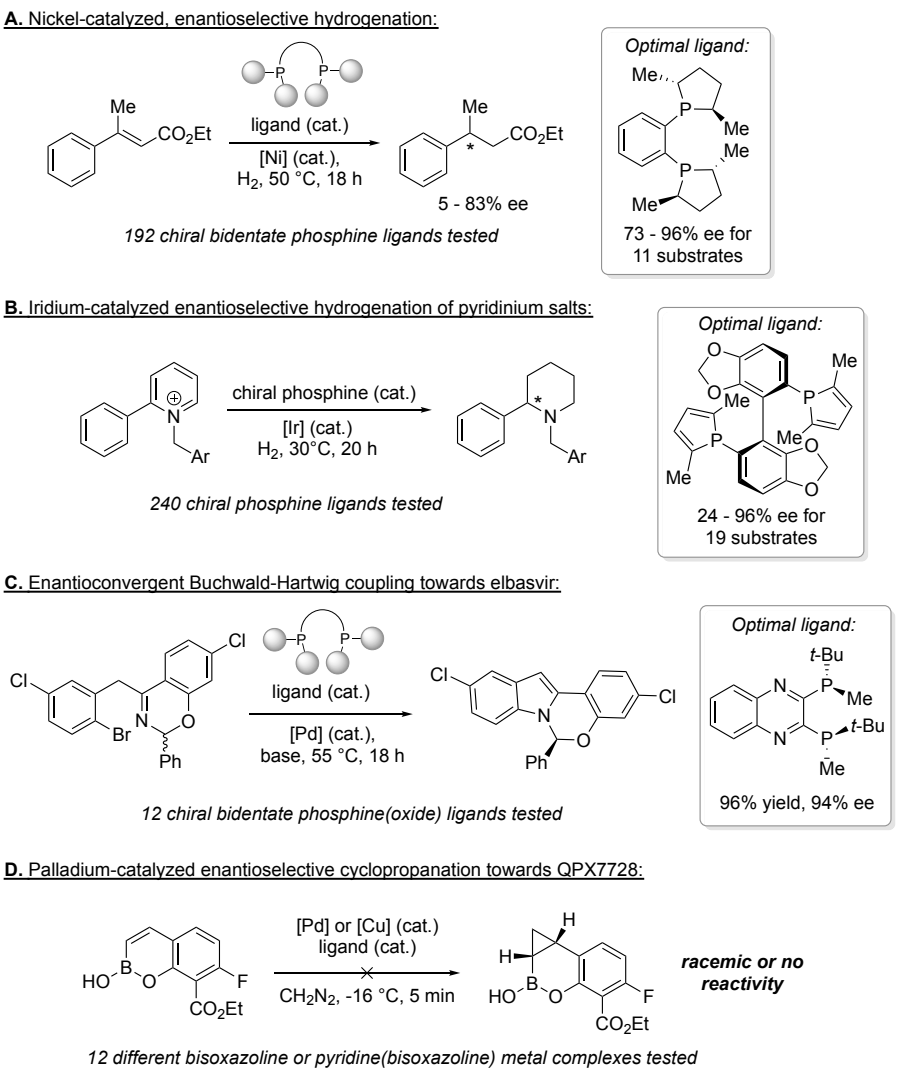

Figure 1: (**A**) Successful HTE-driven optimization of phosphine ligand identity in enantioselective hydrogenation of α,β-unsaturated esters. (**B**) Iridium-catalyzed hydrogenation of pyridinium salts. (**C**) Enantioconvergent Buchwald-Hartwig amination. (**D**) Attempted catalytic enantioselective cyclopropanation toward QPX7728.

hydrogenation, achieving good-to-excellent enantioselectivities by evaluating a 192-member chiral, bidentate phosphine ligand library.[9] This approach was later leveraged for the discovery and optimization of the nickel-bisphosphine-catalyzed hydrogenation of α,β-unsaturated esters (Figure 1A).[10] A 240-member chiral phosphine library was evaluated by Zhang *et al*. for the iridium-catalyzed, enantioselective hydrogenation of 2-substituted pyridinium salts (Figure 1B),[11] which was then extended to 3-substituted substrates by adapting a rhodium-based catalyst system.[12] Enantioselective carbon-carbon bond forming reactions have received less attention, and notable examples include the enantioconvergent Buchwald-Hartwig amination disclosed by Campos *et al.* for forging the benzoxazinoindole core of the antiviral drug elbasvir (Figure 1C),[13] and the development of a palladium-catalyzed, enantioselective cyclopropanation reaction on route to β-lactamase inhibitor QPX7728.[14]

In these examples, the chiral ligands considered for HTE design were either commercially available, or readily prepared from commercial sources. Designing HTE campaigns around ligand availability is a sensible starting point for optimization, yet this approach necessarily biases the domain of stereodifferentiating reaction components toward commercially available entities.[8] In many cases, highly stereoselective catalyst systems derived from commercially available ligands can be created and this remains a fruitful area of investigation. However, for particularly challenging transformations or if commercial sources are limited, this approach may not lead to the discovery of an optimal catalyst. For example, a catalytic, enantioselective route to QPX7728 was pursued initially, but evaluation of commercially available bis(oxazoline), pyridine-bis(oxazoline), and salen-derived chiral ligands in combination with palladium or copper sources did not lead to adequate enantioselectivity (Figure 1D).[13] Instead, the final conditions incorporated a chiral auxiliary that delivered the product in 78% ee, incurring a penalty in mass efficiency with

only moderate enantioselectivity.[15] Although the chiral auxiliary was a readily-available, chiral pool material, one must consider the possibility that a highly stereoselective, *catalytic* method could have been discovered had a broader domain of stereodifferentiating reaction components been considered. Furthermore, these examples illustrate an important trade-off in experimental design between intensive optimization of a single substrate, versus optimization of a general class of substrates. HTE applications toward substrate generality have seen increasing interest,[16] while many industrial applications remain highly specific for the relevant process development.

The success of an HTE campaign is fundamentally dependent on the reaction parameter spaces defined by the experimentalist for empirical exploration.[17] Catalyst identity is a discrete reaction parameter, which presents a unique challenge when defining suitable parameter domains. Whereas well-defined continuous domains (*e.g.* temperature or concentration) necessarily contain locally optimal parameters and are inherently suitable for statistically designed, experimental exploration (*i.e.* Design of Experiments, DoE),[3a,18] discrete parameter domains are not associated with a chemically-meaningful distance metric (*i.e.* measure of similarity) *a priori*, complicating their rational exploration. Expert knowledge is required to define a suitable mapping from each discrete catalyst into a continuous vector space in which locality can be assessed mathematically. A continuous vector description (or "featurization") of a catalyst in this manner[19] provides a representation that is suitable for DoE[20] and statistical modelling of a quantitative structure-selectivity relationship (QSSR) between catalyst identity and empirical stereoselectivity.[2a,21] In practice, catalyst featurization can be constructed on the basis of empirical data (*e.g.* linear free energy relationships)[22] or calculated properties intended to capture the relevant chemical characteristics of the species. Within the realm of transition-metal catalysis, phosphine ligand space featurization has been extensively developed, spurred initially by the foundational

calculation of Tolman cone angles,[23] progressing through computed properties of various phosphine-metal complexes,[24] and ultimately leading to the recent disclosure of the large *kraken* database of quantum-mechanically described monodentate phosphines.[25] Sigman *et al*. have identified a representative, mostly achiral ligand subset in this database,[26] and recently disclosed chiral bisphosphine ligand QSSR models for Hayashi-Heck reactions.[27] Whereas the featurization, exploration, and modelling of phosphine ligand space has received significant attention, we believe that data-driven workflows with a general form of: (1) featurizing broad chemical spaces, (2) selecting diverse representative subsets, (3) acquiring empirical data in some reaction manifold, and (4) constructing QSSR models for the potential prediction of new optimal catalysts,[28] should find application to myriad ligand classes and reactions.

The Denmark group has a long-standing interest in the behavior of privileged oxazoline-containing ligand scaffolds, particularly the bis(oxazoline) (BOX) class.[29] Encouraged by our success in applying conformer-averaged grid-based (CAGB) descriptors for catalyst featurization, subset selection, and QSSR modelling of chiral, Bronsted acid-catalyzed enantioselective *N*-acyl imine addition reactions[30] and atroposelective iodination reactions,[16c] we next sought to apply these methods to the optimization of enantioselectivity in a copper-BOX-catalyzed, vinylogous, Mukaiyama aldol reaction (VMA). This academic-industrial collaboration describes an effective workflow for optimizing catalyst identity in a challenging reaction manifold and originated as a response to the failure of an initial HTE design that considered commercial or commercially-derived catalysts only. Herein we describe CAGB average steric occupancy (ASO) descriptors for the featurization of a large *in silico* library of BOX ligands, disclose a diverse subset of BOX ligands for empirical evaluation, and demonstrate a semi-supervised modification of our previously-described chemoinformatic workflow for the optimization of the VMA reaction

between silyloxyfuran **1** and aldehydes **2a** – **2e** toward PRMT5 inhibitor analogues (Figure 2).

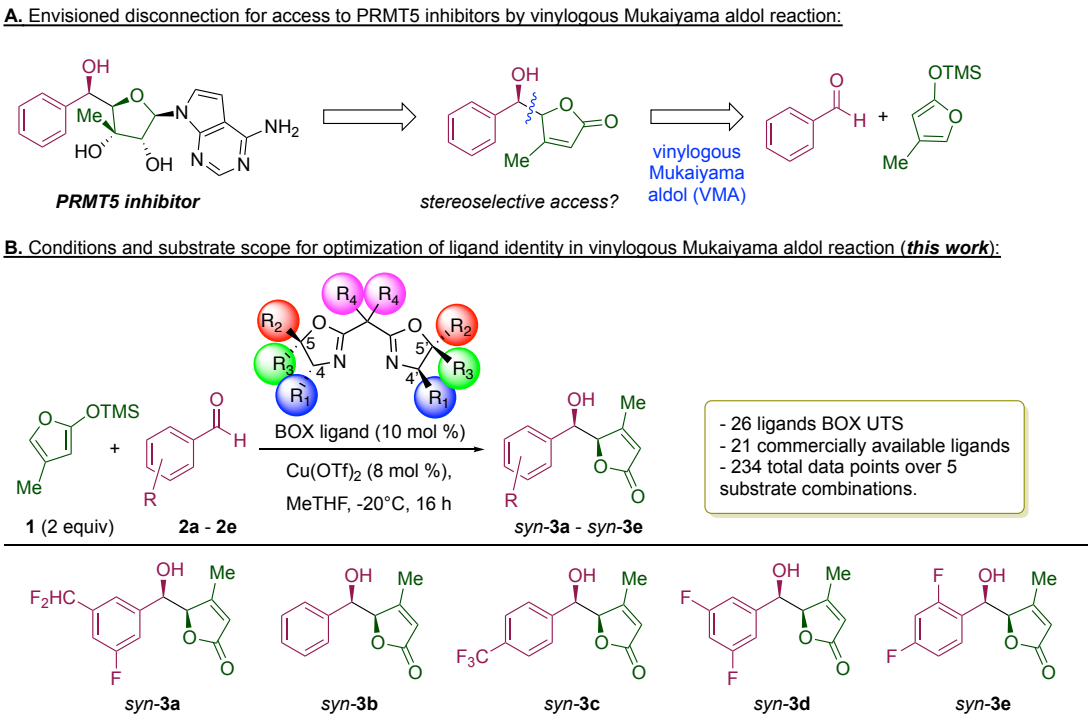

Figure 2: (**A**) Access to PRMT5 inhibitor by enantioselective vinylogous Mukaiyama aldol reaction. (**B**) Reaction conditions and substrate scope for optimization of copper-bis(oxazoline)-catalyzed vinylogous Mukaiyama aldol reaction.

Protein arginine methyl transferase 5 (PRMT5) is a post-translational modification enzyme that catalyzes the formation of symmetric dimethylarginine (sDMA).[31] PRMT5 plays an essential role in maintaining cancer pathogenesis, with high protein expression attributed to increased tumor size in lung,[32] breast,[33] and liver[34] cancers, including increased cancer metastasis.[35] Upregulation of PRMT5 has been directly linked to poor prognosis in a variety of malignancies.[36] Given the consequences of PRMT5 overexpression in cancer treatment, PRMT5 inhibitors have recently emerged as potential antineoplastic agents.[37]

BACKGROUND

Methods for the Catalytic Asymmetric Synthesis of Chiral Butenolides. Chiral

butenolides are important intermediates in the synthesis of natural products[38] and medicinal compounds.[39] Butenolides are commonly accessed in an enantioenriched form by the VMA reaction of a 2-siloxyfuran with an aldehyde, which can be stereochemically controlled using an appropriate chiral catalyst.[40] Enantioselective VMA reactions to forge chiral butenolides have been investigated with a variety of chiral organic-[41] and transition-metal based catalysts with various ligands, including those derived from oxazoline,[42] salen,[43] and binaphthol (BINOL)[44] scaffolds. However, challenges remain in realizing a general method that provides high diastereo- and enantioselectivity while accommodating variable substitution patterns on the 2-siloxyfuran and aldehyde reactants.

Catalyst Featurization and QSSR Modelling in the Cu-BOX Manifold. The BOX ligand scaffold is privileged for a variety of organic transformations and presents multiple points of combinatorial diversity permitting access to extensive analogue generation.[45] Early approaches to modelling Cu-BOX catalysis in enantioselective Diels-Alder reactions used a comparative molecular field analysis (CoMFA) approach,[46] which is conceptually related to CAGB descriptors, but which requires more computationally expensive probe calculations. It was also shown that simple, linear correlations can be drawn between the continuous chirality measure of a BOX ligand and the corresponding enantioselectivity in the Diels-Alder reaction.[47] More recently, linear regression models employing DFT descriptors have been developed to predict enantioselectivity in several different Cu-BOX-catalyzed, reaction manifolds.[48]

Failure of Commercial Catalysts in the HTE Optimization of a VMA Reaction on Route to a Class of PRMT5 Inhibitors. In the course of developing a scalable route to access stereochemical analogues of PRMT5 inhibitor therapeutic targets (Figure 2A),[49] we encountered a VMA reaction particularly refractory to optimization of stereoselectivity in the synthesis of chiral

butenolide product *syn*-**3a** from 4-methyl substituted 2-siloxyfuran **1** and aldehyde **2a** (Figure 3A). Initial high-throughput screening campaigns considered 147 different chiral catalysts, including organocatalysts of various classes (squaramides, thioureas, ureas, diols, phosphoric acids, and bis-triflimides) and chiral titanium-BINOL catalysts. A squaramide catalyst that provided *syn*-**3a** with 82% ee and a 3.6:1 *syn*/*anti* ratio was discovered (Figure 3B).

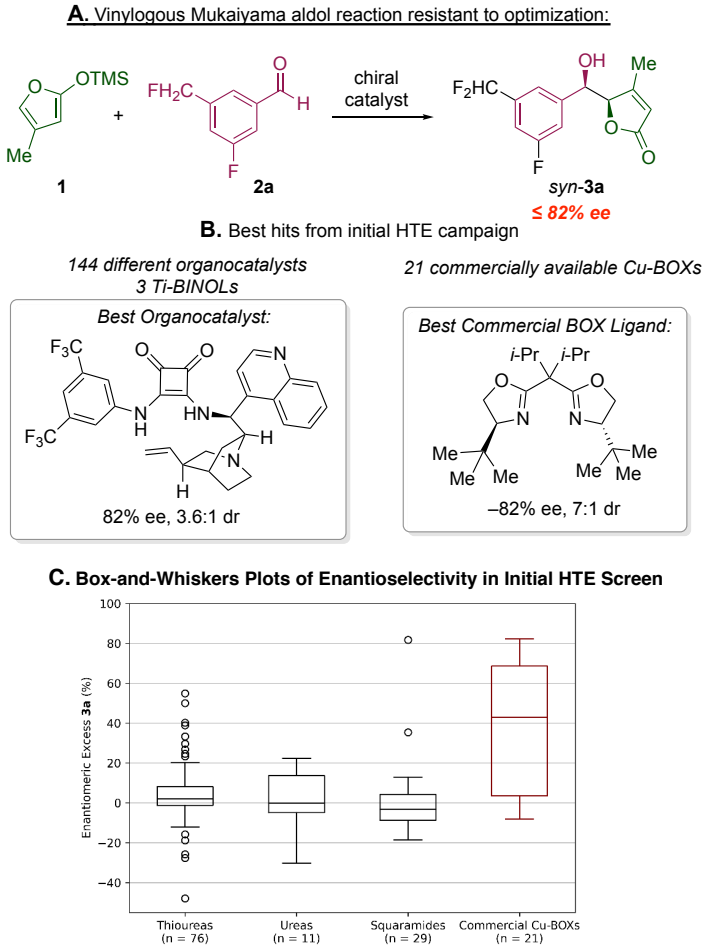

Figure 3: (**A**) Vinylogous Mukaiyama aldol reaction between **1** and **2a** and summary of HTE catalyst screen**.** (**B**) Optimal catalysts discovered in initial HTE campaigns testing commercial catalysts. (**C**) Box-and-whiskers plots showing selected data acquired in initial HTE campaign.

A subsequent high-throughput survey of copper-BOX catalysts derived from commercially available BOX ligands identified a 4,4'-*tert*-butyl-substituted BOX ligand that provided *syn*-**3a** with 82% ee and a 6:1 *syn*/*anti* ratio (Figure 3B). The improved diastereoselectivity provided by

the copper-BOX system suggested that a more extensive evaluation of this platform was warranted. Furthermore, the reaction enantioselectivity displayed significant response to changes in BOX ligand structure (Figure 3C), suggesting that this scaffold could be tuned to affect higher enantioselectivity. However, further optimization of the copper-BOX catalyzed VMA reaction was limited by the commercial availability of testable ligands (see the Supporting Information). Because the BOX scaffold has multiple points of diversification, it was clearly amenable to extensive analogue generation including 4,4'-substitution, 5,5'-substitution, bridging methylene substitution, and relative stereochemical arrangements which permit an enormous collection of possible catalysts (Figure 4). This feature provided the opportunity to explore reactivity and selectivity in a large chemical space of synthetically accessible copper-BOX catalysts. Additionally, multistep syntheses (7-9 steps) are required to access more challenging ligands, so an informatics-guided approach is desirable to rationally identify ligand candidates *in silico* prior to their synthesis and testing. In considering the synthetic challenges and inherent combinatorial complexity of the BOX scaffold, we sought to apply our chemoinformatic catalyst optimization workflow that had previously been successful in a proof-of-principle optimization of chiral Bronsted acid-catalyzed synthesis of *N,S*-acetals.[27]

First, a large *in silico* library of BOX ligands was constructed that captured the breadth of possible catalyst structures accessible by a given robust synthetic route. Conformer-averaged grid-based (CAGB) average steric occupancy (ASO) descriptors were calculated to capture the steric effects of each catalyst. Whole-molecule featurization with ASO provides granularity in representing the potentially subtle differences in three-dimensional catalyst shapes. The catalyst chemical space was analyzed by unsupervised clustering analysis to identify diverse ligands, which were further triaged for synthetic accessibility (Figure 4A). The catalysts were synthesized, and

data was collected over various substrate combinations with this universal training set (UTS), to provide labelled, catalyst-substrate-enantioselectivity data points that can be used for machine learning methods (Figure 4B, 4C).

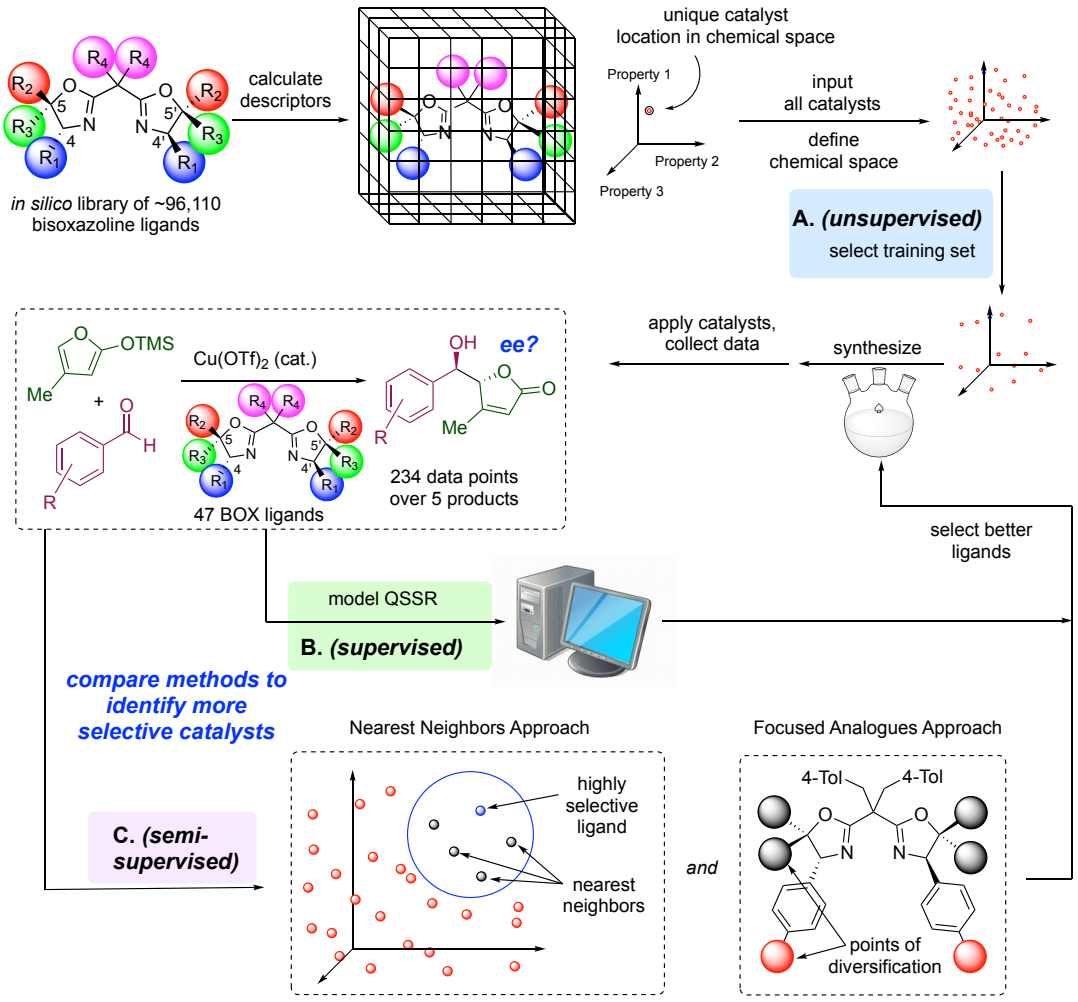

Figure 4: Chemoinformatic workflow for this work. (**A**) Universal training set selection from bis(oxazoline) *in silico* library. (**B**) Supervised QSSR modelling approach. (**C**) Semi-supervised nearest neighbors and focused analogues approaches.

RESULTS

Universal Training Set Selection from Bis(oxazoline) In Silico Library. A combinatorial, *in silico* library of chiral, C_2-symmetric BOX ligands was constructed comprising 267 unique substituents at the oxazoline 4,4'-positions, three homodisubstituted and three

heterodisubstituted combinations at the oxazoline 5,5'-positions including *syn/anti* stereochemical arrangements relative to the 4,4'-position, and 40 geminal bridging group substituents. This was supplemented with 40 indenyl BOX ligands with variable bridging group substitution. Structures that could not be automatically checked for stereochemical configuration were omitted, producing a final *in silico* library of 96,110 members. Conformers were generated for each catalyst using the distance geometry method implemented in RDKit and ASO descriptors were calculated from each conformer distribution. A subset of the in silico library was selected by a *k*-means analysis. The cluster medoids were triaged for synthetic considerations and substitutes were chosen with a focus on chemical diversity while maintaining adequate coverage of the ASO chemical space (see the Supporting Information, Figure 5A, 5C). Twenty six ligands were identified by this approach and synthesized using the methods recently described by these laboratories (**L1a**-**L1z**, Figure 5A).[50] These ligands were tested for enantioselectivity in the VMA reaction between 4-methyl-2-siloxyfuran **1** and five different aldehyde electrophiles in parallel with a supplementary set of 21 BOX ligands that are either commercially available or accessible from commercial ligands in one step (see the Supporting Information, referred to henceforth as 'commercial ligands').

Data Acquisition and Analysis of Results in the Vinylogous Mukaiyama Aldol Reaction. Reactions were run in duplicate in a Para-dox® 96-well photoredox aluminum reaction block. Enantiomeric excesses were assessed by chiral SFC following quench of the crude reaction mixture. Duplicate data were averaged, providing 234 trustworthy data points over five different substrate combinations. To directly compare the magnitude of enantioinduction from ligands with different stereochemical configurations, enantiomeric excesses were normalized based on the configuration at the oxazoline 4,4'-positions of the corresponding BOX ligand. Enantiomeric excesses corresponding to ligands with 4,4'-(*S,S*) configuration were inverted, while those

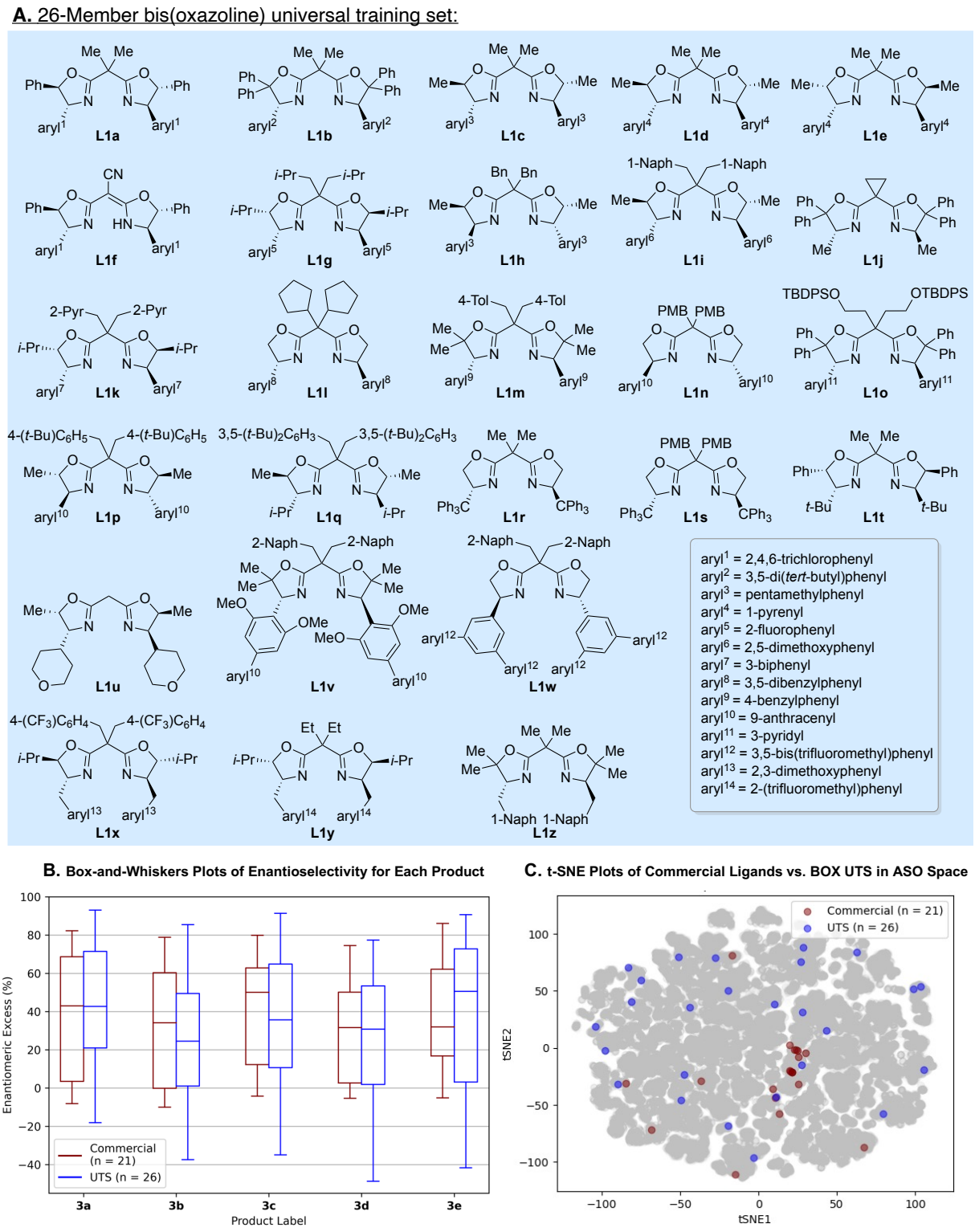

Figure 5: (**A**) 26-member BOX training set synthesized and tested in the VMA reaction. (**B**) Box-and-whiskers plots of enantiomeric excesses for BOX training set ligands versus commercially available BOX ligands for each product. (**C**) t-SNE plots of BOX training set versus commercial ligands in ASO chemical space.

from 4,4'-(*R,R*) ligands were left unchanged (see the Supporting Information). BOX ligand **L1m**, a UTS member, was identified as a new optimal ligand for each of the five products, achieving 88% average enantiomeric excess. It is worth noting that *syn*/*anti* diastereoselectivity was also determined for each ligand (see the Supporting Information), however optimization of enantioselectivity for the *syn* diastereomer was the primary focus of this work. **L1m** afforded product **3a** with excellent 22:1 *syn/anti* diastereoselectivity.

The selectivity spread of the UTS ligands exceeded that of the 21 commercially available ligands for all products (Figure 5B) and provided more even coverage of BOX ASO chemical space (Figure 5C). A broader UTS interquartile range was observed for products **3c**, **3d**, and **3e**. The normality of the distributions of enantiomeric excesses were assessed by Anderson-Darling tests[51] at a 95% confidence level (CL), indicating that the UTS data were approximately normally distributed for products **3a**, **3b**, **3c**, and **3d**, whereas product **3e** was non-normal. The commercial ligand set produced approximately normal data for only products **3d** and **3e**. The variance of the data obtained from the UTS ligands was compared to those obtained from commercial ligands by a one-tailed F-test. The UTS data was significantly more varied than the commercial data for product **3e** only (90% CL, $p = 0.08$).

Supervised Model Development and Validation. Supervised models for ligand- and substrate-dependent QSSR relationships were constructed using all acquired data and employing leave-one-catalyst-out validation (LOCO). Attempts at featurizing substrate combinations and products with CAGB ASO and average electronic indicator field (AEIF), sterimol descriptors, and various DFT descriptors did not result in more predictive models compared to simple one-hot encoded substrate descriptors (see the Supporting Information). A substrate-agnostic feature selection protocol was used to identify and aggregate catalyst CAGB descriptors relevant to

catalyst-dependent QSSR. The resulting models had modest predictive performance with high absolute errors (> 0.2 kcal/mol) and suffered from significant underprediction of the most selective ligand **L1m** (Figure 6). In view of the low reliability of the supervised models, only one ligand was synthesized from the model prediction (**L2**). Given the relative lack of high selectivity data for model training and considering the limited size of the dataset overall, it was expected that acquiring additional training data with more selective ligands might provide models with more accuracy in the range of high enantioselectivities. *In silico* ligands proximal to the UTS hit **L1m** in ASO feature space should engender similar steric enantiocontrol in the reaction and would likely provide similarly high enantioselectivity. **L1m** therefore served as a starting point for a nearest neighbors analysis in the ASO feature space.

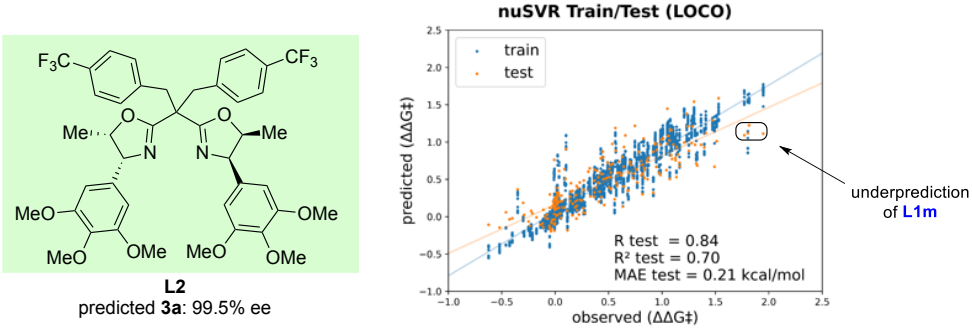

Figure 6: Underprediction of highly selective ligands for supervised model and ligand selected for out-of-sample model testing.

Nearest Neighbors and Focused Analogue Ligand Selection. Two, simple, semi-supervised, catalyst selection approaches were pursued: a nearest neighbors (NN) analysis and a focused analogue (FA) method. In the NN approach, ligands were selected based on their proximity in chemical feature space to the UTS hit **L1m** (Figure 7A). The Euclidean distance in ASO chemical space from **L1m** to each member of the *in silico* library was calculated. Ligands proximal to **L1m** with unique 4- and 5- position substitution were triaged, and five new BOX cores were selected for synthesis and testing (Figure 7A, **L3a**-**L3e**). Two analogues with variation at the

methylene bridging position were also synthesized for the first two nearest neighbors (**L3a'**, **L3b'**).

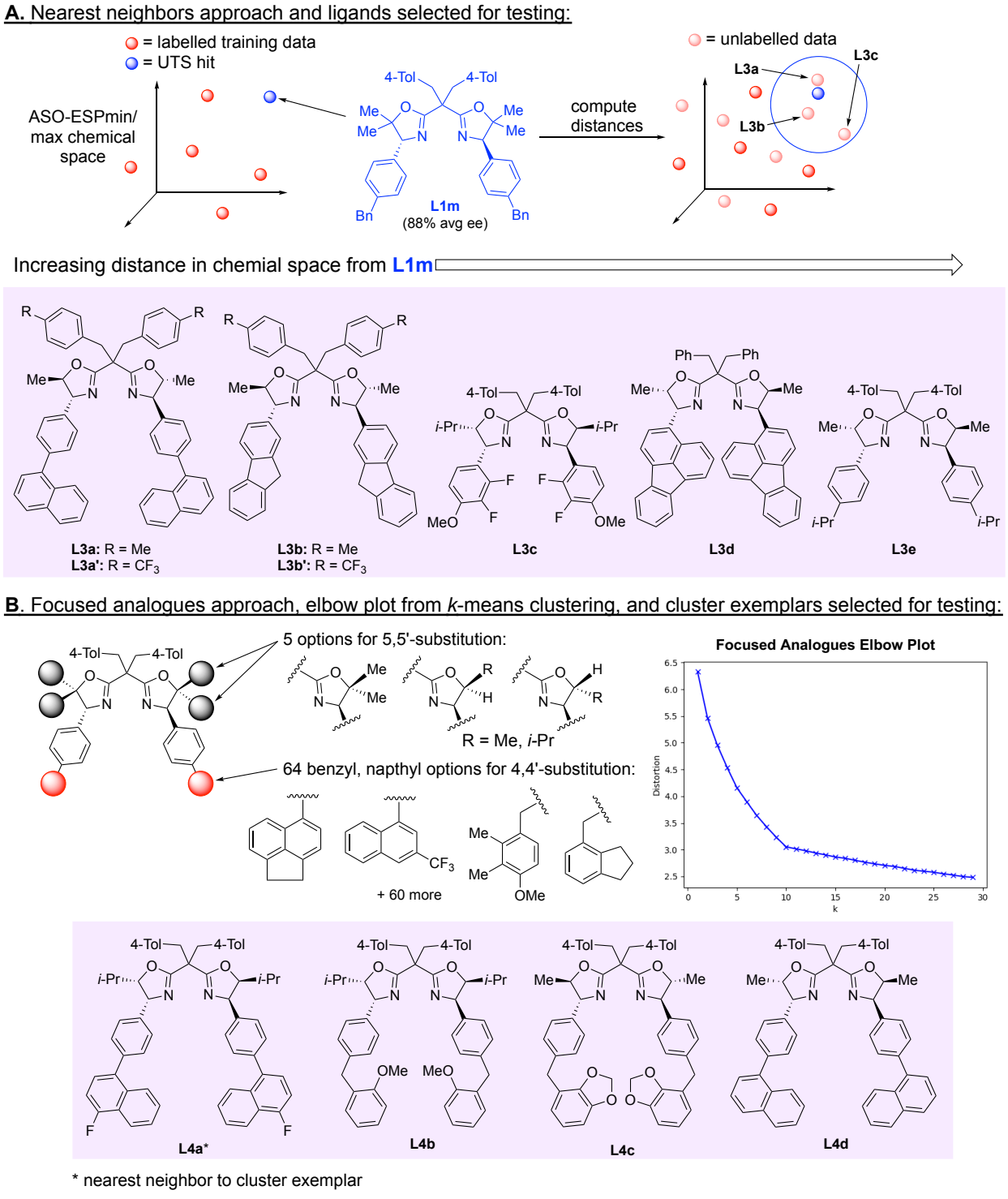

Figure 7. (**A**) Nearest neighbors approach and analogues selected for synthesis and testing. (**B**) Focused analogues approach, elbow plot from k-means clustering, and cluster exemplars selected for synthesis and testing.

In the FA approach, a new in silico library was constructed containing focused variation around the oxazoline 4-positions of **L1p** and the first nearest neighbor **L3a** (Figure 7B). Synthetic concerns and commercial availability were considered to construct a 320-member combinatorial library comprising 64 options for 4,4'-BOX substitution and five options for 5,5'-BOX substitution, while conserving the 4-tolylmethyl bridging substituent present in **L1m**. The focused analogue library was then clustered using a *k*-means clustering algorithm with six clusters. Exemplars of the four clusters not represented by **L1m** or **L3a** were selected for synthesis and testing in the reaction (Figure 7B, **L4a-L4d**).

Altogether, one ligand was identified by supervised QSSR modelling (**L2**), seven ligands were identified by the semi-supervised NN approach (**L3a-L3e**, **L3a'**, **L3b'**) and four ligands were identified by the semi-supervised FA approach (**L4a-L4d**). All 12 ligands were tested in duplicate with the five substrate combinations of interest (Figure 8). The ligand from supervised modeling, **L2**, performed less selectively than predicted, providing a modest 82% ee for product **3a** with an absolute error of 2.06 kcal/mol.

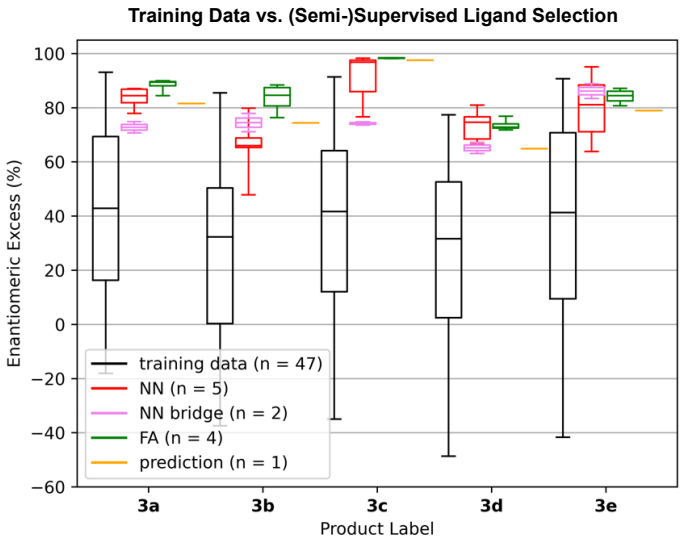

Figure 8. Box-and-whiskers plots of enantioselectivity data for model prediction, nearest neighbors, and focused analogue selections over five products.

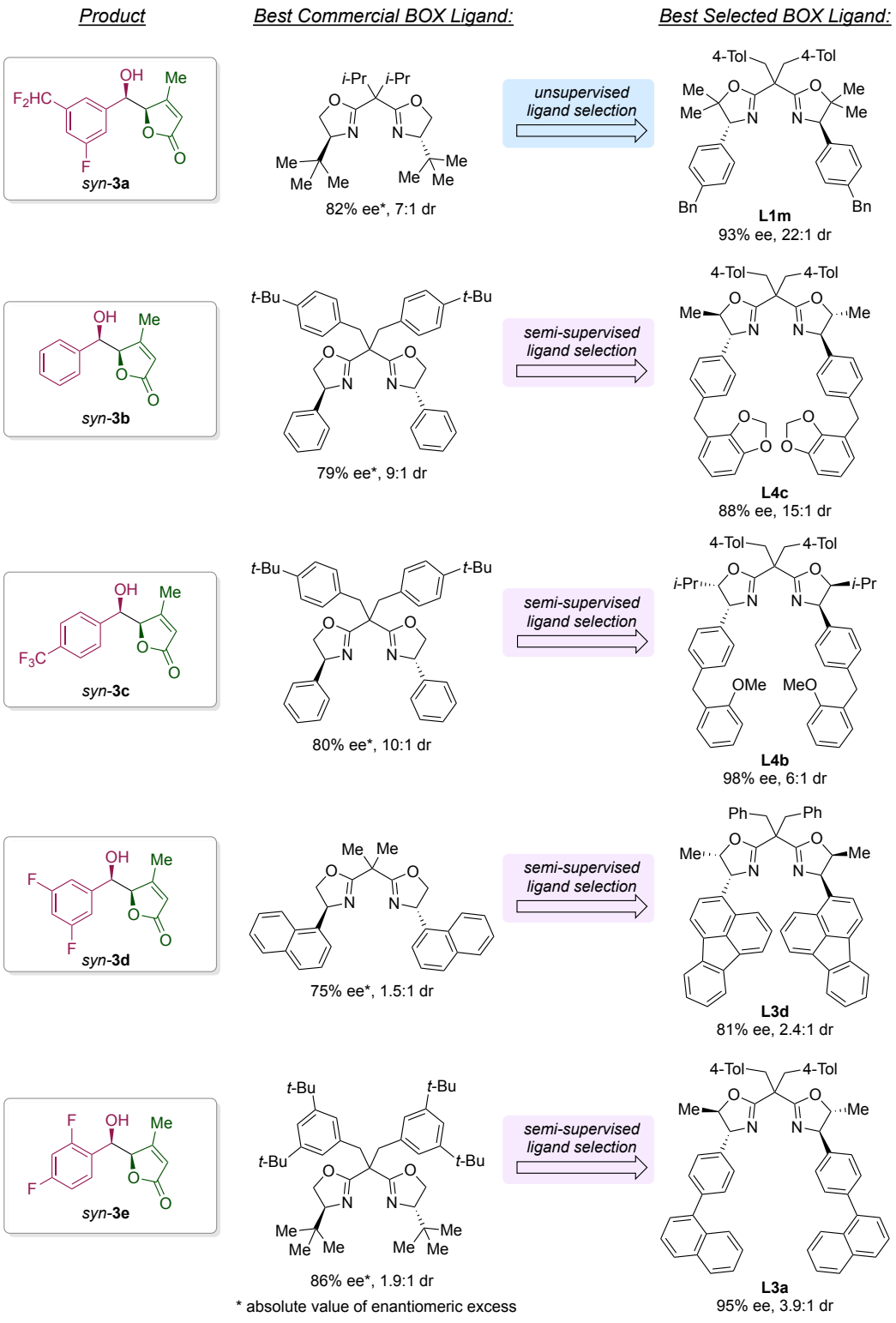

Figure 9. Improvement in reaction enantioselectivity from initial commercial ligand screen to final optimal ligands for each product.

The semi-supervised approaches were more fruitful.Focused analog exemplars **L4c** and **L4d** emerged as new optimal ligands for substrate combinations **3b** and **3c**, delivering 88% ee and 98% ee, respectively. Nearest neighbor ligands **L3d** and **L3a** were new optima for **3d** and **3e** at 81% ee and 96% ee, respectively (Figure 9). Nearest neighbor bridging group analogues performed well for **3b** and **3e**, but underperformed the actual nearest neighbors for products **3a**, **3c**, and **3d** and in no cases emerged as new optimal ligands. Overall, four out of five substrate combinations were optimized past the initial UTS hit by the semi-supervised approaches. All substrates were delivered in good to excellent enantioselectivity (81 – 98% ee) following UTS selection and one iteration of the semi-supervised selection methods (Figure 9).

DISCUSSION

Steric Diversity of BOX UTS and Limitations of Commercially Available BOX Ligands. The BOX ligand UTS was selected with the intention of capturing more chemical diversity in the ASO chemical space than would be provided by ligands that are currently commercially available. The subset selection was informed by a *k*-means analysis and the elbow method, although expert triaging for synthetic considerations led to some deviation from the optimal *k*-means solution (see the Supporting Information). Even so, when compared with a representative subset of ligands that are commercially available or accessible from a commercial ligand in one step, the selected UTS displayed greater total variance in ASO chemical feature space. Although synthetic methods to access BOX ligands have improved in recent years,[45b, 50] the number of commercial ligands is still rather limited in their representation of the steric space accessible with the BOX ligand scaffold. Indeed, as of submission of this work the diversity of 4,4'-substitution displayed in commercially available BOX ligands is limited primarily to simple

alkyl- and aryl- substituted ligands and their corresponding methylene homologues, with 4,4'-phenyl substitution being the most common (Figure 10A). Diversity in 5,5'-substitution is even more limited, with most examples displaying unsubstituted, non-stereogenic 5,5'- substitution. Substitution at the methylene bridging position in commercial ligands is somewhat more diverse, but mostly limited to benzylic arene substituents and some spirocycles (for full details of commercial search, see the Supporting Information). Rational *in silico* library construction considering synthetic accessibility ensures that the breadth of accessible BOX ligand steric effects is represented in the ASO chemical feature space. Coupled with a selection method emphasizing diversity, the BOX UTS represents unique steric contributions that are not captured in commercially available ligands alone (Figure 10B). It is worth noting that owing to the sparse sampling of the very large BOX *in silico* library, one might expect a random selection of 26 ligands

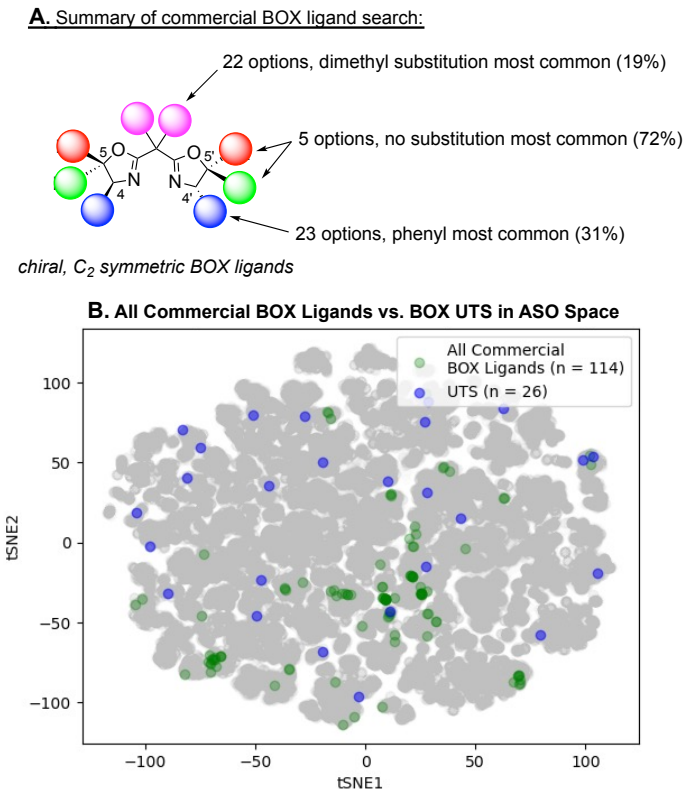

Figure 10. (**A**) Summary of SciFinder search results for commercial, enantiopure C_2-symmetric BOX ligands. (**B**) All commercial BOX ligands versus BOX UTS in ASO feature space.

to be as diverse as the ligands presented herein. Empirical analysis of 1,000 different random subsets from the BOX ligand in silico library demonstrates that random selection tends to be less diverse than an informed selection, although representation of the space is quite similar, a consequence of the deviation from the optimal *k*-means solution due to synthetic considerations (see the Supporting Information).

The diversity of the BOX ligand UTS in ASO feature space is mirrored in the wide range of empirically observed enantiomeric excesses in the VMA reaction (Figure 5B). This outcome is expected given the high variance of the ligands in ASO feature space. The identity of the aldehyde and the corresponding butenolide product can bias selectivity across all catalysts. Product **3d** bearing 3,5-difluoro substitution on the arene ring tended to be less selective with all catalysts and did not show selectivities > 80% in the initial UTS screening. Product **3e** with 2,4-difluoro substitution tended to be more selective, as evidenced by the observed highest median UTS selectivity compared to the other products (Figure 5B). It is worthy of note that 4-methyl substitution on the 2-siloxyfuran starting material **1** had a significant impact on stereoselectivity. Omitting this substituent led to racemic products for most catalysts (see the Supporting Information). Finally, ligand **L1o** tended to show a generally inverted sense of stereoinduction compared to the other catalysts for all products. This ligand, containing a 3-pyridyl substituent at the stereogenic oxazoline 4,4' positions, could permit a different ligand binding mode than the other UTS members, leading to an inverted sense of stereoinduction.

Motivation for Semi-Supervised Selection Methods and Considerations for Substrate-Catalyst Interactions. Given that **L1m** was the most selective UTS ligand for all five products, the nearest neighbors and focused analogues approaches emerged as logical next steps to explore the space around this selective ligand. Initially, it seemed sensible to explore the local space around

L1m to understand the limitations encountered with supervised QSSR modelling. Very few ligands in the initial screening campaign exceeded 80% ee, so there were few labelled examples for model training in the range of selectivities of interest for further ligand optimization. The nearest neighbors and focused analogues approaches were intended to provide more data within this region to improve the robustness of the models. Furthermore, given that the nearest neighbors and focused analogues are rather similar in feature space to the highly selective ligand (see the Supporting Information), these approaches serve to probe the continuity of the selectivity-response surface within the chemical feature space. A continuous selectivity-response surface is appropriate for predictive modelling, and large discrepancies in observed data between a given catalyst and its nearest neighbor ('activity cliffs,' or in this case, 'selectivity cliffs')[52] indicate that the choice of catalyst featurization may be inadequate for regression model development (Figure 11A). In our approach, the distance in chemical space to **L1m** was used to prioritize ligands for synthesis, which were further triaged on the basis of structural considerations by the experimentalist. The resulting selection deviates from an exact nearest neighbors approach, but the response surface is still generally conserved for products **3a**, **3b**, and **3e**, where **L1m** and the closest nearest neighbor ligand **L3a** emerge as the first or second-most selective ligands when focused analogues are omitted. Products **3c** and **3e** have a discontinuous response surface, as optimal nearest neighbor enantioselectivities are achieved with more distal ligands **L3d** or **L3o**.

The nearest neighbors, selected from the larger *in silico* library, tended to be more proximal to **L1m** in ASO chemical space than the focused analogues (Figure 11B and 11C). For substrates **2a**, **2b**, and **2c**, the focused analogue ligands generally outperformed the nearest neighbors, indicating that at the BOX 4,4'-positions, 4-(1-naphthyl)phenyl and 4-benzylphenyl-substitution

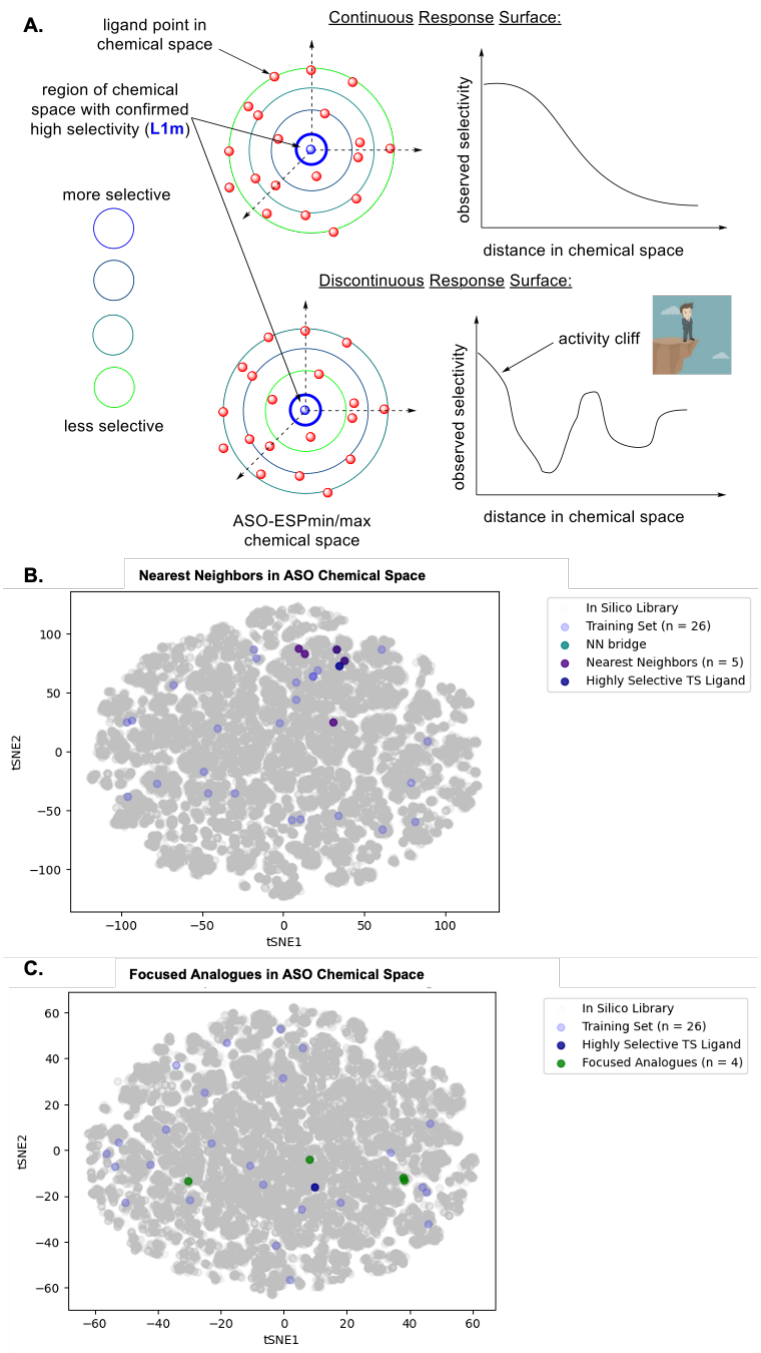

Figure 11. (**A**) Graphical representation of discontinuous versus continuous structure-selectivity response surfaces. Location of nearest neighbors (**B**) and focused analogues (**C**) in ASO chemical space.

is especially advantageous for these substrates and some continuity exists in the structure-selectivity response surface around ligand **L1m**. Indeed, **3c** is produced with excellent

enantioselectivity with all the focused analogues. This outcome indicates that the focused analogue scaffold engenders significant steric control in the stereo-determining step of this reaction, perhaps owing to favorable π-stacking interactions between the pendant arenes of the ligand and the electron-deficient π system of aldehyde **2c**. In contrast, optimal stereoselectivities for products **3d** and **3e** were achieved with nearest neighbors. Aldehydes **2d** and **2e** are more π-rich than **2a**, **2b**, or **2c**, so it is not surprising that a more significant deviation from the **L1m** scaffold is necessary to achieve higher selectivities with these products. Overall, these results demonstrate that multiple 'islands' of high selectivity exist within the ASO feature space. Products **3a** – **3c** tend to reach optima in a more similar region to the UTS hit **L1m**, whereas the more dissimilar products **3d** and **3e** require more dissimilar ligands from **L1m**. Therefore, it can be concluded that selectivity in this reaction is determined not only by ligand effects, but also by significant substrate-ligand interactions. New descriptors capturing ligand electronic and dispersion effects could provide a more meaningful feature space for examining these effects, and efforts in this area are underway.

Considerations for Applying Semi-Supervised and Supervised Catalyst Selection Methods. Although respectable QSSR models could be constructed with the BOX UTS training data, the models could not be satisfactorily validated in out-of-sample ligand predictions, particularly for the highly selective ligand **L1m**. This behavior was assessed by a leave-one-catalyst-out validation scheme, both with and without substrate descriptors. The CAGB descriptors used for modelling outperformed one-hot and random descriptors in control tests, but the resulting QSSR models chronically underpredicted ligands that were empirically demonstrated to have high selectivity (see the Supporting Information). This result is likely obtained because: (1) the bias in the data set toward ligands that have moderate to low selectivity and (2) discontinuity in the SSR surface for these intermediate ligands and the highly selective ligand **L1m**. QSSR modelling is

further complicated by the high dimensionality of the CAGB descriptors which require feature selection or dimensionality reduction techniques to produce a lower-dimensional feature space amenable to modelling the small dataset generated in this study (see the Supporting Information). Efforts to realize effective, unsupervised dimensionality reduction techniques for these descriptors are underway with the intention of developing a workflow that is more appropriate for the smaller datasets that are typically encountered in catalyst optimization campaigns.

The nearest neighbors and focused analogues approaches can be seen as very simple, semi-supervised catalyst selection techniques. Although these techniques may simply suggest ligands that are intuitively similar to the ligand that spurred the semi-supervised exploration (*i.e.* **L1m**), the additional benefits of these methods are contingent on the composition of the *in silico* library and the choice of catalyst featurization technique both implemented by the designer. The full BOX library and the focused analogues library are constructed with attention to synthetic accessibility and precursor commercial availability. The libraries are intended to be broadly representative of possible BOX ligands, without including intuitive human biases. Therefore, the prospect of selecting ligands that are non-intuitive for the experimentalist is bolstered by an unbiased *in silico* library. These semi-supervised methods are also easily actionable. QSSR modelling requires significant time, effort, and specialized knowledge, whereas the nearest neighbors and focused analogues approaches described herein can be applied by an experimentalist without requiring significant troubleshooting and specialist knowledge. In combination with our newly developed chemoinformatics package designed for processing large *in silico* libraries, the entire workflow, starting from a CDXML file of combinatorial ligand substituent fragments to ligand recommendation, can be accomplished with minimal programming expertise.

CONCLUSION AND OUTLOOK

Unsupervised subset selection was used to identify a diverse set of 26 BOX ligands from a large *in silico* library of synthetically accessible ligands. The diversity of the BOX UTS led to the identification of a highly selective ligand **L1m** for the VMA reaction between 2-siloxy-4-methylfuran **1** and five different aldehydes substrates **2a** – **2e**. The BOX UTS disclosed herein is significantly more varied than commercially available ligands and thus represents ligands that have unique steric properties compared to those which have been disclosed to date. While initial HTE campaigns considering commercial ligands did not reveal highly selective ligands, BOX UTS selection provided a subset with a wide range of observed enantiomeric excess and identified a ligand with better selectivity than any of the commercial ligands considered in the first round of HTE optimization. A nearest neighbors analysis and a focused analogues exemplar selection technique applied in the vicinity of the selective ligand **L1m** led to the identification of 11 new BOX ligands. These ligands reached new optimal selectivities for four out of five VMA products, including achieving 93% ee and 22:1 *syn/anti* for **3a**, an intermediate in the synthesis of a PRMT5 inhibitor under development. These approaches are simple, easily actionable, and effective at identifying more selective ligands for the VMA reaction. With BOX ligands **L1a-L1z** on hand, future optimization campaigns using this ligand scaffold will be greatly accelerated. Future work will leverage the data generated from the semi-supervised approaches for the development of more accurate QSSR models. Graph-based dimensionality reduction techniques are being developed to produce information-rich lower-dimensional feature spaces from the multidimensional CAGB descriptors that may be more appropriate for QSSR modelling in data limited scenarios.

■ ASSOCIATED CONTENT

SUPPORTING INFORMATION

The Supporting Information is available free of charge on the ACS Publications website at DOI: **XX**. Summary of experimental procedures, characterization data for all new compounds along with copies of ^{1}H, ^{13}C, and ^{19}F NMR spectra, HTE catalyst survey data tables, X-ray crystallographic data, descriptor calculations, clustering analysis, and computational methods are provided (PDF).

AUTHOR INFORMATION

Corresponding Author

*Email: neil_strotman@merck.com

*Email: sdenmark@illinois.edu

ORCID

Casey L. Olen: 0000-0002-8621-2973

Andrew F. Zahrt: 0000-0002-1835-5163

Sean W. Reilly: 0000-0002-1656-1895

Danielle Schultz: 0000-0002-4811-2218

Neil A. Strotman: 0000-0002-5350-8735

Scott E. Denmark: 0000-0002-1099-9765

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We are grateful to the National Science Foundation for financial support of the Molecular Maker Laboratory Institute (NSF CHE2019897) as well as for NSF CHE1900617. We would like to thank Steve Silverman, Kevin Belyk, and Ryan Cohen for their assistance and Bijay Shrestha and Adon Kwong for providing samples of **L1b, L1d, L1e, L1n, L1w, L1z**.

.

References

[1] (a) Noyori, R. Asymmetric Catalysis: Science and Opportunities (Nobel Lecture). *Angew. Chem. Int. Ed.* **2002**, 41, 2008–2022. (b) Carreira, E. M.; Yamamoto, H. *Comprehensive Chirality*, 1st ed.; Oxford: Elsevier Science, 2012. (c) *Asymmetric Organocatalysis*, 1st ed.; List, B., Maruoka, K., Eds.; In *Science of Synthesis*; Thieme Chemistry, 2012. (d) *Stereoselective Synthesis*, 1st ed.; de Vries, J. G., Molander, G. A., Evans, A. P., Eds.; In *Science of Synthesis*; Thieme Chemistry, 2011.

[2] (a) Zahrt, A. F.; Athavale, S. V.; Denmark, S. E. Quantitative Structure–Selectivity Relationships in Enantioselective Catalysis: Past, Present, and Future. *Chem. Rev.* **2020**, 120, 1620–1689. (b) Sigman, M. S.; Harper, K. C.; Bess, E. N.; Milo, A. The Development of Multidimensional Analysis Tools for Asymmetric Catalysis and Beyond. *Acc. Chem. Res.* **2016**, 49, 1292–1301. (c) 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. (d) Taylor, C. J.; Pomberger, A.; Felton, K. C.; Grainger, R.; Barecka, M.; Chamberlain, T. W.; Bourne, R. A.; Johnson, C. N.; Lapkin, A. A. A Brief Introduction to Chemical Reaction Optimization. *Chem. Rev.* **2023**, 123, 3089–3126.

[3] (a) Isbrandt, E. S.; Sullivan, R. J.; Newman, S. G. High Throughput Strategies for the Discovery and Optimization of Catalytic Reactions. *Angew. Chem. Int. Ed.* **2019**, 58, 7180–7191. (b) Mennen, S. M.; Alhambra, C.; Allen, C. L.; Barberis, M.; Berritt, S.; Brandt, T. A.; Campbell, A. D.; Castañón, J.; Cherney, A. H.; Christensen, M.; Damon, D. B.; Eugenio De Diego, J.; García-Cerrada, S.; García-Losada, P.; Haro, R.; Janey, J.; Leitch, D. C.; Li, L.; Liu, F.; Lobben, P. C.; MacMillan, D. W. C.; Magano, J.; McInturff, E.; Monfette, S.; Post,

R. J.; Schultz, D.; Sitter, B. J.; Stevens, J. M.; Strambeanu, I. I.; Twilton, J.; Wang, K.; Zajac, M. A. The Evolution of High-Throughput Experimentation in Pharmaceutical Development and Perspectives on the Future. *Org. Process Res. Dev.* **2019**, 23, 1213–1242. (c) Krska, S. W.; DiRocco, D. A.; Dreher, S. D.; Shevlin, M. The Evolution of Chemical High-Throughput Experimentation To Address Challenging Problems in Pharmaceutical Synthesis. *Acc. Chem. Res.* **2017**, 50, 2976–2985. (d) Collins, K. D.; Gensch, T.; Glorius, F. Contemporary Screening Approaches to Reaction Discovery and Development. *Nat. Chem.* **2014**, 6, 859–871.

[4] (a) Adlington, N. K.; Agnew, L. R.; Campbell, A. D.; Cox, R. J.; Dobson, A.; Barrat, C. F.; Gall, M. A. Y.; Hicks, W.; Howell, G. P.; Jawor-Baczynska, A.; Miller-Potucka, L.; Pilling, M.; Shepherd, K.; Tassone, R.; Taylor, B. A.; Williams, A. Process Design and Optimization in the Pharmaceutical Industry: A Suzuki–Miyaura Procedure for the Synthesis of Savolitinib. *J. Org. Chem.* **2019**, 84, 4735–4747. (b) Perera, D.; Tucker, J. W.; Brahmbhatt, S.; Helal, C. J.; Chong, A.; Farrell, W.; Richardson, P.; Sach, N. W. A Platform for Automated Nanomole-Scale Reaction Screening and Micromole-Scale Synthesis in Flow. *Science* **2018**, 359, 429–434.

[5] Cook, A.; Clément, R.; Newman, S. G. Reaction Screening in Multiwell Plates: High-Throughput Optimization of a Buchwald–Hartwig Amination. *Nat. Protoc.* **2021**, 16, 1152–1169.

[6] (a) Lovering, F., Bikker, J., Humblet, C. Escape from flatland: Increasing saturation as an approach to improving clinical success. *J. Med. Chem.* **2009**, 52, 6752-6756. (b) Meanwell, N. A. Improving drug candidates by design: A focus on physicochemical properties as a means of improving compound disposition and safety. *Chem. Res. Toxicol.* **2011**, 24, 1420–

1456. (c) Lovering, F. Escape from Flatland 2: Complexity and Promiscuity. *Med. Chem. Commun.* **2013**, 4, 515.

[7] (a) Walsh, P. J.; Li, H.; De Parrodi, C. A. A Green Chemistry Approach to Asymmetric Catalysis: Solvent-Free and Highly Concentrated Reactions. *Chem. Rev.* **2007**, 107, 2503–2545. (b) Dunn, P. J. The Importance of Green Chemistry in Process Research and Development. *Chem. Soc. Rev.* **2012**, 41, 1452–1461.

[8] Renom-Carrasco, M.; Lefort, L. Ligand Libraries for High Throughput Screening of Homogeneous Catalysts. *Chem. Soc. Rev.* **2018**, 47, 5038–5060.

[9] Friedfeld, M. R.; Shevlin, M.; Hoyt, J. M.; Krska, S. W.; Tudge, M. T.; Chirik, P. J. Cobalt Precursors for High-Throughput Discovery of Base Metal Asymmetric Alkene Hydrogenation Catalysts. *Science* **2013**, 342, 1076–1080.

[10] Shevlin, M.; Friedfeld, M. R.; Sheng, H.; Pierson, N. A.; Hoyt, J. M.; Campeau, L.-C.; Chirik, P. J. Nickel-Catalyzed Asymmetric Alkene Hydrogenation of α,β-Unsaturated Esters: High-Throughput Experimentation-Enabled Reaction Discovery, Optimization, and Mechanistic Elucidation. *J. Am. Chem. Soc.* **2016**, 138, 3562–3569.

[11] Chang, M.; Huang, Y.; Liu, S.; Chen, Y.; Krska, S. W.; Davies, I. W.; Zhang, X. Asymmetric Hydrogenation of Pyridinium Salts with an Iridium Phosphole Catalyst. *Angew. Chem. Int. Ed.* **2014**, 53, 12761–12764.

[12] (a) Renom-Carrasco, M.; Gajewski, P.; Pignataro, L.; de Vries, J. G.; Piarulli, U.; Gennari, C.; Lefort, L. Asymmetric Hydrogenation of 3-Substituted Pyridinium Salts. *Chem. Eur. J.* **2016**, 22, 9528–9532.

[13] Li, H.; Belyk, K. M.; Yin, J.; Chen, Q.; Hyde, A.; Ji, Y.; Oliver, S.; Tudge, M. T.; Campeau, L.-C.; Campos, K. R. Enantioselective Synthesis of Hemiaminals via Pd-Catalyzed C–N

Coupling with Chiral Bisphosphine Mono-Oxides. *J. Am. Chem. Soc.* **2015**, 137, 13728–13731.

[14] Boyer, S. H.; Gonzalez-de-Castro, A.; Dielemans, J. A. H.; Lefort, L.; Zhu, Z.; Gnahn, M.; Schörghuber, J.; Steinhofer, S.; de Vries, A. H. M.; Hecker, S. J. Scalable Synthesis of β-Lactamase Inhibitor QPX7728 by Sequential Nickel-Catalyzed Boron Insertion into a Benzofuran Substrate and Enantioselective Cyclopropanation of the Resulting Vinylboronate. *Org. Process Res. Dev.* **2022**, 26, 925–935.

[15] Jimenez-Gonzalez, C.; Ponder, C. S.; Broxterman, Q. B.; Manley, J. B. Using the Right Green Yardstick: Why Process Mass Intensity Is Used in the Pharmaceutical Industry To Drive More Sustainable Processes. *Org. Process Res. Dev.* **2011**, 15, 912–917.

[16] (a) Wagen, C. C.; McMinn, S. E.; Kwan, E. E.; Jacobsen, E. N. Screening for Generality in Asymmetric Catalysis. *Nature* **2022**, 610, 680–686. (b) Prieto Kullmer, C. N.; Kautzky, J. A.; Krska, S. W.; Nowak, T.; Dreher, S. D.; MacMillan, D. W. C. Accelerating Reaction Generality and Mechanistic Insight through Additive Mapping. *Science* **2022**, 376, 532–539. (c) Rose, B. T.; Timmerman, J. C.; Bawel, S. A.; Chin, S.; Zhang, H.; Denmark, S. E. High-Level Data Fusion Enables the Chemoinformatically Guided Discovery of Chiral Disulfonimide Catalysts for Atropselective Iodination of 2-Amino-6-Arylpyridines. *J. Am. Chem. Soc.* **2022**, 144, 22950–22964.

[17] Shevlin, M. Practical High-Throughput Experimentation for Chemists. *ACS Med. Chem. Lett.* **2017**, 8, 601–607.

[18] (a) Mead, R. *The Design of Experiments: Statistical Principles for Practical Applications*; Cambridge University Press, 1990. (b) Murray, P. M.; Bellany, F.; Benhamou, L.; Bučar, D.-K.; Tabor, A. B.; Sheppard, T. D. The Application of Design of Experiments (DoE)

Reaction Optimisation and Solvent Selection in the Development of New Synthetic Chemistry. *Org. Biomol. Chem.* **2016**, 14, 2373–2384. (c) Weissman, S. A.; Anderson, N. G. Design of Experiments (DoE) and Process Optimization. A Review of Recent Publications. *Org. Process Res. Dev.* **2015**, 19, 1605–1633.

[19] Durand, D. J.; Fey, N. Computational Ligand Descriptors for Catalyst Design. *Chem. Rev.* **2019**, 119, 6561–6594.

[20] For examples of DoE in catalyst optimization see: (a) Harper, K. C.; Sigman, M. S. Predicting and Optimizing Asymmetric Catalyst Performance Using the Principles of Experimental Design and Steric Parameters. *Proc. Natl. Acad. Sci U.S.A.* **2011**, 108, 2179–2183. (b) Murray, P. M.; Tyler, S. N. G.; Moseley, J. D. Beyond the Numbers: Charting Chemical Reaction Space. *Org. Process Res. Dev.* **2013**, 17, 40–46.

[21] Reid, J. P.; Sigman, M. S. Holistic Prediction of Enantioselectivity in Asymmetric Catalysis. *Nature* **2019**, 571, 343–348.

[22] Harper, K. C.; Sigman, M. S. Using Physical Organic Parameters To Correlate Asymmetric Catalyst Performance. *J. Org. Chem.* **2013**, 78, 2813–2818.

[23] Tolman, C. A. Phosphorus Ligand Exchange Equilibriums on Zerovalent Nickel. Dominant Role for Steric Effects. *J. Am. Chem. Soc.* **1970**, 92, 2956–2965.

[24] (a) Jover, J.; Fey, N.; Harvey, J. N.; Lloyd-Jones, G. C.; Orpen, A. G.; Owen-Smith, G. J. J.; Murray, P.; Hose, D. R. J.; Osborne, R.; Purdie, M. Expansion of the Ligand Knowledge Base for Chelating P,P-Donor Ligands (LKB-PP). *Organometallics* **2012**, 31, 5302–5306. (b) Jover, J.; Fey, N.; Harvey, J. N.; Lloyd-Jones, G. C.; Orpen, A. G.; Owen-Smith, G. J. J.; Murray, P.; Hose, D. R. J.; Osborne, R.; Purdie, M. Expansion of the Ligand Knowledge Base for Monodentate P-Donor Ligands (LKB-P). *Organometallics* **2010**, 29, 6245–6258.

[25] 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.; Aspuru-Guzik, A. A Comprehensive Discovery Platform for Organophosphorus Ligands for Catalysis. *J. Am. Chem. Soc.* **2022**, 144, 1205–1217.

[26] 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, 7773–7780.

[27] Dotson, J. J.; Van Dijk, L.; Timmerman, J. C.; Grosslight, S.; Walroth, R. C.; Gosselin, F.; Püntener, K.; Mack, K. A.; Sigman, M. S. Data-Driven Multi-Objective Optimization Tactics for Catalytic Asymmetric Reactions Using Bisphosphine Ligands. *J. Am. Chem. Soc.* **2023**, 145, 110–121.

[28] Henle, J. J.; Zahrt, A. F.; Rose, B. T.; Darrow, W. T.; Wang, Y.; Denmark, S. E. Development of a Computer-Guided Workflow for Catalyst Optimization. Descriptor Validation, Subset Selection, and Training Set Analysis. *J. Am. Chem. Soc.* **2020**, 142, 11578–11592.

[29] (a) Denmark, S. E.; Nakajima, N.; Nicaise, O. J.-C.; Faucher, A.-M.; Edwards, J. P. Preparation of Chiral Bisoxazolines: Observations on the Effect of Substituents. *J. Org. Chem.* **1995**, 60, 4884–4892. (b) Denmark, S. E.; Nakajima, N.; Stiff, C. M.; Nicaise, O. J.-C.; Kranz, M. Studies on the Bisoxazoline- and (−)-Sparteine-Mediated Enantioselective Addition of Organolithium Reagents to Imines. *Adv. Synth. Catal.* **2008**, 350, 1023–1045. (c) Denmark, S. E.; Stiff, C. M. Effect of Ligand Structure in the Bisoxazoline Mediated Asymmetric Addition of Methyllithium to Imines. *J. Org. Chem.* **2000**, 65, 5875–5878.

[30] 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, eaau5631.

[31] Jarrold, J.; Davies, C. C. PRMTs and Arginine Methylation: Cancer's Best-Kept Secret? *Trends Mol. Med.* **2019**, 25, 993–1009.

[32] Shilo, K.; Wu, X.; Sharma, S.; Welliver, M.; Duan, W.; Villalona-Calero, M.; Fukuoka, J.; Sif, S.; Baiocchi, R.; Hitchcock, C. L.; Zhao, W.; Otterson, G. A. Cellular Localization of Protein Arginine Methyltransferase-5 Correlates with Grade of Lung Tumors. *Diagn. Pathol.* **2013**, 8, 201.

[33] Wu, Y.; Wang, Z.; Zhang, J.; Ling, R. Elevated Expression of Protein Arginine Methyltransferase 5 Predicts the Poor Prognosis of Breast Cancer. *Tumour. Biol.* **2017**, 39, 1–11.

[34] Jeon, J.-Y.; Lee, J. S.; Park, E.-R.; Shen, Y. N.; Kim, M.-Y.; Shin, H.-J.; Joo, H.-Y.; Cho, E.-H.; Moon, S. M.; Shin, U. S.; Park, S. H.; Han, C. J.; Choi, D. W.; Gu, M. B.; Kim, S.-B.; Lee, K.-H. Protein Arginine Methyltransferase 5 Is Implicated in the Aggressiveness of Human Hepatocellular Carcinoma and Controls the Invasive Activity of Cancer Cells. *Oncol. Rep*. **2018**, 40, 536–544.

[35] (a) Jing, P.; Zhao, N.; Ye, M.; Zhang, Y.; Zhang, Z.; Sun, J.; Wang, Z.; Zhang, J.; Gu, Z. Protein Arginine Methyltransferase 5 Promotes Lung Cancer Metastasis via the Epigenetic Regulation of MiR-99 Family/FGFR3 Signaling. *Cancer Lett.* **2018**, 427, 38–48. (b) Liu, X.; Zhang, J.; Liu, L.; Jiang, Y.; Ji, J.; Yan, R.; Zhu, Z.; Yu, Y. Protein Arginine Methyltransferase 5-Mediated Epigenetic Silencing of IRX1 Contributes to Tumorigenicity

and Metastasis of Gastric Cancer. *Biochim. Biophys. Acta. Mol. Basis Dis.* **2018**, 1864, 2835–2844.

[36] Shimizu, D.; Kanda, M.; Sugimoto, H.; Shibata, M.; Tanaka, H.; Takami, H.; Iwata, N.; Hayashi, M.; Tanaka, C.; Kobayashi, D.; Yamada, S.; Nakayama, G.; Koike, M.; Fujiwara, M.; Fujii, T.; Kodera, Y. The Protein Arginine Methyltransferase 5 Promotes Malignant Phenotype of Hepatocellular Carcinoma Cells and Is Associated with Adverse Patient Outcomes after Curative Hepatectomy. *Int. J. Oncol.* **2017**, 50, 381–386.

[37] Feustel, K.; Falchook, G. S. Protein Arginine Methyltransferase 5 (PRMT5) Inhibitors in Oncology Clinical Trials: A Review. *J. Immunother. Precis. Oncol.* **2022**, 5, 58–67.

[38] (a) Zafra-Polo, M. C.; Figadere, B.; Gallardo, T.; Tormo, J. R.; Cortes, D. Natural Acetogenins from Annonaceae, Synthesis and Mechanisms of Action. *Phytochemistry* **1998**, 48, 1087. (b) Curti, C.; Brindani, N.; Battistini, L.; Sartori, A.; Pelosi, G.; Mena, P.; Brighenti, F.; Zanardi, F.; Del Rio, D. Catalytic, Enantioselective Vinylogous Mukaiyama Aldol Reaction of Furan-Based Dienoxy Silanes: A Chemodivergent Approach to γ-Valerolactone Flavan-3-ol Metabolites and δ-Lactone Analogues. *Adv. Synth. Catal.* **2015**, 357, 4082–4092. (c) Evans, D. A.; Kværnø, L.; Dunn, T. B.; Beauchemin, A.; Raymer, B.; Mulder, J. A.; Olhava, E. J.; Juhl, M.; Kagechika, K.; Favor, D. A. Total Synthesis of (+)-Azaspiracid-1. An Exhibition of the Intricacies of Complex Molecule Synthesis. *J. Am. Chem. Soc.* **2008**, 130, 16295–16309. (d) de March, P.; Figueredo, M.; Font, J.; Raya, J.; Alvarez-Larena, A.; Piniella, J. F. C 2 -Symmetric Enantiopure Ethanotethered Bis(α,β-Butenolides) as Templates for Asymmetric Synthesis. Application to the Synthesis of (+)-Grandisol. *J. Org. Chem.* **2003**, 68, 2437–2447.

[39] Li, X.; Li, X.-Q.; Liu, H.-M.; Zhou, X.-Z.; Shao, Z.-H. Synthesis and Evaluation of Antitumor Activities of Novel Chiral 1,2,4-Triazole Schiff Bases Bearing γ-Butenolide Moiety. *Org. Med. Chem. Lett.* **2012**, 2, 26.

[40] (a) Mao, B.; Fañanás-Mastral, M.; Feringa, B. L. Catalytic Asymmetric Synthesis of Butenolides and Butyrolactones. *Chem. Rev.* **2017**, 117, 10502–10566. (b) Casiraghi, G.; Battistini, L.; Curti, C.; Rassu, G.; Zanardi, F. The Vinylogous Aldol and Related Addition Reactions: Ten Years of Progress. *Chem. Rev.* **2011**, 111, 3076–3154. (c) Denmark, S. E.; Heemstra, J. R. Jr.; Beutner, G. L. Catalytic, Enantioselective, Vinylogous Aldol Reactions *Angew. Chem., Int. Ed. Engl.* **2005**, *44*, 4682-4698.

[41] (a) Brown, S. P.; Goodwin, N. C.; MacMillan, D. W. C. The First Enantioselective Organocatalytic Mukaiyama−Michael Reaction: A Direct Method for the Synthesis of Enantioenriched γ-Butenolide Architecture. *J. Am. Chem. Soc.* **2003**, 125, 1192–1194. (b) Huang, Y.; Walji, A. M.; Larsen, C. H.; MacMillan, D. W. C. Enantioselective Organo-Cascade Catalysis. *J. Am. Chem. Soc.* **2005**, 127, 15051–15053. (c) Pansare, S. V.; Paul, E. K. The Organocatalytic Vinylogous Aldol Reaction: Recent Advances. *Chem. Eur J.* **2011**, 17, 8770–8779.

[42] (a) Evans, D. A.; Kozlowski, M. C.; Murry, J. A.; Burgey, C. S.; Campos, K. R.; Connell, B. T.; Staples, R. J. C 2 -Symmetric Copper(II) Complexes as Chiral Lewis Acids. Scope and Mechanism of Catalytic Enantioselective Aldol Additions of Enolsilanes to (Benzyloxy)Acetaldehyde. *J. Am. Chem. Soc.* **1999**, 121, 669–685. (b) Sedelmeier, J.; Hammerer, T.; Bolm, C. C 1 -Symmetric Oxazolinyl Sulfoximines as Ligands in Copper-Catalyzed Asymmetric Mukaiyama Aldol Reactions. *Org. Lett.* **2008**, 10, 917–920.

[43 Onitsuka, S.; Matsuoka, Y.; Irie, R.; Katsuki, T. Highly Enantioselective Cr(salen)-catalyzed Reaction of 2-(Trimethylsilyloxy)furan and Aldehydes. Effect of Alcohol on Enantioselectivity. *Chem. Lett.* **2003**, 32, 974.

[44] Szlosek, M.; Figadère, B. Highly Enantioselective Aldol Reaction with 2-Trimethylsilyloxyfuran: The First Catalytic Asymmetric Autoinductive Aldol Reaction. *Angew. Chem. Int. Ed.* **2000**, 39, 1799–1801.

[45] (a) Desimoni, G.; Faita, G.; Jørgensen, K. A. Update 1 of: C 2 -Symmetric Chiral Bis(Oxazoline) Ligands in Asymmetric Catalysis. *Chem. Rev.* **2011**, 111, PR284–PR437. (b) Connon, R.; Roche, B.; Rokade, B. V.; Guiry, P. J. Further Developments and Applications of Oxazoline-Containing Ligands in Asymmetric Catalysis. *Chem. Rev.* **2021**, 121, 6373–6521.

[46] Lipkowitz, K. B.; Pradhan, M. Computational Studies of Chiral Catalysts: A Comparative Molecular Field Analysis of an Asymmetric Diels−Alder Reaction with Catalysts Containing Bisoxazoline or Phosphinooxazoline Ligands. *J. Org. Chem.* **2003**, 68, 4648–4656.

[47] (a) Lipkowitz, K. B.; Schefzick, S.; Avnir, D. Enhancement of Enantiomeric Excess by Ligand Distortion. *J. Am. Chem. Soc.* **2001**, 123, 6710–6711. (b) (1) Alvarez, S.; Alemany, P.; Avnir, D. Continuous Chirality Measures in Transition Metal Chemistry. *Chem. Soc. Rev.* **2005**, 34, 313.

[48] Werth, J.; Sigman, M. S. Linear Regression Model Development for Analysis of Asymmetric Copper-Bisoxazoline Catalysis. *ACS Catal.* **2021**, 11, 3916–3922.

[49] Machacek, M.; Witter, D.; Huang, C.; Kawamura, S.; Schneider, S.; Wan, M. Prmt5 Inhibitors. U.S Patent WO 2020033285, February 13, 2020.

[50] Shrestha, B.; Rose, B. T.; Olen, C. L.; Roth, A.; Kwong, A. C.; Wang, Y.; Denmark, S. E. A Unified Strategy for the Asymmetric Synthesis of Highly Substituted 1,2-Amino Alcohols Leading to Highly Substituted Bisoxazoline Ligands. *J. Org. Chem.* **2021**, 86, 3490–3534.

[51] Stephens, M. A. EDF Statistics for Goodness of Fit and Some Comparisons. *J. Am. Stat. Assoc.* **1974**, 69, 730–737.

[52] Muratov, E. N.; Bajorath, J.; Sheridan, R. P.; Tetko, I. V.; Filimonov, D.; Poroikov, V.; Oprea, T. I.; Baskin, I. I.; Varnek, A.; Roitberg, A.; Isayev, O.; Curtalolo, S.; Fourches, D.; Cohen, Y.; Aspuru-Guzik, A.; Winkler, D. A.; Agrafiotis, D.; Cherkasov, A.; Tropsha, A. QSAR without Borders. *Chem. Soc. Rev.* **2020**, 49, 3525–3564.