Molecular Field Analysis Using Computational-Screening Data in Asymmetric *N*-Heterocyclic Carbene-Copper Catalysis toward Data-driven *in silico* Catalyst Optimization

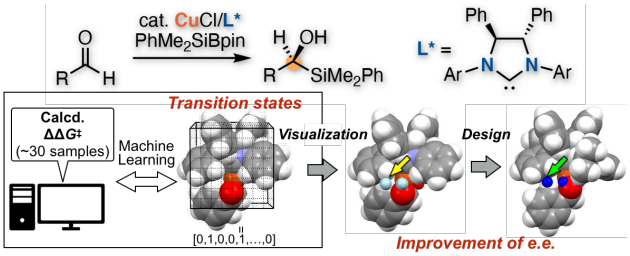
Masakiyo Mukai†, Kazunori Nagao†, Shigeru Yamaguchi‡,*, and Hirohisa Ohmiya†,§,*

†Division of Pharmaceutical Sciences, Graduate School of Medical Sciences, Kanazawa University, Kakuma-machi, Kanazawa 920-1192, Japan.

‡RIKEN Center for Sustainable Resource Science, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan.

§JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan.

Supporting Information Placeholder



ABSTRACT: A molecular-field-based regression analysis using computational screening data for *N*-heterocyclic carbene (NHC)-Cu-catalyzed asymmetric carbonyl additions of a silylboronate to aldehydes is reported. A computational screening was performed to collect enantioselectivity data ($\Delta\Delta G^{\ddagger}$: energy differences between the transitions states leading to each enantiomer) via transition-state (TS) calculations using density functional theory (DFT) methods. A molecular field analysis (MFA) was carried out using the obtained calculated $\Delta\Delta G^{\ddagger}$ values and TS structures (30 samples in total). Important structural information for enantioselectivity extracted by the MFA was visualized on the TS structures, which provided insight into an asymmetric induction mechanism. Based on the obtained information, chiral NHC ligands were designed, which showed improved enantioselectivity in these carbonyl additions.

Data-driven machine-learning-based optimization of molecular catalysis is an emerging and promising research area.[1] Regression analysis, i.e., quantitative structure–property relationships (QSPR) modeling of catalytic reactions typically correlates molecular descriptors with *experimental catalytic activity values*. High quality-experimental data suitable for QSPR modeling are, however, not always available. In some cases, experimental data include non-negligible noise derived from various factors, such as side reactions and experimental errors. The use of such data in regression analysis/QSPR can reduce its predictive performance, and in some cases, the information extracted by the analysis may not provide reliable insights into the reactions. The use of computational-screening data obtained through *ab-initio* calculations avoids this potential shortcoming, as the calculated activity data do not include the aforementioned noise. In addition, computational-screening approaches do not require experiments to collect data, thereby enabling data analysis of reactions that involve expensive and synthetically difficult catalysts. Despite these attractive features, reports of data-driven molecular design in molecular catalysis using the QSPR framework with the calculated catalytic-activity values as the target variables are scarce,[2] although virtual screening using transition-state force fields developed via the quantum-guided molecular-mechanics method has been investigated in asymmetric catalysis.[1e,1f,3] As transition-state (TS) DFT calculations with appropriate calculation conditions have become a reliable method to design asymmetric catalytic reactions,[4,5] the use of the calculated selectivity values as target variables for QSPR or quantitative structure-selectivity relationships (QSSR) can be expected to further accelerate the development of asymmetric catalytic reactions. Recently, a QSSR study of the asymmetric propargylation of aryl aldehydes has been reported,[6] in which more than 600 training samples with literature-reported calculated activation-energy values[5] were used. However, the calculation cost of DFT-based TS modeling is high. Therefore, it would be highly desirable to develop a regression-based method that enables molecular design to improve enantioselectivity by analyzing a relatively small set of DFT-based computational-screening data. Here, we report the data-driven design of chiral catalysts for *N*-heterocyclic carbene (NHC)-Cu-catalyzed asymmetric carbonyl additions of a silylboronate to aldehydes through molecular-field-based regression analysis[7,8] using the calculated $\Delta\Delta G^{\ddagger}$ values and transition-state structures with a total of 30 training samples (Figure 1a).

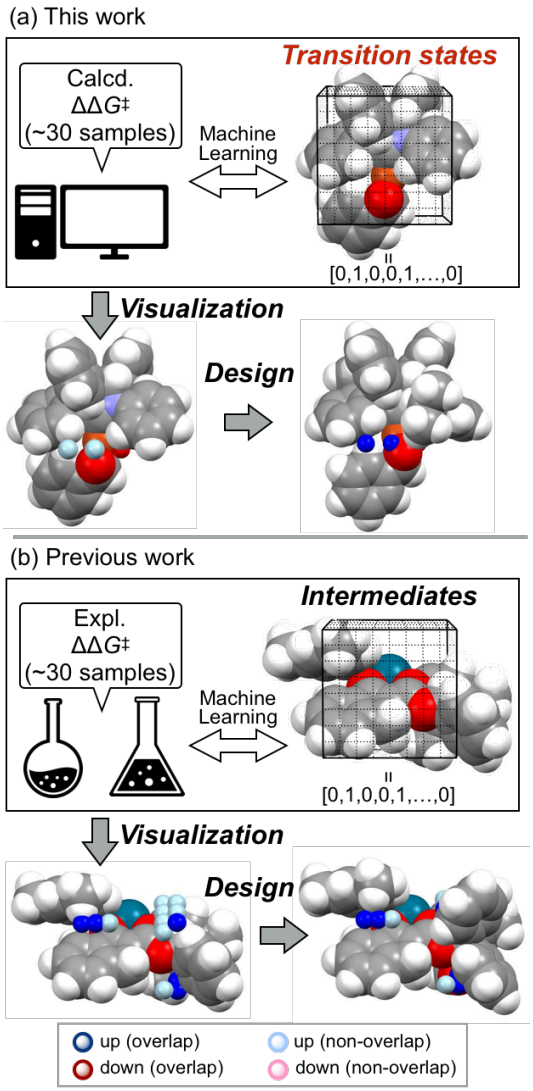
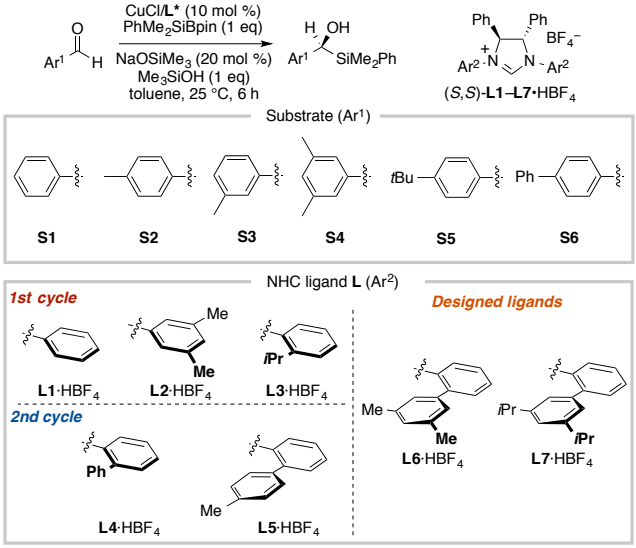


Figure 1. Molecular field analysis (MFA) in asymmetric catalysis and molecular design based on the visualized structural information (introducing substituents that overlap at the light blue points improves the enantioselectivity). (a) This work: MFA using transition-state structures and calculated $\Delta\Delta G^{\ddagger}$ values. (b) Previous work: MFA using intermediate structures and experimental $\Delta\Delta G^{\ddagger}$ values. Blue points correspond to positive regression coefficients where molecular structures overlap [up (overlap)]. Light blue points correspond to positive regression coefficients where molecular structures do not overlap and if substituents are introduced there, selectivity will increase [up (non-overlap)]. Red points correspond to negative regression coefficients where molecular structures overlap [down (overlap)]. Light red points correspond to negative regression coefficients where molecular structures do not overlap and if substituents are introduced there, selectivity will decrease [down (non-overlap)].

For the initial training dataset, we employed three ring-saturated C_2-symmetric NHC ligands (**L1**–**3**), which have shown low to moderate enantioselectivity in a previous investigation (Table 1).[7b] Moreover, six aromatic aldehydes (**S1**–**6**) were selected. We calculated the TSs of a possible enantio-determining step for all combinations of the three catalyst and six substrate structures. We hypothesized that the reaction proceeded via the four-centered reaction mechanism that is frequently observed in related Cu(I)-catalyzed reactions.[8,9] As an example, the TS structure derived from ligand **L3** and substrate **S1** (**L3S1**) is shown in Figure 2. The calculated $\Delta\Delta G^{\ddagger}$ values exhibited a correlation with the experimental $\Delta\Delta G^{\ddagger}$ values (see Figure S9). For data analysis, we employed a 3D-QSSR approach,[1c] which we refer to here as molecular field analysis (MFA). MFA is a regression analysis between reaction outcomes and molecular fields calculated from three-dimensional molecular structures. In our previous study, we found that MFA of the intermediates in the enantio-determining step of asymmetric catalytic reactions enabled the extraction and visualization of structural information that led to the design of molecules with improved enantioselectivity (Figure 1b).[10,11] Thus, we were interested in whether MFA between the calculated $\Delta\Delta G^{\ddagger}$ values and the TS structures would be able to extract similar information (Figure 1a). The seminal report of the use of MFA in asymmetric catalysis by the Kozlowski group employed TS structures; however, the authors did not use the extracted and visualized information for molecular design, and their target variables were the experimental $\Delta\Delta G^{\ddagger}$ values.[12] Our specific goal in this study was to evaluate our approach through the data-driven design of the current optimum NHC ligand **L6** in the targeted carbonyl additions[7] starting from a set of 18 training samples (Table 1).

Table 1. Enantioselectivity of the training samples (%ee).



	L1	L2	L3	L4	L5
S1	21 (0.34)	31 (0.58)	73 (1.77)	87 (1.83)	82 (2.25)
S2	18 (0.59)	41 (0.65)	17 (1.67)	85 (2.29)	85 (2.32)
S3	15 (0.60)	38 (0.71)	63 (1.33)	80 (2.31)	78 (2.81)
S4	11 (0.04)	18 (0.70)	66 (1.82)	78 (2.43)	74 (2.41)
S5	19 (0.16)	27 (0.30)	69 (0.85)	82 (2.18)	81 (2.76)
S6	17 (0.38)	31 (–0.25)	63 (1.28)	88 (2.17)	78 (2.30)

Calculated $\Delta\Delta G^{\ddagger}$ values (kcal/mol) are shown in parentheses. The 18 initial training samples are highlighted in grey.

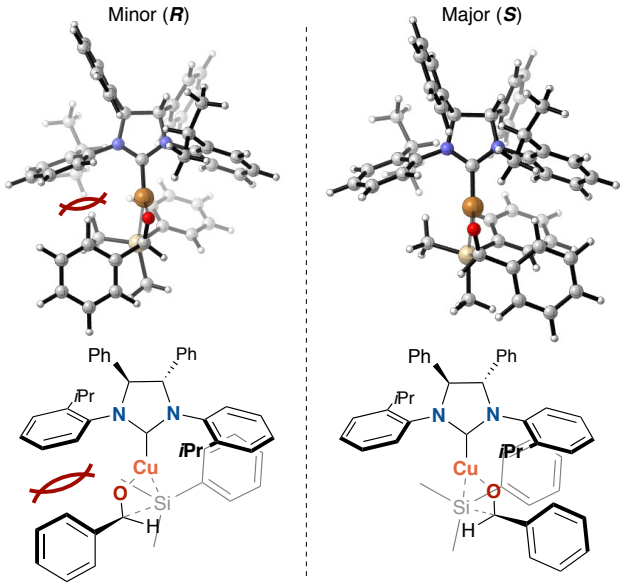


Figure 2. Transition-state structures of **L3S1**; calculations were performed at the ωB97XD/SDD(Cu) & 6-311+G**-SMD(toluene)//B3LYP-D3/LANL2DZ(Cu) & 6-31G* level of theory.

As molecular fields, we employed steric indicator fields composed of indicator variables (with values of 0 or 1), as shown in Figure 1. The indicator fields were calculated for each unit cell of a grid space. A unit cell that includes the van der Waals radius of any atom was assigned a value of 1, otherwise, the cell was given a value of 0 (Figures 1 and S2). The indicator fields were calculated from the TS structures of a minor pathway (the *R* pathway), as we aimed to obtain guidelines for the design of chiral NHC ligands that would destabilize the minor pathway. By correlating the calculated $\Delta\Delta G^{\ddagger}$ values and the indicator fields using LASSO[13] or Elastic Net[14] regression according to the reported procedure[15] employing the R package glmnet[16], we generated a regression model (for more details regarding the MFA, see the SI section 'Details of the molecular field analysis'). The important structural information obtained from the regression coefficients is visualized in the TS structures of **L1S1** and **L3S1** (Figure 3a and 3b). The blue and red points correspond to positive and negative coefficients (for regression coefficient values with coordinates of unit cells, see Table S5). If molecular structures are located on the blue or red points, the enantioselectivity increases or decreases, respectively. We focused on the blue point observed above the phenyl group of the aldehyde. While the phenyl group on the NHC nitrogen atom of the TS structure **L1S1** did not overlap with the blue point, as indicated by the yellow arrow (Figure 3b), the corresponding phenylene moiety of **L3S1** did overlap with the blue point, as indicated by the green arrow (Figure 3a). The difference in the position of the phenylene moiety occurs due to steric repulsion between the *i*Pr group of the NHC ligand and the silyl group of **L3S1** (for details, see Figure S8), which induces a steric interaction between the phenylene moiety of the NHC ligand and the phenyl group of the aldehyde and destabilizes the *R* pathway (Figure 2). On the other hand, this steric effect between the ligand and aldehyde does not exist in the *S* pathway (Figure 2). Inspired by this mechanistic insight, we introduced a phenyl or *p*-tolyl group in the place of the *i*Pr group of NHC ligand **L3** to further destabilize the *R* pathway by increasing the steric repulsion. The transition-state structures **L4S1** and **L5S1** overlap at the blue point described above (Figures 3c and S4). The calculated $\Delta\Delta G^{\ddagger}$ values for **L4S1** (1.8 kcal/mol) and **L5S1** (2.2 kcal/mol) are higher than that of the design template **L1S1** (0.3 kcal/mol). We experimentally confirmed that the enantioselectivity of the reactions **L4S1** (87% ee) and **L5S1** (82% ee) are improved relative to the selectivity of **L3S1** (73% ee), which exhibited the highest value among the 18 training samples (for more details of the molecular design workflow, see the SI, page S4). Thus, we successfully improved the enantioselectivity through data-driven catalyst design using computational-screening data.

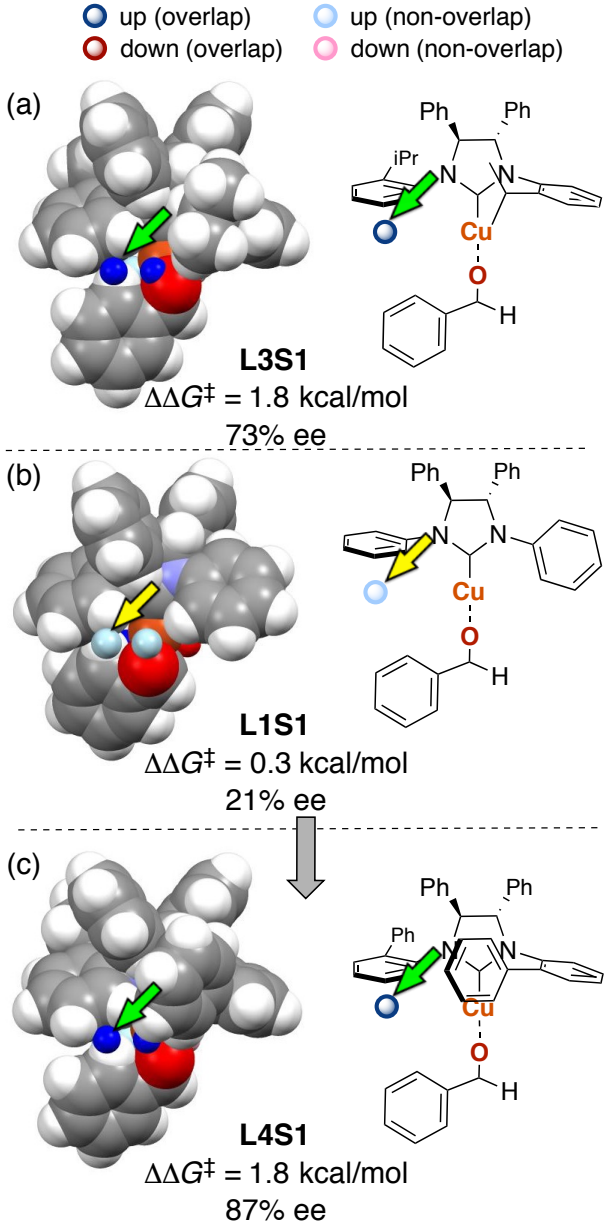


Figure 3. Important structural information visualized in the TS structures. The information extracted by the MFA using 18 samples is shown, together with the TS structures (a) **L3S1**, (b) **L1S1**, and (c) **L4S1**. Calculated $\Delta\Delta G^{\ddagger}$ and experimentally determined enantiomeric excess values are also shown. The silyl group was removed for the MFA (for more details, see Figure S2), and thus, we have omitted the silyl group in Figures 3 and 4.

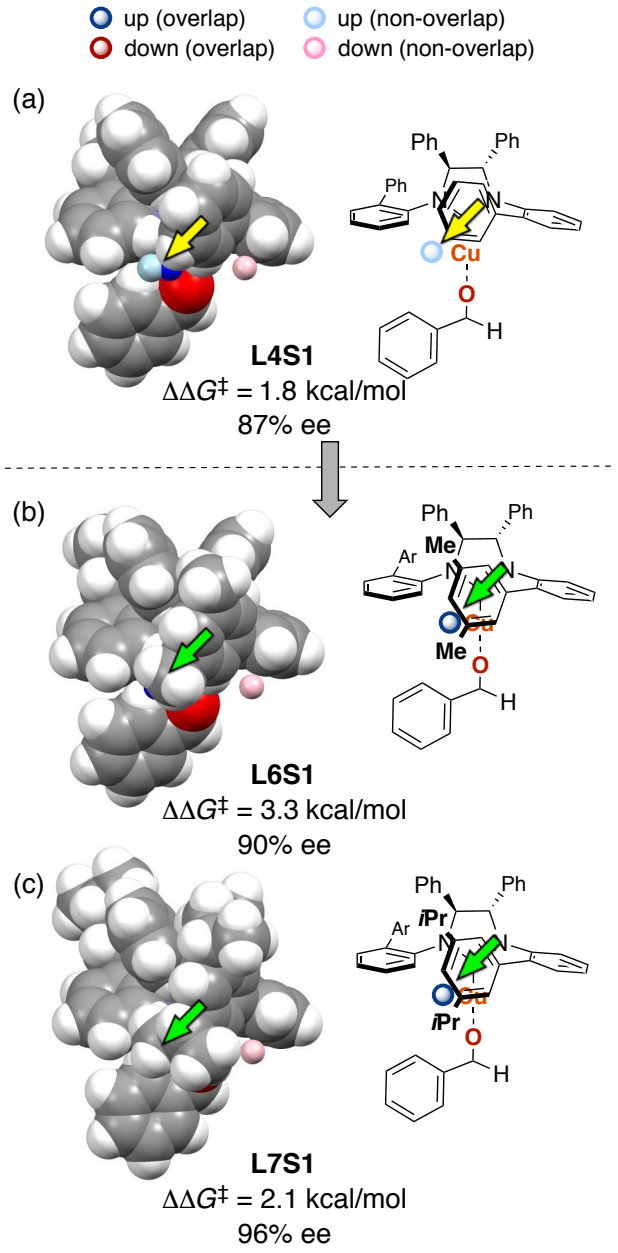


Figure 4. Important structural information visualized in the TS structures. The information extracted by the MFA using 30 samples is shown together with the TS structures (a) **L4S1**, (b) **L6S1**, and (c) **L7S1**. Calculated $\Delta\Delta G^{\ddagger}$ and experimentally determined enantiomeric-excess values are also shown.

We then performed the MFA again, this time including the data calculated from the designed ligands (30 samples, Table 1). Figure 4a shows the structural information in the TS structure **L4S1**, in which a blue point that did not overlap with the structure was identified near the 3-position of the aromatic group of the NHC ligand, as indicated by the yellow arrow. Based on this information, we designed **L6** and **L7**, which bear Me or *i*Pr groups at the 3- and 5-positions of the aromatic ring. The structures **L6S1** and **L7S1** overlapped with the structural information used for the design (Figures 4b, 4c and Figure S5). The calculated $\Delta\Delta G^{\ddagger}$ values are higher than that of the design template **L4S1** (**L4S1**: 1.8 kcal/mol; **L6S1**: 3.3 kcal/mol; **L7S1**: 2.1 kcal/mol). We then performed the reactions **L6S1** and **L7S1**. To our delight, further improvement of the enantioselectivity was observed (**L6S1**: 90% ee; **L7S1**: 96% ee; **L4S6**: 88% ee, which is the highest value among the 30 training samples). Although DFT calculations are useful to design chiral catalysts,[4,5] DFT-based computational predictions vary vastly depending on the calculation conditions and often fail as replacements for experiments.[17] In this case, we employed the ωB97XD functional for the single-point calculations; we also confirmed that MFA using $\Delta\Delta G^{\ddagger}$ values obtained via single-point calculations using other representative functionals for the TS calculations in asymmetric catalysis led to the same molecular design as described above (Figures S10–S12). Although we employed 30 training samples to secure the quality and reliability of the regression models, practically, we can in this case successfully design appropriate molecules using a smaller sample size (for details, see Figure S13).

In summary, a molecular field analysis (MFA) using computational-screening data and the corresponding transition-state (TS) structures with a relatively small dataset (30 samples) enabled the design of chiral NHC ligands with improved enantioselectivity for the Cu-catalyzed carbonyl addition of a silylboronate. The design of ligands **L6** and **L7** was achieved through iterative MFA-based molecular design. Further optimization of asymmetric catalysis using this strategy is currently in progress. As the structural information comprises Cartesian coordinates that computers can understand, the combination of our MFA-based strategy and a structure generator could lead to the future development of efficient computational chiral catalyst optimization programs/software, as has been seen in the field of the computational design of drug-like molecules.[18]

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Experimental details, characterization data for all new compounds, details of calculations and data analysis (PDF)

Input data for the data analysis (Zip)

Output data obtained from the data analysis (Zip)

Coordinates of all the TS structures (XYZ)

AUTHOR INFORMATION

Corresponding Author

Shigeru Yamaguchi − RIKEN Center for Sustainable Resource Science, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan.; orcid.org /0000-0003-2304-8448; Email: shigeru.yamaguchi.hw@riken.jp

Hirohisa Ohmiya − Division of Pharmaceutical Sciences, Graduate School of Medical Sciences, Kanazawa University, Kanazawa 920-1192, Japan; orcid.org/0000-0002-1374-1137; Email: ohmiya@p.kanazawa-u.ac.jp

Authors

Masakiyo Mukai − Division of Pharmaceutical Sciences, Graduate School of Medical Sciences, Kanazawa University, Kanazawa 920-1192, Japan

Kazunori Nagao − Division of Pharmaceutical Sciences, Graduate School of Medical Sciences, Kanazawa University, Kanazawa 920-1192, Japan; orcid.org/0000-0003-3141- 5279

Notes

The authors declare no competing interests.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI grants JP17H06449 (H. O.) and 20H04831 (S. Y.) (Hybrid Catalysis). The DFT calculations were performed on the RIKEN HOKUSAI supercomputer system.

REFERENCES

1. (a) 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. (b) Santiago, C. B.; Guo, J.-Y.; Sigman, M. S. Predictive and Mechanistic Multivariate Linear Regression Models for Reaction Development. *Chem. Sci.* **2018**, *9*, 2398-2412. (c) 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. (d) 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. (e) Toyao, T.; Maeno, Z.; Takakusagi, S; Kamachi, T.; Takigawa, I; Shimizu, K.-i. Machine Learning for Catalysis Informatics: Recent Applications and Prospects *ACS Catal.* **2020**, *10*, 2260-2297. (f) Foscato, M.; Jensen, V. R. Automated *in Silico* Design of Homogeneous Catalysts. *ACS Catal.* **2020**, *10*, 2354-2377.
2. (a) Chu, Y.; Heyndrickx, W.; Occhipinti, G.; Jensen, V. R.; Alsberg B. K. An Evolutionary Algorithm for de Novo Optimization of Functional Transition Metal Compounds *J. Am. Chem. Soc.* **2012**, *134*, 8885–8895. (b) Maley, S. M.; Kwon, D.-H.; Rollins, N.; Stanley, J. C.; Sydora, O. L.; Bischof, S. M.; Ess, D. H. Quantum-mechanical Transition-state Model Combined with Machine Learning Provides Catalyst Design Features for Selective Cr Olefin Oligomerization. *Chem. Sci.* **2020**, *11*, 9665-9674.
3. (a) Rosales, A. R.; Quinn, T. R.; Wahlers, J.; Tomberg, A.; Zhang, X.; Helquist, P.; Wiest, O.; Norrby, P.-O. Application of Q2MM to Predictions in Stereoselective Synthesis. *Chem. Commun.* **2018**, *54*, 8294−8311. (b) Rosales, A. R.; Wahlers, J.; Limé, E.; Meadows, R. E.; Leslie, K. W.; Savin, R.; Bell, F.; Hansen, E.; Helquist, P.; Munday, R. H.; Wiest, O.; Norrby, P.-O. Rapid Virtual Screening of Enantioselective Catalysts Using CatVS. *Nat. Cat.* **2019**, *2*, 41−45.
4. (a) Houk, K. N.; Cheong, P. H.-Y. Computational Prediction of Small-Molecule Catalysts. *Nature* **2008**, *455*, 309−313. (b) Cheong, P. H.-Y.; Legault, C. Y.; Um, J. M.; Çelebi-Ölcüm, N.; Houk, K. N. Quantum Mechanical Investigations of Organocatalysis: Mechanisms, Reactivities, and Selectivities. *Chem. Rev.* **2011**, *111*, 5042−5137. (c) Peng, Q; Duarte, F; Paton, R. S. Computing Organic Stereoselectivity – from Concepts to Quantitative Calculations and Predictions. *Chem. Soc. Rev.* **2016**, *45*, 6093–6107. (d) Iwamoto, H.; Imamoto, T.; Ito, H. Computational Design of High-performance Ligand for Enantioselective Markovnikov Hydroboration of Aliphatic Terminal Alkenes *Nat. Commun.* **2018**, *9*, 2290. (e) Meng, S.-S.; Yu, P.; Yu, Y.-Z.; Liang, Y.; Houk, K. N.; Zheng, W.-H. Computational Design of Enhanced Enantioselectivity in Chiral Phosphoric Acid-Catalyzed Oxidative Desymmetrization of 1,3-Diol Acetals. *J. Am. Chem. Soc.* **2020**, *142*, 8506−8513.
5. (a) Sepúlveda, D.; Lu, T.; Wheeler, S. E. Performance of DFT Methods and Origin of Stereoselectivity in Bipyridine N,N′-dioxide Catalyzed Allylation and Propargylation Reactions *Org. Biomol. Chem.* **2014**, *12*, 8346‑8353. (b) Doney, A. C.; Rooks, B. J.; Lu, T.; Wheeler, S. E. Design of Organocatalysts for Asymmetric Propargylations through Computational Screening. *ACS Catal.* **2016**, *6*, 7948‑7955. (c) Lu, T.; Porterfield, M. A.; Wheeler, S. E. Explaining the Disparate Stereoselectivities of N-Oxide Catalyzed Allylations and Propargylations of Aldehydes. *Org. Lett.* **2012**, *14*, 5310‑5313.
6. Gallarati, S.; Fabregat, R.; Laplaza, R; Bhattacharjee, S.; Wodrich, M. D.; Corminboeuf, C. Reaction-based Machine Learning Representations for Predicting the Enantioselectivity of Organocatalysts. *Chem. Sci.* **2021**, *12*, 6879–6889.
7. (a) Takeda, M.; Yabushita, K.; Yasuda, S.; Ohmiya, H. Synergistic Palladium/Copper-Catalyzed Csp^3–Csp^2 Cross-Couplings Using Aldehydes as Latent α-Alkoxyalkyl Anion Equivalents. *Chem. Commun.* **2018**, *54*, 6776−6779. (b) Yabushita, K.; Yuasa, A.; Nagao, K.; Ohmiya, H. Asymmetric Catalysis Using Aromatic Aldehydes as Chiral α-Alkoxyalkyl Anions. *J. Am. Chem. Soc.* **2019**, *141*, 113−117. (c) Takeda, M.; Mitsui, A.; Nagao, K.; Ohmiya, H. Reductive Coupling between Aromatic Aldehydes and Ketones or Imines by Copper Catalysis. *J. Am. Chem. Soc.* **2019**, *141*, 3664−3669. (d) Mitsui, A; Nagao, K; Ohmiya, H. Copper-Catalyzed Enantioselective Reductive Cross-Coupling of Aldehydes and Imines. *Org. Lett.* **2020**, *22*, 800-803. (e) Yuasa, A; Nagao, K.; Ohmiya, H. Allylic Cross-Coupling Using Aromatic Aldehydes as α-Alkoxyalkyl Anions. *Beilstein J. Org. Chem.* **2020**, *16*, 185–189. (f) Kondo, Y.; Nagao, K.; Ohmiya, H. Reductive Umpolung for Asymmetric Synthesis of Chiral α-Allenic Alcohols. *Chem. Commun.* **2020**, *56,* 7471–7474.
8. (a) Zhao, H.; Dang, L.; Marder, T. B.; Lin, Z. DFT Studies on the Mechanism of the Diboration of Aldehydes Catalyzed by Copper(I) Boryl Complexes. *J. Am. Chem. Soc.* **2008**, *130*, 5586–5594. (b) Kubota, K.; Osaki, S.; Jin, M.; Ito, H. Copper(I)-Catalyzed Enantioselective Nucleophilic Borylation of Aliphatic Ketones: Synthesis of Enantioenriched Chiral Tertiary α-Hydroxyboronates. *Angew. Chem., Int. Ed.* **2017**, *56*, 6646−6650.
9. (a) Cirriez, V.; Rasson, C.; Hermant, T.; Petrignet, J.; Díaz Álvarez, J.; Robeyns, K.; Riant, O. Copper-Catalyzed Addition of Nucleophilic Silicon to Aldehydes. *Angew. Chem., Int. Ed.* **2013**, *52*, 1785–1788. (b) Oestreich, M., Hartmann, E., Mewald, M. Activation of the Si–B Interelement Bond: Mechanism, Catalysis, and Synthesis. *Chem. Rev*. **2013**, *113*, 402–441. (c) Delvos, L. B.; Hensel, A.; Oestreich, M. McQuade's Six-Membered NHC–Copper(I) Complexes for Catalytic Asymmetric Silyl Transfer. *Synthesis* **2014**, *46*, 2957–2964. (d) Hensel, A.; Oestreich, M. Asymmetric Addition of Boron and Silicon Nucleophiles, in Progress in Enantioselective Cu(I)-catalyzed Formation of Stereogenic Centers (Ed.: Harutyunyan, S. R.), Springer International Publishing Switzerland, *Top. Organomet. Chem.* **2016**, *58*, 135–167.
10. (a) Yamaguchi, S; Sodeoka, M. Molecular Field Analysis Using Intermediates in Enantio-Determining Steps Can Extract Information for Data-Driven Molecular Design in Asymmetric Catalysis. *Bull. Chem. Soc. Jpn.* **2019**, *92*, 1701-1706. (b) Chen, H.; Yamaguchi, S.; Morita, Y.; Nakao, H.; Zhai, X.; Shimizu, Y.; Mitsunuma, H.; Kanai, M. Data-driven Catalyst Optimization for Stereodivergent Asymmetric Synthesis of α-allyl Carboxylic Acids by Iridium/boron Hybrid Catalysis. *ChemRxiv* https://doi.org/10.26434/chemrxiv.14579169.v1
11. Although careful selection of training samples is required for predictive purposes as the Demark group has demonstrated (*cf.* ref. 1d), our methodology aims at a mechanistic interpretation. The molecular design is performed using a combination of the researcher's intuition and obtained insights (not predicted values), thus allowing a rough sample selection

based on practical considerations such as availability as described in this manuscript.

12. Kozlowski, M. C.; Dixon, S. L.; Panda, M.; Lauri, G. Quantum Mechanical Models Correlating Structure with Selectivity: Predicting the Enantioselectivity of *β*-Amino Alcohol Catalysts in Aldehyde Alkylation. *J. Am. Chem. Soc.* **2003**, *125*, 6614-6615.
13. Tibshirani, R. Regression Shrinkage and Selection via the LASSO. *J. R. Stat. Soc. B* **1996**, *58*, 267-288.
14. Zou, H.; Hastie, T. Regularization and Variable Selection via the Elastic Net. *J. R. Stat. Soc. B.* **2005**, *67*, 301–322.
15. Yamaguchi, S.; Nishimura T.; Hibe, Y.; Nagai, M.; Sato, H.; Johnston, I. Regularized Regression Analysis of Digitized Molecular Structures in Organic Reactions for Quantification of Steric Effects. *J. Comp. Chem.* **2017**, *38*, 1825-1833.
16. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1–22.
17. Plata, R. E.; Singleton, D. A. A Case Study of the Mechanism of Alcohol-Mediated Morita Baylis−Hillman Reactions. The Importance of Experimental Observations. *J. Am. Chem. Soc*. **2015**, *137*, 3811–3826.
18. Schneider, G.; Fechner, U. Computer-based De Novo Design of Drug-like Molecules. *Nat. Rev. Drug Discov.* **2005**, *4*, 649-663.