Prediction of Reaction Orthogonality using Machine Learning

Hootan Roshandel,[‡] Santiago Vargas,[‡] Amy Lai, Anastassia N. Alexandrova,* and Paula L. Diaconescu*

ABSTRACT: We present a statistical learning model relying on a small dataset to predict the selectivity of a two state system toward the same substrate, specifically of redox-switchable metal complexes in the ring opening polymerization of ε-caprolactone or trimethylene carbonate. We mapped the descriptor space of several switchable metal complexes and surveyed a set of supervised machine learning algorithms using different train/test validation methods on a limited dataset based on experimental studies of ca. 10 metal complexes. Linear discriminant analysis showed an accuracy of >80% and a F1 score of 0.86 on a test mixture of experimental and predicted molecules, and successfully predicted the reactivity of three new metal complexes. The established method will be used to guide future studies in recommending promising new metal complexes for related substrates, reducing the need for blind synthetic trial and error efforts.

Predictive models can determine complex relationships between chemical structure and activity, and streamline the proposal and synthesis of new compounds with optimized properties. Machine learning has become an attractive tool in chemistry to make predictions such as reactivity,[1-6] optimal reaction conditions,[7-10] molecular properties,[11] and mechanistic information.[12-13] These cases typically require large datasets, generated either through systematic high-throughput experiments[14-16] or large-scale computational studies.[17-18] It is, however, less feasible for organometallic systems[19-24] to generate large quantities of data, given the complexity of syntheses that may require multiple steps along with catalytic mechanisms involving several intermediates and transition states. As such, a general method of producing a predictive model with a small dataset is desirable.

Small datasets require a simple model relying on a limited set of interpretable descriptors. The importance of appropriate descriptors is also key, as models with many irrelevant descriptors can struggle to generalize,[25] leading to poor performance. Additionally, highly complex models with a large space of input descriptors tend to be less human-interpretable, making them less pragmatic in building new theories of physical systems.

Despite the difficulty in building datasets of organometallic complexes, quantitative structure-activity relationships (QSARs) have been previously used;[26] for example, reaction productivity of a data set of 51 zirconocene catalysts was modeled with 6 descriptors.[3] Another notable study built a regressive model by incorporating the rate of 19 zirconocene catalysts using a combined set of steric and electronic features. Furthermore, QSAR has been used to predict computationally determined values such as DFT-derived reactivity[27] or supporting ligand effects.[3] Recently, Tong and coworkers reported a Bayesian optimization workflow on a subset of literature results for stereoselective lactide ring-opening polymerization, and identified multiple new aluminum complexes that perform stereoselective polymerization.[28]

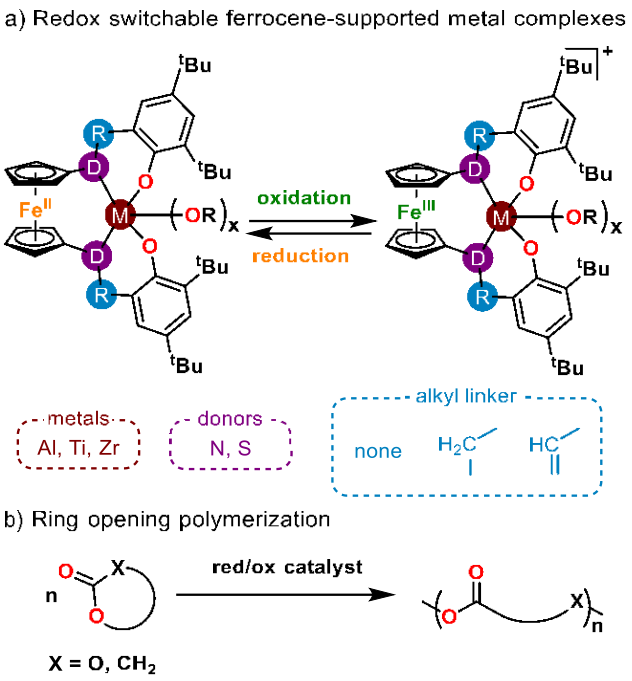


Figure 1. a) General description of previously studied ring opening polymerization (ROP) ferrocene catalysts; b) ROP of cyclic carbonates (X = O) and lactones (X = CH_2).

Over the years, our group has studied several redox active ferrocene-supported metal complexes for their catalytic activity toward the ring opening polymerization (ROP) of monomers such as lactones and cyclic carbonates (**Figure 1**).[29-32] Some metal complexes stood out for their selective, orthogonal reactivity between neutral and oxidized states; we define orthogonal reactivity as a minimum 50% monomer conversion difference between the two oxidation states. This selectivity is important as with a controllable, switchable, and selective catalytic system, diverse block copolymers can be synthesized.[33-38]

Table 1. Redox switchable metal complexes and their reactivity and selectivity toward the ROP of ε-caprolactone (CL) and trimethylene carbonate (TMC) used as training sets.

Entry	Compounds	CL activity		TMC activity	
		Reduced	Oxidized	Reduced	Oxidized
1	(salfan-H_2)Ti(O^iPr)$_2$	< 1%	30%	95%	< 1%
2	(salfan-H_2)Zr(O^tBu)$_2$	< 1%	< 1%	95%	< 1%
3	(salfan)Zr(O^tBu)$_2$[39]	5%	98%	67%*	81%**
4	(salfen)Ti(O^iPr)$_2$	90%	< 1%	19%	40%
5	(salfen)Zr(O^tBu)$_2$[35, 40]	< 1%	< 1%	92%	88%
6	(salfen)Al(O^iPr)[34]	92%	92%	98%	< 1%
7	(thiolfan*)Ti(O^iPr)$_2$[36]	83%	90%	5%	91%
8	(thiolfan*)Zr(O^tBu)$_2$[36]	89%	84%	55%	56%
9	(thiolfan*)Al(O^iPr)[37]	95%	95%	95%	74%
10	(thiolfan)Zr(O^tBu)$_2$[39]	57%	92%	10%	80%

*Polymerization was performed with a 6 mM solution of catalyst in C_6D_6 at room temperature with [monomer]:[catalyst] = 100:1. Conversion was determined through ^{1}H NMR spectroscopy with 1,3,5-trimethoxybenzene as the internal standard. **Catalyst reacted with 1 equivalent of oxidant ([acetylferrocenium][BAr^F]) (Ar^F = 3,5-bis(trifluoromethyl)phenyl)borate) for 30 minutes prior to use, otherwise the conditions are identical to the reduced state experiments.

However, the correlation between a metal complex and orthogonal behavior toward a specific monomer was not obvious. For example, in the case for ROP of ε-caprolactone (CL), only 3 out of 10 metal complexes displayed an orthogonal activity between their reduced and oxidized states (Table 1, entry 1, (salfan-H_2)Ti(O^iPr)$_2$ (salfan-H_2 = 1,1′-di(2-(aminomethyl)-4,6-di-*tert*-butylphenoxy)ferrocene), entry 3, (salfan)Zr(O^tBu)$_2$ (salfan = 1,1′-di(2-*tert*-butyl-6-N-methylmethylenephenoxy)ferrocene),[39] and entry 4, (salfen)Ti(O^iPr)$_2$ (salfen = 1,1′-di(2,4-bis-*tert*-butyl-salicylimino)ferrocene). All other examples[34-37, 39-40] showed non-orthogonal selectivity, where both oxidation states of the catalyst showed either high or low monomer conversions. Similarly, for trimethylene carbonate (TMC), of the 10 metal complexes studied, 5 showed switchable behavior (Table 1, entry 1, (salfan-H_2)Ti(O^iPr)$_2$, entry 2, (salfan-H_2)Zr(O^tBu)$_2$, entry 6, (salfen)Al(O^iPr), entry 7, (thiolfan*)Ti(O^iPr)$_2$ (thiolfan* = 1,1′-bis(2,4-di-*tert*-butyl-6-thiophenoxy)ferrocene), and entry 10, (thiolfan)Zr(O^tBu)$_2$ (thiolfan = 1,1′-bis(2,4-di-*tert*-butyl-6-thiomethylenephenoxy)ferrocene)).[34-37, 39-40] Herein, we aim to demonstrate that a statistical learning approach could be leveraged to understand and predict these obfuscated trends.

Improvement on monomer selectivity and activity is a target that can be optimized by continuous studies in metal and ligand modifications of the precatalysts. Such efforts come with a synthetic trial and error process of each new variation that are costly in time and resources. The outlook for this approach is to iterate between experiment and a statistical model to minimize the number of trials for experimentalists.

Our potential chemical space of interest expands to 15 unique examples, including 5 different ferrocene supporting ligands (Figure 2), which are based on N-type donors (salfan-H_2, salfan, and salfen) and S type donors (thiolfan* and thiolfan), and 3 catalytically active metal centers (Al, Ti, and Zr). Of the possible combinations, the 10 metal complexes whose reactivity with CL and TMC was previously reported are shown in Table 1 and were used as the training set. DFT geometry optimizations were performed for all 10 metal complexes in both the neutral and oxidized states.

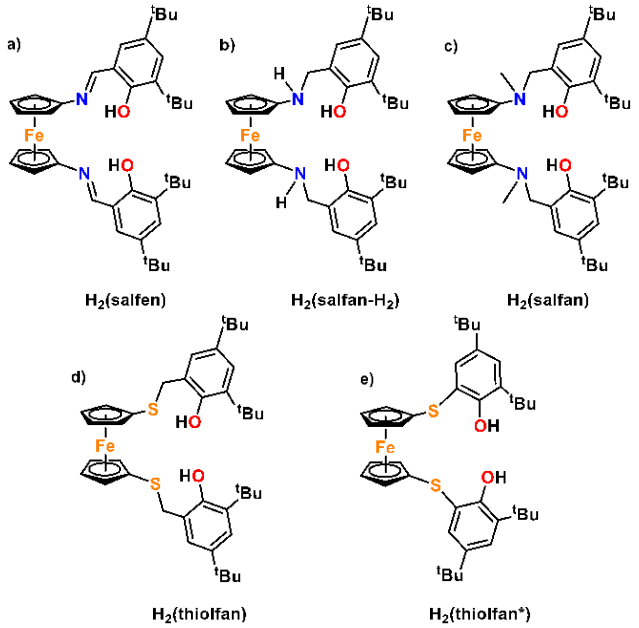


Figure 2. List of frameworks explored in this work.

Chemical descriptors (atomic number, charge, i.e., Gasteiger-Marsili sigma charges in OpenBabel),[41] Pauling electronegativity, coordination number, atomic radius, percent buried volume (%V_{bur}; bond radii and H atoms were included),[42-43] and natural population analysis (NPA) for each metal complex were generated for the metal center and up to 3 atoms away from the metal center (Figure 3b). Atomic identity, charge, electronegativity, coordination number, and atomic radius were passed through an autocorrelation function to reduce the input space of our models. NPA at the DFT level (see SI for details) was also used. This set was the initial set of potential descriptors, which was narrowed down by removing low variance and highly correlated features, and iteratively removing features with train-testing and assessing where performance in accuracy and generalizability decreased. In particular, different ways of expressing the %V_{bur} were highly correlated, as well as atomic identity with atom connectivity. Low variance features included atom connectivity, identity, and electronegativity beyond the immediate coordination sphere. Trained regression models can also be used to perform feature selection of the input space of variables, thereby allowing us to determine which features are critical for predictions. We found that descriptors of the immediate coordination sphere were highly consequential, in line with standard chemical logic. Furthermore, the use of %V_{bur} was also important, in agreement with this feature's relevance to other studied organometallic systems.[42, 44] The feature importance was also used to simplify models to three descriptors associated with %V_{bur}, along with the first-degree autocorrelation functions for charge, electronegativity, atomic radius, and coordination number prior to using principal component analysis (PCA).

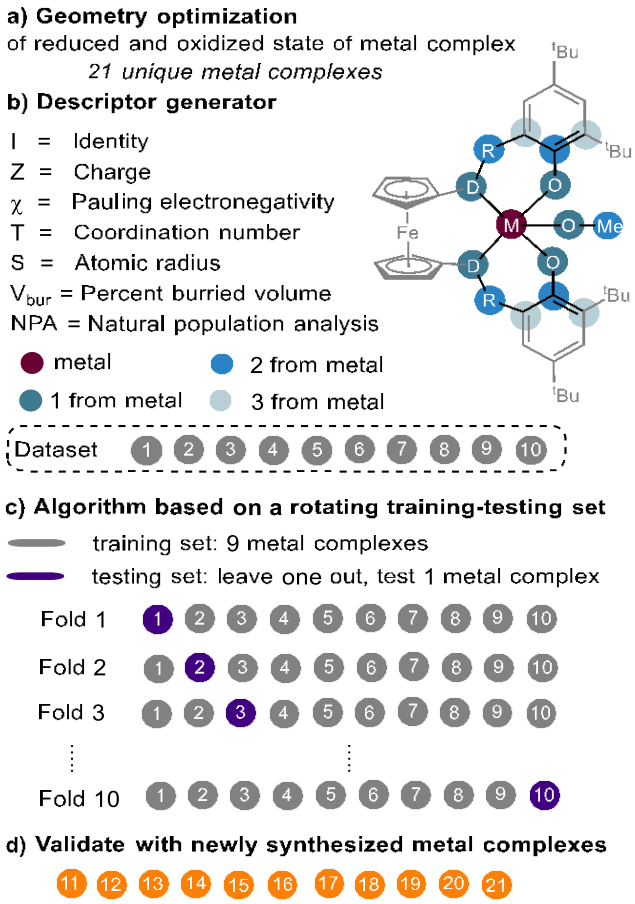


Figure 3. General protocol for developing a ML classifier starting from **a)** DFT calculations to generate structures, **b)** descriptor extraction, **c)** training ML algorithms based on a small dataset, and finally **d)** use of the best model to predict selectivity for the remaining metal complexes.

Four statistical learning algorithms (linear discriminant analysis, quadratic discriminant analysis, logistic regression, and support vector machine)[45-46] were trained and tested. In the process of model selection, we opted to use a leave-one out approach[46] to accommodate our small dataset. For further testing, we also used a stratified testing and training scheme that included a minimum of one type of compound in each set, i.e., at least one orthogonal and one non-orthogonal metal complex. Compounds were partitioned randomly in four different trials and average statistics were reported. Metrics for accuracy/F1/ROC were then taken as the average of all train/testing runs (Figure 3). Our finalized model uses quadratic discriminant analysis for prediction, however, we tested other top performing classification models and found predictions to be consistent with linear discriminant analysis with identical features. Due to the large data imbalance between orthogonal/non-orthogonal examples, we used the synthetic minority oversampling technique (SMOTE)[47] to increase parity when training/testing our model. This algorithm interpolates between points in descriptor space to generate points from the minority class.

Table 2. List of predicted compounds and their experimentally determined reactivity toward ε-caprolactone (CL) and trimethylene carbonate (TMC).

Entry	Catalyst	CL activity			TMC activity		
		Prediction	Reduced	Oxidized	Prediction	Reduced	Oxidized
1	(salfan-H_2)Al(O^{i}Pr)	Non orthogonal	95%	90%	Non orthogonal	> 99%	> 99%
2	(thiolfan)Al(OBn)	Orthogonal	> 99%	45%	Non orthogonal	> 99%	> 99%

PCA was also used to simplify our descriptor space as input to classification models where we used four components. Previous work used PCA to decompose the descriptors of different compounds to simple, visual plots that still encoded the differences between those compounds.[48] Here we found that the metal identity largely determined the mapping of metal complexes in the PCA latent spaces (Figure S3-5). It appears that there is no clear separation between the orthogonal/non-orthogonal examples. This is, however, not the case when we project to the 3D component space where the dividing plane of separation between orthogonal/non-orthogonal complexes is clear.

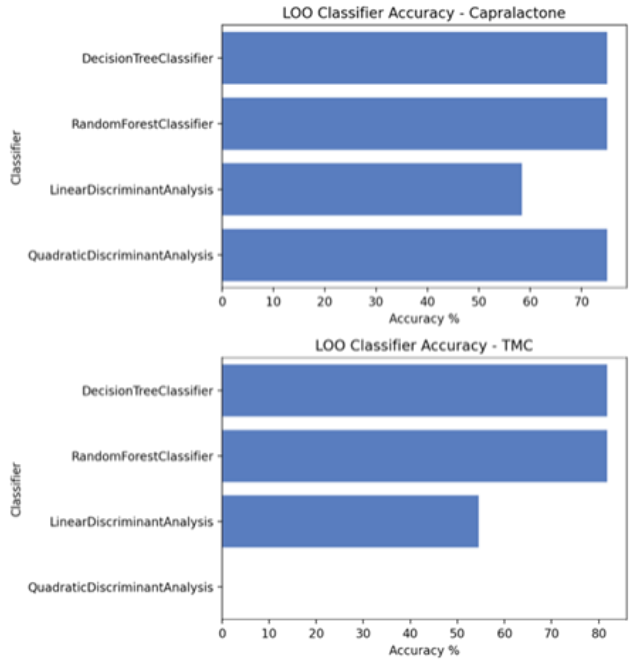


Figure 4. The performance of the model with both leave-one-out testing and stratified testing, where it was ensured an orthogonal and non-orthogonal example were in the testing set. F1 score evenly weights different categories to offset the effects of imbalanced datasets.

Furthermore, we used the algorithm to predict the CL and TMC reactivity and selectivity of 3 metal complexes: (salfan)Ti(OR)$_2$, (salfan-H_2)Al(OR), and (thiolfan)Al(OR); R was Me for the computational models and ^{i}Pr or Bn = benzyl for the experimental studies. Our group then synthesized and tested the orthogonality of the new compounds; while we were conducting our studies, the reactivity of (thiolfan)Al(OBn) with CL was reported.[49] The algorithm predicted that an orthogonal selectivity will be observed between the reduced and oxidized states of (thiofan)Al(OR) for CL and TMC, which was experimentally validated as true. It also predicted non-orthogonal selectivity for (salfan-H_2)Al(OR) and (salfan)Ti(OR)$_2$ toward CL and TMC, also confirmed to be correct (**Table 2**).

In summary, a predictive machine learning model has been achieved through a small data set of 10 metal complexes. The model was validated through the reactivity studies of 3 compounds for CL and TMC. We are also currently investigating ML for predicting orthogonal selectivity toward other monomers.

ASSOCIATED CONTENT

Supporting Information: Computational methods, statistics, experimental studies (pdf).

AUTHOR INFORMATION

Corresponding Authors

Anastassia N. Alexandrova − Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095-1569, United States; California NanoSystems Institute, Los Angeles, California 90095, United States; orcid.org/0000-0002-3003-1911; email: ana@chem.ucla.edu

Paula L. Diaconescu − Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095-1569, United States; orcid.org/0000-0003-2732-4155; email: pld@chem.ucla.edu

Author Information

Hootan Roshandel ‡ − Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095-1569, United States; orcid.org/0000-0002-7744-0958; Email: hootan@chem.ucla.edu

Santiago Vargas ‡ − Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095-1569, United States; orcid.org/0000-0002-1634-0945; Email: santiagovargas@chem.ucla.edu

Amy Lai − Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095-1569, United States; Email: alai7039@gmail.com

Author Contributions
‡ Denotes equal contribution from the authors

Funding Sources

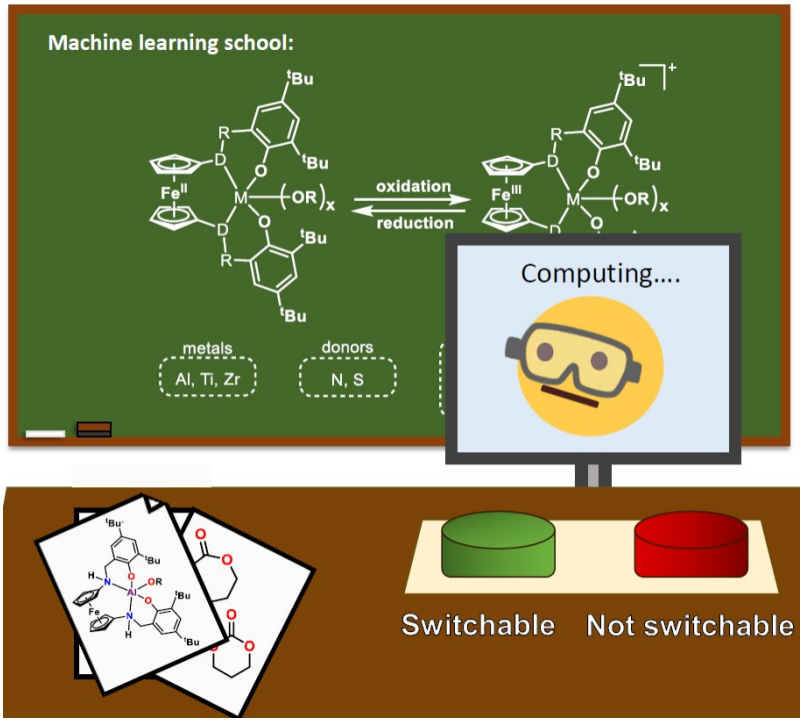
This work was supported by the National Science Foundation grant CHE-1809116 to PLD, Department of Energy, Office of Basic Sciences grant DE-SC0019152 to ANA. S.V. is thankful for a Department of Energy Computational Science Graduate Fellowship, which supported this work under DE-SC0021110. This work used Bridges-2 at Pittsburgh Supercomputing Center from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation grants 2138259, 2138286, 2138307, 2137603, and 2138296.

REFERENCES

1. Maity, B.; Cao, Z.; Kumawat, J.; Gupta, V.; Cavallo, L., A Multivariate Linear Regression Approach to Predict Ethene/1-Olefin Copolymerization Statistics Promoted by Group 4 Catalysts. *ACS Catal.* **2021,** *11* (7), 4061-4070.
2. Fang, C.; Fantin, M.; Pan, X.; de Fiebre, K.; Coote, M. L.; Matyjaszewski, K.; Liu, P., Mechanistically Guided Predictive Models for Ligand and Initiator Effects in Copper-Catalyzed Atom Transfer Radical Polymerization (Cu-ATRP). *J. Am. Chem. Soc.* **2019,** *141* (18), 7486-7497.
3. Parveen, R.; Cundari, T. R.; Younker, J. M.; Rodriguez, G.; McCullough, L., DFT and QSAR Studies of Ethylene Polymerization by Zirconocene Catalysts. *ACS Catal.* **2019**, 9339-9349.
4. Żurański, A. M.; Martinez Alvarado, J. I.; Shields, B. J.; Doyle, A. G., Predicting Reaction Yields via Supervised Learning. *Acc. Chem. Res.* **2021,** *54* (8), 1856-1865.
5. Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G., Predicting reaction performance in C-N cross-coupling using machine learning. *Science* **2018,** *360* (6385), 186-190.
6. Vargas, S.; Hennefarth, M. R.; Liu, Z.; Alexandrova, A. N., Machine Learning to Predict Diels–Alder Reaction Barriers from the Reactant State Electron Density. *J. Chem. Theory Comput.* **2021,** *17* (10), 6203-6213.
7. Gao, H.; Struble, T. J.; Coley, C. W.; Wang, Y.; Green, W. H.; Jensen, K. F., Using Machine Learning To Predict Suitable Conditions for Organic Reactions. *ACS Cent. Sci.* **2018,** *4* (11), 1465-1476.
8. Saebi, M.; Nan, B.; Herr, J. E.; Wahlers, J.; Guo, Z.; Zurański, A. M.; Kogej, T.; Norrby, P.-O.; Doyle, A. G.; Chawla, N. V.; Wiest, O., On the use of real-world datasets for reaction yield prediction. *Chem. Sci.* **2023,** *14* (19), 4997-5005.
9. 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), eaau5631.
10. Reid, J. P.; Sigman, M. S., Holistic prediction of enantioselectivity in asymmetric catalysis. *Nature* **2019,** *571* (7765), 343-348.
11. Janet, J. P.; Liu, F.; Nandy, A.; Duan, C.; Yang, T.; Lin, S.; Kulik, H. J., Designing in the Face of Uncertainty: Exploiting Electronic Structure and Machine Learning Models for Discovery in Inorganic Chemistry. *Inorg. Chem.* **2019,** *58* (16), 10592-10606.
12. Lustosa, D. M.; Milo, A., Mechanistic Inference from Statistical Models at Different Data-Size Regimes. *ACS Catal.* **2022,** *12* (13), 7886-7906.
13. Żurański, A. M.; Gandhi, S. S.; Doyle, A. G., A Machine Learning Approach to Model Interaction Effects: Development and Application to Alcohol Deoxyfluorination. *J. Am. Chem. Soc.* **2023,** *145* (14), 7898-7909.
14. Robbins, D. W.; Hartwig, J. F., A Simple, Multidimensional Approach to High-Throughput Discovery of Catalytic Reactions. *Science* **2011,** *333* (6048), 1423-1427.
15. Collins, K. D.; Gensch, T.; Glorius, F., Contemporary screening approaches to reaction discovery and development. *Nat. Chem.* **2014,** *6* (10), 859-871.
16. Cannavacciuolo, F. D.; Yadav, R.; Esper, A.; Vittoria, A.; Antinucci, G.; Zaccaria, F.; Cipullo, R.; Budzelaar, P. H. M.; Busico, V.; Goryunov, G. P.; Uborsky, D. V.; Voskoboynikov, A. Z.; Searles, K.; Ehm, C.; Veige, A. S., A High-Throughput Approach to Repurposing Olefin Polymerization Catalysts for Polymer Upcycling. *Angew. Chem. Int. Ed.* **2022,** *61* (23), e202202258.
17. Friederich, P.; dos Passos Gomes, G.; De Bin, R.; Aspuru-Guzik, A.; Balcells, D., Machine learning dihydrogen activation in the chemical space surrounding Vaska's complex. *Chem. Sci.* **2020,** *11* (18), 4584-4601.
18. 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* (1), 110-121.
19. Nandy, A.; Duan, C.; Taylor, M. G.; Liu, F.; Steeves, A. H.; Kulik, H. J., Computational Discovery of Transition-metal Complexes: From High-throughput Screening to Machine Learning. *Chem. Rev.* **2021,** *121* (16), 9927-10000.
20. Durand, D. J.; Fey, N., Computational Ligand Descriptors for Catalyst Design. *Chem. Rev.* **2019,** *119* (11), 6561-6594.
21. Durand, D. J.; Fey, N., Building a Toolbox for the Analysis and Prediction of Ligand and Catalyst Effects in Organometallic Catalysis. *Acc. Chem. Res.* **2021,** *54* (4), 837-848.
22. Matsuoka, W.; Harabuchi, Y.; Maeda, S., Virtual Ligand Strategy in Transition Metal Catalysis Toward Highly Efficient Elucidation of Reaction Mechanisms and Computational Catalyst Design. *ACS Catal.* **2023,** *13* (8), 5697-5711.
23. Fey, N.; Lynam, J. M., Computational mechanistic study in organometallic catalysis: Why prediction is still a challenge. *WIREs Comput. Mol. Sci.* **2022,** *12* (4), e1590.
24. Duan, C.; Nandy, A.; Meyer, R.; Arunachalam, N.; Kulik, H. J., A transferable recommender approach for selecting the best density functional approximations in chemical discovery. *Nat. Comput. Sci.* **2023,** *3* (1), 38-47.
25. Tom, G.; Hickman, R. J.; Zinzuwadia, A.; Mohajeri, A.; Sanchez-Lengeling, B.; Aspuru-Guzik, A., Calibration and generalizability of probabilistic models on low-data chemical datasets with DIONYSUS. *Digital Discovery* **2023,** *2* (3), 759-774.
26. Cruz, V. L.; Martinez, S.; Ramos, J.; Martinez-Salazar, J., 3D-QSAR as a Tool for Understanding and Improving Single-Site Polymerization Catalysts. A Review. *Organometallics* **2014,** *33* (12), 2944-2959.
27. Yang, B.; Truhlar, D. G., Computational Design of an Iron Catalyst for Olefin Metathesis. *Organometallics* **2018,** *37* (21), 3917-3927.
28. Wang, X.; Huang, Y.; Xie, X.; Liu, Y.; Huo, Z.; Lin, M.; Xin, H.; Tong, R., Bayesian-optimization-assisted discovery of stereoselective aluminum complexes for ring-opening polymerization of racemic lactide. *Nat. Commun.* **2023,** *14* (1), 3647.
29. Lai, A.; Hern, Z. C.; Shen, Y.; Dai, R.; Diaconescu, P. L., 9.07 - Metal Complexes for Redox Switching and Control of Reactivity. In *Comprehensive Coordination Chemistry III*, Constable, E. C.; Parkin, G.; Que Jr, L., Eds. Elsevier: Oxford, 2021; pp 155-180.
30. Deng, S.; Diaconescu, P. L., A switchable dimeric yttrium complex and its three catalytic states in ring opening polymerization. *Inorg. Chem. Front.* **2021,** *8* (8), 2088-2096.

31. Xu, X.; Luo, G.; Hou, Z.; Diaconescu, P. L.; Luo, Y., Theoretical insight into the redox-switchable activity of group 4 metal complexes for the ring-opening polymerization of ε-caprolactone. *Inorg. Chem. Front.* **2020,** *7* (4), 961-971.
32. Wei, J.; Diaconescu, P. L., Redox-switchable Ring-opening Polymerization with Ferrocene Derivatives. *Acc. Chem. Res.* **2019,** *52* (2), 415-424.
33. Hern, Z. C.; Quan, S. M.; Dai, R.; Lai, A.; Wang, Y.; Liu, C.; Diaconescu, P. L., ABC and ABAB Block Copolymers by Electrochemically Controlled Ring-Opening Polymerization. *J. Am. Chem. Soc.* **2021,** *143* (47), 19802-19808.
34. Lai, A.; Hern, Z. C.; Diaconescu, P. L., Switchable Ring-Opening Polymerization by a Ferrocene Supported Aluminum Complex. *ChemCatChem* **2019,** *11* (16), 4210-4218.
35. Dai, R.; Diaconescu, P. L., Investigation of a Zirconium Compound for Redox Switchable Ring Opening Polymerization. *Dalton Trans.* **2019,** *48*, 2996-3002.
36. Lowe, M. Y.; Shu, S.; Quan, S. M.; Diaconescu, P. L., Investigation of redox switchable titanium and zirconium catalysts for the ring opening polymerization of cyclic esters and epoxides. *Inorg. Chem. Front.* **2017,** *4*, 1798-1805.
37. Wei, J.; Riffel, M. N.; Diaconescu, P. L., Redox Control of Aluminum Ring-Opening Polymerization: A Combined Experimental and DFT Investigation. *Macromolecules* **2017,** *50* (5), 1847-1861.
38. Quan, S. M.; Wang, X.; Zhang, R.; Diaconescu, P. L., Redox Switchable Copolymerization of Cyclic Esters and Epoxides by a Zirconium Complex. *Macromolecules* **2016,** *49* (18), 6768-6778.
39. Wang, X.; Thevenon, A.; Brosmer, J. L.; Yu, I.; Khan, S. I.; Mehrkhodavandi, P.; Diaconescu, P. L., Redox Control of Group 4 Metal Ring-Opening Polymerization Activity toward l-Lactide and ε-Caprolactone. *J. Am. Chem. Soc.* **2014,** *136* (32), 11264-11267.
40. Dai, R.; Lai, A.; Alexandrova, A. N.; Diaconescu, P. L., Geometry Change in a Series of Zirconium Compounds during Lactide Ring-Opening Polymerization. *Organometallics* **2018,** *37* (21), 4040-4047.
41. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R., Open Babel: An open chemical toolbox. *J. Cheminform.* **2011,** *3* (1), 33.
42. Falivene, L.; Cao, Z.; Petta, A.; Serra, L.; Poater, A.; Oliva, R.; Scarano, V.; Cavallo, L., Towards the online computer-aided design of catalytic pockets. *Nat. Chem.* **2019,** *11* (10), 872-879.
43. Antinucci, G.; Dereli, B.; Vittoria, A.; Budzelaar, P. H. M.; Cipullo, R.; Goryunov, G. P.; Kulyabin, P. S.; Uborsky, D. V.; Cavallo, L.; Ehm, C.; Voskoboynikov, A. Z.; Busico, V., Selection of Low-Dimensional 3-D Geometric Descriptors for Accurate Enantioselectivity Prediction. *ACS Catal.* **2022,** *12* (12), 6934-6945.
44. 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.
45. Platt, J., Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. *Adv. Large Margin Classif.* **2000,** *10*.
46. Hastie, T.; Tibshirani, R.; Friedman, J., *The Elements of Statistical Learning*. 2nd ed.; Springer New York, NY, 2009.
47. Chawla, N.; Bowyer, K.; Hall, L.; Kegelmeyer, W., SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **2002,** *16*, 321-357.
48. Lai, A.; Clifton, J.; Diaconescu, P. L.; Fey, N., Computational mapping of redox-switchable metal complexes based on ferrocene derivatives. *Chem. Commun.* **2019,** *55,* , 7021-7024
49. Kosuru, S. R.; Chang, Y.-L.; Chen, P.-Y.; Lee, W.; Lai, Y.-C.; Ding, S.; Chen, H.-Y.; Chen, H.-Y.; Chang, Y.-C., Ring-Opening Polymerization of ε-Caprolactone by Using Aluminum Complexes Bearing Aryl Thioether Phenolates: Labile Thioether Chelation. *Inorg. Chem.* **2022,** *61* (9), 3997-4008.

TOC Graphic


Machine learning school:
oxidation
reduction
metals
Al, Ti, Zr
donors
N, S
Computing....
Switchable
Not switchable