Quantum-Mechanical Transition-State Model Combined with Machine Learning Provides Catalyst Design Features for Selective Cr Olefin Oligomerization

Steven M. Maley,[‡Δ] Doo-Hyun Kwon,[‡Δ] Nick Rollins,[‡] Johnathan C. Stanley,[‡] Orson L. Sydora,[†] Steven M. Bischof,[†*] and Daniel H. Ess[‡*]

[‡]Department of Chemistry and Biochemistry, Brigham Young University, Provo, Utah, 84602 (USA); [†]Research and Technology, Chevron Phillips Chemical Company LP, 1862, Kingwood Drive, Kingwood, Texas 77339 (USA)

* dhe@chem.byu.edu, bischs@cpchem.com

[Δ]Authors contributed equally to this work.

Abstract

The use of data science tools to provide the emergence of nontrivial chemical features for catalyst design is an important goal in catalysis science. Additionally, there is currently no general strategy for computational homogeneous, molecular catalyst design. Here we report the unique combination of an experimentally verified DFT-transition-state model with a random forest machine learning model in a campaign to design new molecular Cr phosphine imine (Cr(P,N)) catalysts for selective ethylene oligomerization, specifically to increase 1-octene selectivity. This involved the calculation of 1-hexene:1-octene transition-state selectivity for 105 (P,N) ligands and the harvesting of 14 descriptors, which were then used to build a random forest regression model. This model showed the emergence of several key design features, such as Cr–N distance, Cr–α distance, and Cr distance out of pocket, which were then used to rapidly design a new generation of Cr(P,N) catalyst ligands that are predicted to give >95% selectivity for 1-octene.

Introduction

Computational chemistry now plays an active role in molecular catalyst design and optimization by either testing chemical hypotheses or directly evaluating catalyst candidates.[1] However, no general strategy for virtual catalyst design or optimization has emerged,[2,3,12–14,4–11] and there are only a few cases of specific catalyst prediction followed by experimental realization.[15–20] Due to the complex electronic structure of metal centers and ligands, one strategy for computational homogeneous, molecular catalyst design is to use quantum-mechanical methods to model transition states.[21–27] While quantum-mechanical transition states can often replicate experiment and be used for catalyst prediction, it is not always straightforward to identify simple chemical features that control catalysis,[28,29] especially for selectivity where small energy quantities can impart significant influence. In this type of scenario, a catalyst design workflow that combines quantum-mechanical transition state modeling with machine learning has the potential to reveal critical catalyst design features.

We recently reported the development and use of a density functional theory (DFT) transition-state model that provided quantitative prediction of molecular Cr catalysts for controllable selective ethylene trimerization and tetramerization (Scheme 1).[30] This selective catalyst design effort is important because the ubiquity of polyethylene resulting from robust Phillips[31] and Ziegler-Natta catalysts[32,33] has led to an increase in global need for linear α-olefin (LAO) polymerization comonomers.[34] Additionally, these short-chain LAOs, specifically 1-hexene and 1-octene, are used in the manufacture of plasticizers, lubricants, detergents, and plastomers/elastomers. Prior to our work, Sydora and co-workers reported a series of aryl and benzyl substituted phosphine imine (P,N) catalysts that experimentally produced only ~30% 1-octene.[35] In our computational catalyst design effort, we used the cationic high-spin transition-states **TS1** and **TS2** shown in Scheme 1b to develop a linear correlation model between DFT computed values and experimental 1-hexene:1-octene ratios. This allowed us to use transition-state calculations to computationally design a new general class of phosphine monocyclic imine Cr(P,N) catalysts where changes in the ligand structure control 1-hexene versus 1-octene selectivity (Scheme 1b). Experimental ligand and catalyst synthesis, and reaction testing, quantitatively confirmed our transition-state predictions.
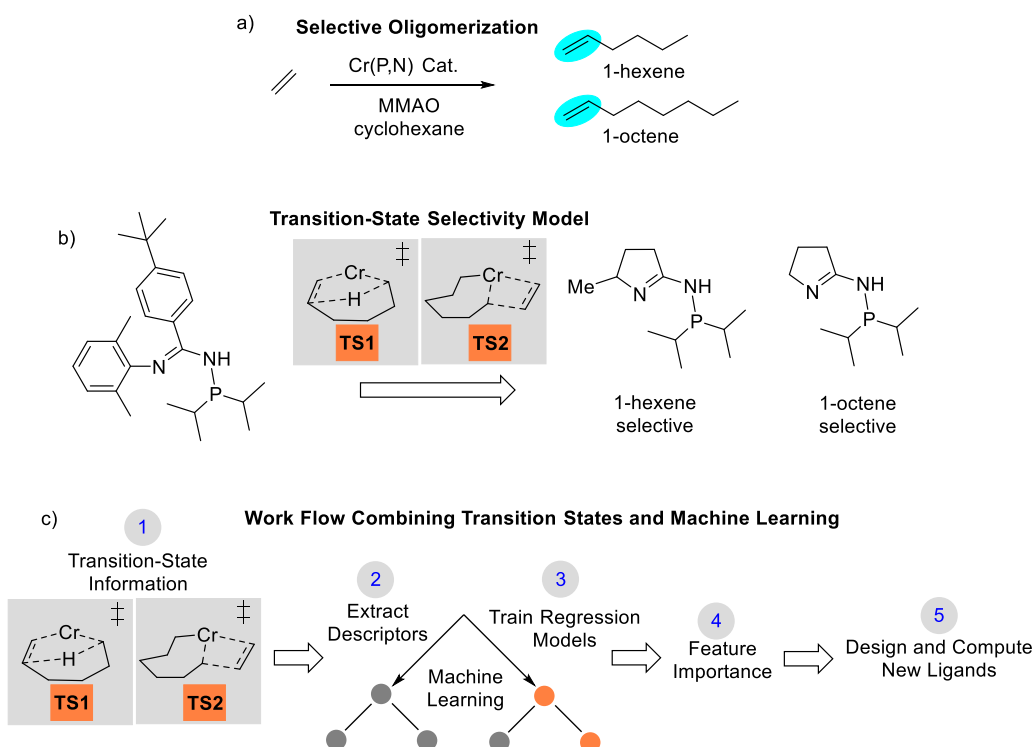
Scheme 1. a) Outline of Cr catalyzed selective ethylene oligomerization reaction conditions with targets of 1-hexene and 1-octene. The catalyst involves a Cr metal center with phosphine and imine ligand coordination. MMAO is typically used to activate the pre-catalyst complex. b) Outline of using **TS1** and **TS2** as a selectivity model to design new monocyclic imine Cr(P,N) catalysts that are 1-hexene and 1-octene selective. c) The work presented here involves using our 1-hexene/1-octene transition-state selectivity model combined with machine learning models to reveal selectivity controlling features that are then used for virtual design new catalyst ligands.

While our DFT transition-state model is practical, accurate, and successfully identified new ligands that were experimentally validated, it did not offer general catalyst design guidance to enhance 1-octene selectivity. Moreover, the interpretation of singular controlling transition-state features is unobvious since the energy difference between 1-hexene and 1-octene selectivity is relatively small. Therefore, we decided to combine our transition-state model with quantitative data science methods that can potentially provide the emergence of chemical features to enhance 1-octene selectivity. While machine learning and related multi-dimensional methods are beginning to be used for molecular catalyst design,[36] there are currently no examples of an experimentally verified quantum-mechanical transition model merged with machine learning methods. Generally, data science approaches to molecular catalyst design emphasize ground-state properties of either pre-catalysts or ligands without metal centers. For example, Fey and Pringle developed

databases of ground-state ligand properties[37–42] that enabled the prediction and experimental verification of new fluorophosphine ligands for hydroformylation and hydrocyanation.[43] Recently Denmark reported a workflow where a subset of a library of synthetically accessible catalysts are selected and tested. The data obtained was then used to train statistical learning models to optimize a chiral catalyst for thiol addition to N-acylimines.[44] Machine learning methods have also been used to predict reaction barrier heights in heterogeneous catalyst applications.[45–47]

Outlined in Scheme 1c, our approach reported here involves using DFT-computed transition-state features and selectivities for machine learning analysis. The analysis of >100 Cr(P,N) catalysts and 14 molecular descriptors through machine learning regression algorithms with multifold cross validation resulted in a low root mean square error (RMSE) and emergence of three critical design elements to enhance 1-octene selectivity. The utility of these machine-learning identified selectivity features was demonstrated by the design and calculation of several new ligands that are predicted to give >90% 1-octene selectivity.

Results and Discussion

Companies such as Shell, Ineos, Idemitsu, SABIC, and Chevron Phillips Chemical Company LP (CPChem) use metal-catalysed ethylene oligomerization to produce a wide distribution of LAOs from 1-butene to 1-eicosene (C_{20}).[48] CPChem was the first to commercialize a selective ethylene trimerization to 1-hexene system using 2-ethylhexanoate Cr along with 2,5-dimethylpyrrole and aluminum co-activators.[49] Sasol later developed selective ethylene oligomerization technology that produces both 1-hexene and 1-octene using a Cr-diphosphinoamine catalyst.[50] However, an important and major unsolved challenge is to develop a general set of molecular catalysts and catalyst design principles that result in highly selective ethylene tetramerization to 1-octene.

As discussed in the introduction, in our recent computational campaign, we developed a DFT transition-state model that provided quantitative prediction of molecular Cr(P,N) catalysts (Scheme 1).[30] This resulted in the computational design and experimental verification of a new general class of phosphine monocyclic imine Cr(P,N) catalysts where slightly more than 50% 1-octene was formed. Importantly, this

DFT transition-state model is accurate and relatively fast to use. New ligands can be virtually screened by calculating the energies of transition-state conformation ensembles for **TS1** and **TS2** (Scheme 1b),[51,52,61,62,53–60] which can then be used in our linear correlation model to provide a predicted 1-hexene:1-octene ratio for CPChem reaction conditions. While this DFT transition-state model is extremely useful, because it is a correlation model with small energy differences, it did not reveal general catalyst design principles that would easily lead to the design of extremely high 1-octene selectivity. Because our transition-state selectivity model is accurate, this provided a platform for combining it with machine learning data science methods that can provide the emergence of general chemical features to enhance 1-octene selectivity for Cr(P,N) catalysts. To our knowledge, there has not been a previous report of combining an experimentally verified transition-state model with machine learning analysis as a workflow to design new catalysts.

While machine learning is beginning to find significant application as a tool to aid organic synthesis,[63,64,73,65–72] there are relatively few examples of machine learning applied to inorganic or organometallic reactions, especially heterogeneous[45,74,83–86,75–82] and homogeneous catalysis.[87,88] Recently, Kulik trained an artificial neural network to predict the high-spin to low-spin splitting energies of ~2700 transition metal complexes.[89] They also demonstrated the usage of a kernel ridge regression model for predicting spin-splitting, bond lengths, and redox potentials for a relatively large collection of transition metal complexes.[90] Related, Corminbouef trained machine learning models to screen over 18000 potential homogeneous catalysts for the Suzuki-Miyaura C-C cross-coupling reaction.[91] Sunoj used a combination of a neural network and random forest model to identify the regioselectivity of catalytic difluorination of alkenes.[92] Brgoch screened over 100,000 compounds using a support vector machine regression to identify novel highly compressible metal materials,[93] and Buehler used convolutional neural networks to search for new composite metal materials.[94] Xin used artificial neural networks to identify heterogeneous metallic catalysts for CO capture and reduction.[95]

With the success of previous machine learning studies predicting spin-splitting energies and redox potentials,[89,90] we were relatively confident that one or more machine learning algorithms would perform well for our workflow. Similar to our previous computational studies,[30] and computational assessments by McGuinness suggesting the good performance of M06L,[51,52] we used the unrestricted M06L density functional[96] for describing the electronic structure of Cr(P,N) catalysts. The UM06L/Def2-TZVPP//UM06L/6-31G**[LANL2DZ] level of theory was combined with the SMD continuum model[97] for cyclohexane to estimate the free energies of **TS1** and **TS2** (see SI for computational details). In this transition state model, the relative free energies of **TS1** and **TS2** provide selectivity under the assumption of Curtin-Hammett type conditions. All transition-state structures were optimised, and vibrational frequencies were computed to verify the stationary points as first-order saddle points. Normal rigid-rotor harmonic oscillator approximations were applied with free energies at 1 atm and 298 K. Because the transition-state model is a linear correlation scheme, no temperature or pressure corrections were applied. All DFT calculations were performed using Gaussian 09.[98] Machine learning analyses were performed using scipy,[99] numpy,[100] pandas[101] and scikit-learn[102] in Python 2.7. The source code is available in the SI.

We used 105 unique (P,N) ligands in our transition-state training data set, which included 14 experimentally measured ligands. Scheme 2 outlines these (P,N) ligands, which have a variety of different functional groups, but retain the phosphine and imine, or imine-like, ligand coordination for which the DFT transition-state model was developed. A major motivation in the selection of these 105 ligands was to further optimize the five-membered imine ring system we previously designed and to stay within the bounds of accuracy for our correlation model. As shown in Scheme 2, this set includes a variety of substituted heterocycles such as pyrrols, imidazoles, and oxazoles. In each of these cases we also examined their combination with alkyl, fluoroalkyl, aryl, and amido phosphines as well as phospholanes.

X and Y: H, F, CH$_3$, CF$_3$, OMe

Z: H or CH$_3$

R: CH$_3$, CF$_3$, Ph, NMeC$_2$H$_4$NMe, OC$_2$H$_4$O, C$_4$H$_8$
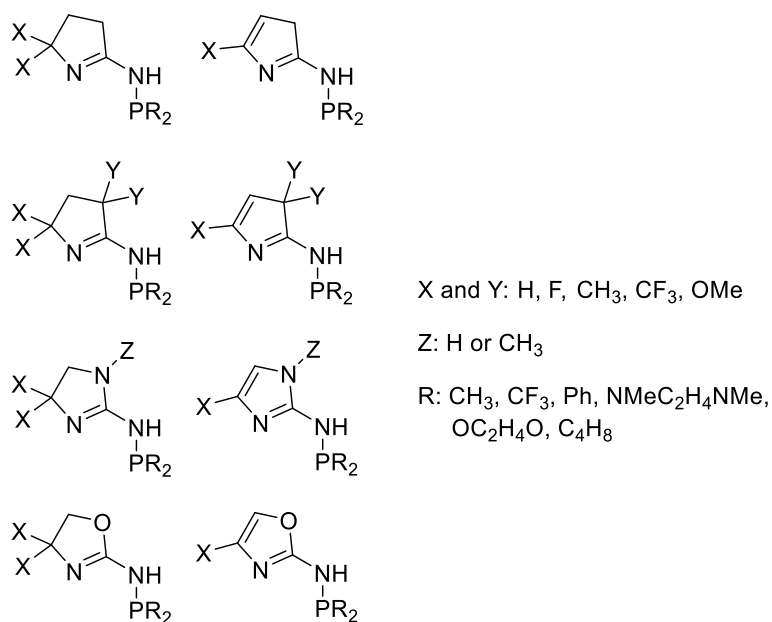
Scheme 2. Outline of 105 unique (P,N) ligands in our transition-state training data set. These ligands were used to calculate selectivity based on **TS1** and **TS2**. Transition-state features were then harvested from the electronic structure and geometries of **TS1** and **TS2**.
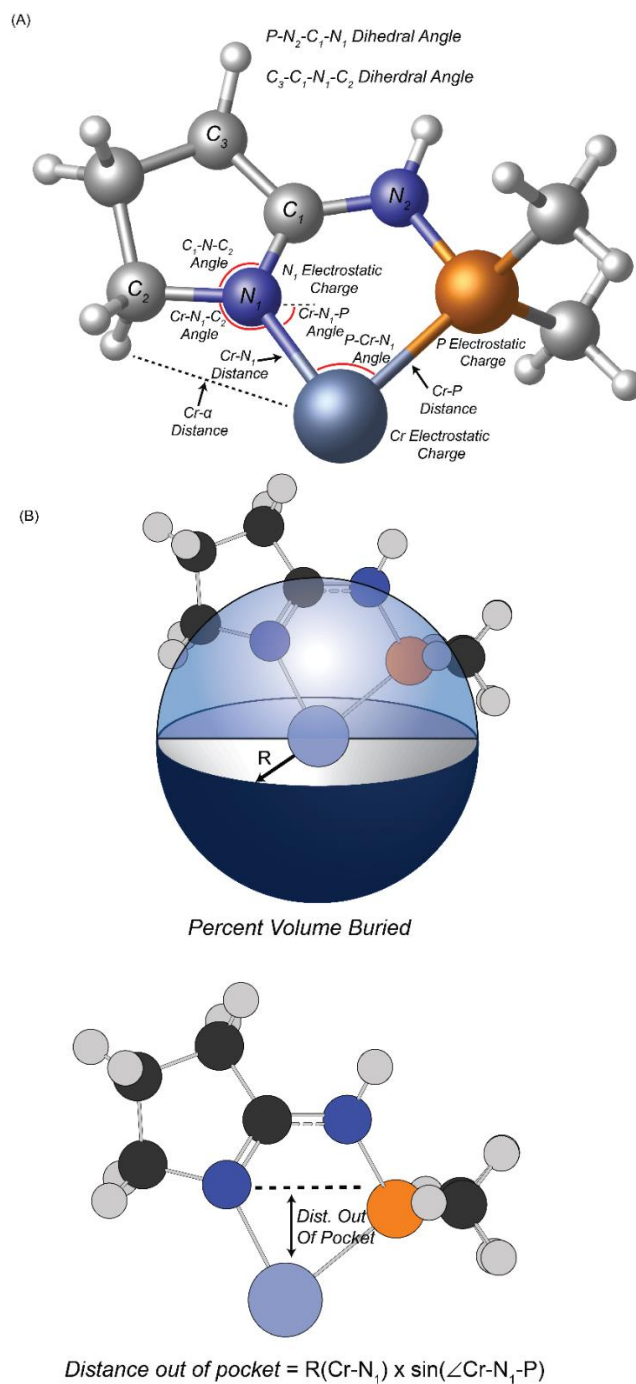
(A)

P-N₂-C₁-N₁ Dihedral Angle

C₃-C₁-N₁-C₂ Diherdral Angle

Figure 1. Descriptors extracted for machine learning analysis. a) Geometric descriptors and electrostatic charges. b) Definition of percent volume buried. c) Definition of distance out of pocket.

14 atomic and molecular descriptors/features were extracted from **TS1** and **TS2** for each of the 105 ligands shown in Scheme 2. Described in Figure 1, extracted features included geometric parameters such

as bond lengths, angles, dihedrals, percent volume buried, and Cr metal center distance out pocket. Several electronic features were also harvested, such as electrostatic-based atomic charges. Percent volume buried describes the extent to which the first coordination sphere of the Cr metal center is occupied by a (P,N) ligand.[103] The distance out of pocket describes the how far the Cr metal is situated from the (P,N) ligand.

The Scikit-Learn python library was used to set up and train regressors on this transition-state data set, which was split into 25% training and 75% testing sets. Seven regression algorithms were tested including: random forest, Gaussian process regression, LASSO, elastic-net, ridge regression and support vector regression with both a linear and radial basis function kernel. Multifold cross validation was performed to protect against model overfitting common in small datasets. This random sampling was performed 10 times and 20-fold cross validation was used at each iteration to determine regression accuracy. The RMSE of each model determined using cross validation averaged across iterations is shown in Figure 2.
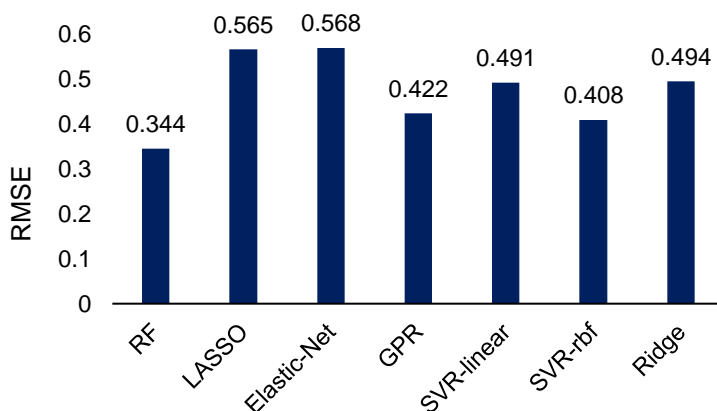

Figure 2. Root mean square error (RMSE) for machine learning regression algorithms to quantitatively predict **TS1** an **TS2** energy differences using 14 atomic and molecular features. RF = random forest, LASSO = least absolute shrinkage and selection operator, GPR = Gaussian process regression, SVR = support vector regression.

The machine learning regression algorithms were used to evaluate the use of the 14 atomic and molecular features to quantitatively correlate with the DFT calculated energy differences between **TS1** and **TS2**. The RMSE of the regression algorithms ranged from 0.344 to 0.568 (Figure 2). The best performing

model was random forest (RMSE = 0.344) and the poorest performing model was elastic-net (RMSE = 0.568). Unsurprisingly the LASSO and ridge algorithms, which are related to elastic-net, also performed poorly with RMSEs of 0.565 and 0.494, respectively. The performance of support vector regression improved by almost 10% when changing from a linear (RMSE = 0.491) to a radial basis function (RMSE = 0.408) kernel. Gaussian process regression performed comparable to SVR-rbf.

Related to the top performance of random forest for our correlation of 1-hexene:1-octene ratios, Doyle reported that random forest outperformed linear models, SVR, k-nearest neighbor, and artificial neural networks at predicting yields of Pd-catalysed Buchwald-Hartwig cross-coupling of aryl halides with 4-methylaniline.[68] Also, for spin-splitting energies and bond lengths of transition metal complexes, Kulik achieved exceptionally low mean unsigned errors using random forest.[90] Palmer et al. showed that random forest outperforms SVR and artificial neural networks in predicting the aqueous solubility of organic molecules.[104]

Because the random forest algorithm performed well in our case, and the applications mentioned above, this algorithm was chosen for further hyperparameter optimization using the GridSearch CV method from SciKit-Learn. Different permutations of hyperparameters and five-fold cross validation were tested in order to determine the set of hyperparameters that maximized the performance of the model. The number of trees in each forest was varied from 20 to 210 and the trees were split from 5 to 125 times. Both mean signed error and mean absolute error were considered when determining the quality of each split and between three and 14 features were examined when considering the best split. The optimized random forest model was then re-fit to the training data to validate the hyperparameter optimization. The RMSE of the random forest model decreased from 0.344 to 0.272 after optimization. The RMSE of the 1-hexene to 1-octene selectivities are 0.275 and 0.269, respectively.

The top of Figure 3 plots the selectivities determined from the optimized random model are plotted against those determined from the DFT selectivity model. In our data set, overall 1-hexene selective (i.e. >50% 1-hexene vs. 1-octene) is labeled as a negative value and overall 1-octene selective is labeled as a

positive value (i.e. >50% 1-octene vs. 1-hexene). The random forest model correctly predicted the overall 1-hexene versus 1-octene selectivity for 83 ligands and incorrectly predicted the overall selectivity for 22 ligands. This incorrect assignment occurs in cases where the DFT computed 1-hexene selectivity of a ligand is less than 1%. The random forest model tends to perform best for ligands ranging from 20:80 to 50:50 1-hexene:1-octene selectivity (See SI-Figure 1). Importantly, the majority of ligands in our data set are overall 1-octene selective, which is useful for our goal of improving the percentage of 1-octene, but it is likely that the RMSE of the random forest model would be reduced if the data set were more evenly distributed between overall 1-hexene and overall 1-octene selective.

In the bottom of Figure 3, the 1-octene selectivities calculated using the DFT selectivity model and the optimized random forest model are plotted against the experimentally determined selectivities. The DFT selectivity model agreed with experiment very well ($R^2 = 0.91$, mean absolute deviation = 4.4%). The DFT model underestimates the experimental selectivity, however, this is overall advantageous with the goal to increase 1-octene production. The lack of very high quantitative correlation between random forest and experiment values is likely due to the relatively small sample size of experimentally studied ligands. Despite this, the random forest can be used to determine critical chemical features that are responsible for enhancing 1-octene selectivity. The relative importance of the 14 features included in our dataset can be determined by replacing data with random values and observing the impact on the RMSE value. If replacing data of a feature with random values results in a small change to the RMSE then it has a low degree of importance. Conversely, if there is a large change in the RMSE it this feature has a large importance. Figure 4 displays this feature importance analysis using the optimized random forest model.
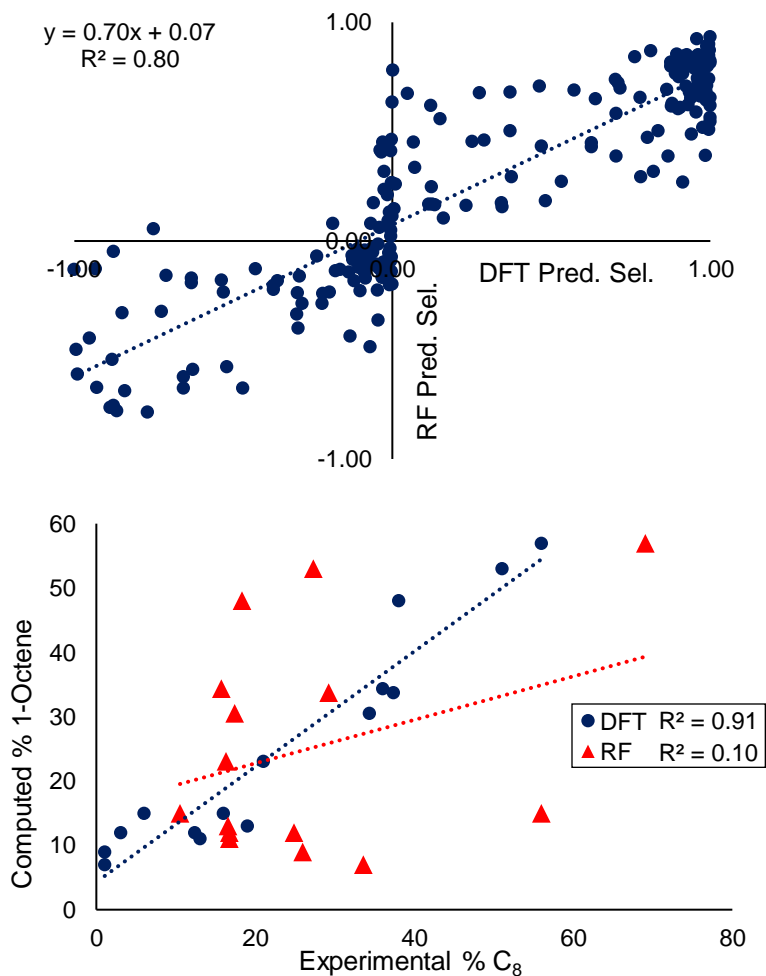
Figure 3. Top: Linear regression of selectivities predicted by DFT selectivity model (x-axis) and optimized random forest (RF) model (y-axis). Negative values correspond to high 1-hexene selectivity, positive values to high 1-octene selectivity. Bottom: Linear regressions of 1-octene selectivities predicted by DFT (circles) and random forest (triangles) models compared to experimental values.
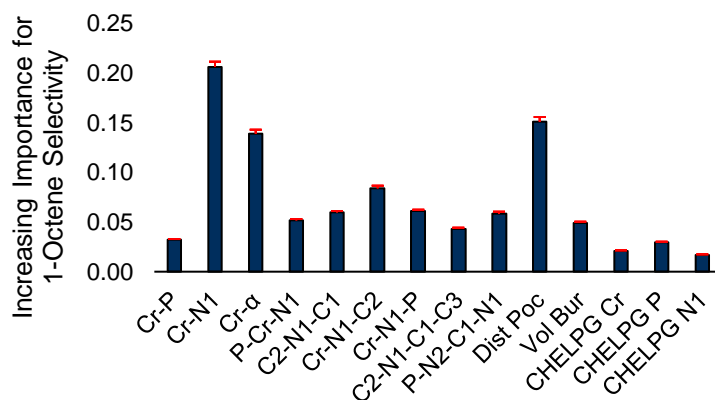
Figure 4. Normalized feature importance determined from random forest model with 95% confidence intervals (red bars).

Inspection of Figure 4 shows that the Cr–N distance, Cr–α distance, and distance out of pocket were identified as being most important in enhancing 1-octene selectivity. The Cr–N–C$_2$ distance, which is related to the Cr–α distance, was also identified as an important 1-octene enhancing feature. Interestingly, despite the proposed importance of the ligand bite angle, especially for phosphine catalysts,[105] we found that the P-Cr-N$_1$ ligand bite angle is among several lesser important features, which is consistent with Sydora previously showing that for Cr-phosphinoamidine catalyst ligands with similar bite angles resulted in significantly different 1-hexene:1-octene selectivities.[35] After we completed this machine learning analysis and identified the importance of the Cr-distance out of pocket for the 105 ligands examined here, Liu reported a DFT analysis of ethylene oligomerization by a Cr(2,2'-dipicolyamine) catalyst and also found that enhanced tetramerization was likely due to Cr-distance out of pocket was also important.

With the emergence of chemical features by the random forest machine learning model, we then turned to the final step in our workflow, which is using this information to virtually identify new catalyst ligands (Scheme 1c). Based on the important features, we very rapdily designed ligands **L9**-**L15** shown in Figure 5. The machine learning features directed us to change the 4-membered (P,N) ligand scaffold found in generations 1 (**L1**-**L5**) and 2 (**L6**-**L8**) to a 3-membered (P,N) ligand, which would potentially alter the Cr–N distance, Cr–α distance, and Cr distance out of ligand pocket. This machine-learning driven

modifcation led to the proposal of ligands **L9**-**L11**, and with our transition-state model they are predicted to be 97-99% 1-octene selective. With the very rapid success of this new generation 3 type of (P,N) ligands we further decreased the the ligand to have direct phosphine-nitrogen connection, which naturally led to the proposal of ligands **L12**-**L14**. Based on using our transition-state model, ligands **L12**-**L14** have predicted selectivities of >95% for 1-octene. Importantly, as plotted at the bottom of Figure 5, the use of our transition-state model combined with translation of machine learning features to new catalyst ligands, resulted in increasing the prediction of 1-octene from between <35% for generation 1 and ~50% for generation 2 to >95% for generation 3.
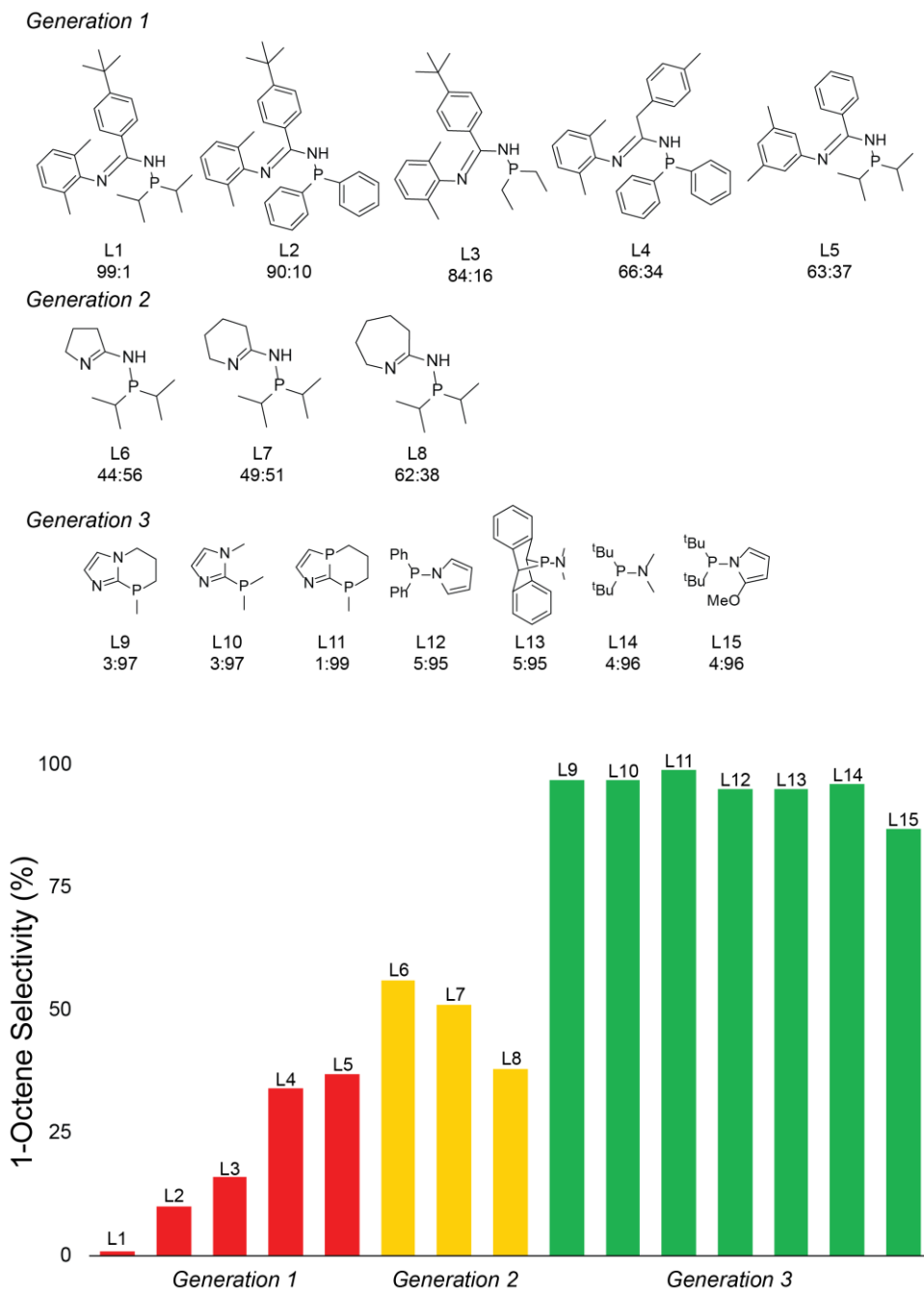
Figure 5. Top: Strucutres for previous (P,N) ligand generations and the new proposed ligands (generation 3) based on machine-learing identified features. The 1-hexene:1-octene selectivity (predicted) is given below each strucutre. Bottom: Plot of 1-octene selectivity for previous (P,N) ligand generations and new proposed ligands.

With the rapid success of designing ligands **L9**-**L15** it is clear that there are a number of new candidates that are now availible for experimental testing, and several more ligands can now be virtually

and rapidly designed. As one experimental confirmation of our results, subsequent to our desing of **L9-L15**, a literature search of all reported Cr-phosphine catlaysts for ethylene oligomerization revealed that Yang reported that **L12** is indeed highly 1-octene selective.[106] In this experimental test, which is signficantly different than the CPChem reaction conditions that our transition-state model was developed for, in situ catalyst formation with the combination of $Cr(acac)_3$, MAO, and **L12** and oligomerized under 50 bar of ethylene gave a 1-hexene:1-octene ratio of 28:70 with trace production of polyethylene.

Conclusions

For Cr(P,N) catalysed ethylene oligomerization, we combined our previously developed experimentally verified DFT-transition-state model with a random forest machine learning model. This workflow involved the calculation of transition-state 1-hexene:1-octene selectivity for 105 ligands and the harvesting of 14 descriptors, which were then used to build a random forest regression model with a low RMSE. This model revealed that Cr–N distance, Cr–α distance, and Cr distance out of pocket were key features for enhancing 1-octene selectivity. This then allowed the rapid design of several generation 3 Cr(P,N) catalyst ligands that are predicted to give >95% selectivity for 1-octene. Overall, this work demonstrated the utility of combining an accurate quantum-mechanical transition state model with machine learning to propel molecular catalyst design.

References

1. Ess, D. H., Gagliardi, L. & Hammes-Schiffer, S. Introduction: Computational Design of Catalysts from Molecules to Materials. *Chem. Rev.* **119**, 6507–6508 (2019).

2. Ahn, S., Hong, M., Sundararajan, M., Ess, D. H. & Baik, M.-H. Design and Optimization of Catalysts Based on Mechanistic Insights Derived from Quantum Chemical Reaction Modeling. *Chem. Rev.* **119**, 6509–6560 (2019).

3. Golab, J. T. Making industrial decisions with computational chemistry. *Chemtech* **28**, 17–23 (1998).

4. Allemann, C., Gordillo, R., Clemente, F. R., Cheong, P. H. Y. & Houk, K. N. Theory of asymmetric organocatalysis of aldol and related reactions: Rationalizations and predictions. *Acc. Chem. Res.* **37**, 558–569 (2004).

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

6. Tantillo, D. J. Using theory and experiment to discover catalysts for electrocyclizations. *Angew. Chemie - Int. Ed.* **48**, 31–32 (2009).

7. Sunoj, R. B. Proline-derived organocatalysis and synergism between theory and experiments. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**, 920–931 (2011).

8. Jover, J. & Fey, N. The computational road to better catalysts. *Chem. - An Asian J.* **9**, 1714–1723 (2014).

9. Nguyen, Q. N. N. & Tantillo, D. J. The many roles of quantum chemical predictions in synthetic organic chemistry. *Chem. - An Asian J.* **9**, 674–680 (2014).

10. Tsang, A. S. K., Sanhueza, I. A. & Schoenebeck, F. Combining Experimental and Computational Studies to Understand and Predict Reactivities of Relevance to Homogeneous Catalysis. *Chem. - A Eur. J.* **20**, 16432–16441 (2014).

11. Tantillo, D. J. Faster, Catalyst! React! React! Exploiting Computational Chemistry for Catalyst Development and Design. *Acc. Chem. Res.* **49**, 1079 (2016).

12. Sperger, T., Sanhueza, I. A. & Schoenebeck, F. Computation and Experiment: A Powerful Combination to Understand and Predict Reactivities. *Acc. Chem. Res.* **49**, 1311–1319 (2016).

13. Hammes-Schiffer, S. Catalysts by design: The power of theory. *Acc. Chem. Res.* **50**, 561–566 (2017).

14. Poree, C. & Schoenebeck, F. A holy grail in chemistry: Computational catalyst design: Feasible or fiction? *Acc. Chem. Res.* **50**, 605–608 (2017).

15. Kheirabadi, M. *et al.* Spiroligozymes for transesterifications: Design and relationship of structure to activity. *J. Am. Chem. Soc.* **134**, 18345–18353 (2012).

16. Doney, A. C., Rooks, B. J., Lu, T. & Wheeler, S. E. Design of organocatalysts for asymmetric propargylations through computational screening. *ACS Catal.* **6**, 7948–7955 (2016).

17. 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.* **125**, 6614–6615 (2003).

18.   Ianni, J. C., Annamalai, V., Phuan, P. W., Panda, M. & Kozlowski, M. C. A priori theoretical prediction of selectivity in asymmetric catalysis: Design of chiral catalysts by using quantum molecular interaction fields. *Angew. Chemie - Int. Ed.* **45**, 5502–5505 (2006).

19.   Mitsumori, S. *et al.* Direct asymmetric anti-Mannich-type reactions catalyzed by a designed amino acid. *J. Am. Chem. Soc.* **128**, 1040–1041 (2006).

20.   Jindal, G. & Sunoj, R. B. Rational design of catalysts for asymmetric diamination reaction using transition state modeling. *Org. Biomol. Chem.* **12**, 2745–2753 (2014).

21.   Wang, Y. *et al.* A computationally designed Rh(I)-catalyzed two-component [5+2+1] cycloaddition of ene-vinylcyclopropanes and CO for the synthesis of cyclooctenones. *J. Am. Chem. Soc.* **129**, 10060–10061 (2007).

22.   Donoghue, P. J., Helquist, P., Norrby, P. O. & Wiest, O. Prediction of enantioselectivity in rhodium catalyzed hydrogenations. *J. Am. Chem. Soc.* **131**, 410–411 (2009).

23.   Rowley, C. N. & Woo, T. K. Computational design of ruthenium hydride olefin-hydrogenation catalysts containing hemilabile ligands1,2. *Can. J. Chem.* **87**, 1030–1038 (2009).

24.   Baik, M. H. *et al.* Computationally designed and experimentally confirmed diastereoselective rhodium-catalyzed Pauson-Khand reaction at room temperature. *J. Am. Chem. Soc.* **133**, 7621–7623 (2011).

25.   Fernandez, L. E., Horvath, S. & Hammes-Schiffer, S. Theoretical design of molecular electrocatalysts with flexible pendant amines for hydrogen production and oxidation. *J. Phys. Chem. Lett.* **4**, 542–546 (2013).

26.   Nielsen, M. C., Bonney, K. J. & Schoenebeck, F. Computational ligand design for the reductive elimination of ArCF 3 from a small bite angle PdII complex: Remarkable effect of a perfluoroalkyl phosphine. *Angew. Chemie - Int. Ed.* **53**, 5903–5906 (2014).

27.   Bernales, V. *et al.* Computationally Guided Discovery of a Catalytic Cobalt-Decorated Metal-Organic Framework for Ethylene Dimerization. *J. Phys. Chem. C* **120**, 23576–23583 (2016).

28.   Tang, S. *et al.* 2D-QSPR/DFT studies of aryl-substituted PNP-Cr-based catalyst systems for highly selective ethylene oligomerization. *J. Mol. Model.* **20**, (2014).

29.   Karelson, M., Lobanov, V. S. & Katritzky, A. R. Quantum-chemical descriptors in QSAR/QSPR studies. *Chem. Rev.* **96**, 1027–1043 (1996).

30.   Kwon, D. H. *et al.* Computational Transition-State Design Provides Experimentally Verified Cr(P,N) Catalysts for Control of Ethylene Trimerization and Tetramerization. *ACS Catal.* **8**, 1138–1142 (2018).

31.   Hogan, P. J. & Banks, R. L. History of Crystalline Polypropylene. in *History of Polyolefins: The World's Most Widley Used Polymers* (eds. Seymour, R. B. & Cheng, T.) 103–115 (Springer Netherlands, 1986).

32.   Cecchin, G., Morini, G. & Piemontesi, F. Ziegler-Natta Catalysts. in *Kirk-Othmer Encyclopedia of Chemical Technology* (John Wiley and Sons).

33.   Vaughan, A., Davis, D. S. & Hagadorn, J. R. Industrial Catalysts for Alkene Polymerization. *Polym. Sci. A Compr. Ref. 10 Vol. Set* **3**, 657–672 (2012).

34.   Alferov, K. A., Belov, G. P. & Meng, Y. Chromium catalysts for selective ethylene oligomerization to 1-hexene and 1-octene: Recent results. *Applied Catalysis A: General* (2017)

doi:10.1016/j.apcata.2017.05.014.

35. Sydora, O. L. *et al.* Selective ethylene tri-/tetramerization catalysts. *ACS Catal.* **2**, 2452–2455 (2012).

36. 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.* **49**, 1292–1301 (2016).

37. Fey, N. *et al.* Development of a ligand knowledge base, Part 1: Computational descriptors for phosphorus donor ligands. *Chem. - A Eur. J.* **12**, 291–302 (2005).

38. Fey, N. *et al.* Computational descriptors for chelating P,P- And P,N-donor ligands. *Organometallics* **27**, 1372–1383 (2008).

39. Fey, N., Orpen, A. G. & Harvey, J. N. Building ligand knowledge bases for organometallic chemistry: Computational description of phosphorus(III)-donor ligands and the metal-phosphorus bond. *Coord. Chem. Rev.* **253**, 704–722 (2009).

40. Jover, J. *et al.* Expansion of the ligand knowledge base for monodentate P-donor ligands (LKB-P). *Organometallics* **29**, 6245–6258 (2010).

41. Jover, J. *et al.* Expansion of the ligand knowledge base for chelating P,P-donor ligands (LKB-PP). *Organometallics* **31**, 5302–5306 (2012).

42. Durand, D. J. & Fey, N. Computational Ligand Descriptors for Catalyst Design. *Chem. Rev.* **119**, 6561–6594 (2019).

43. Fey, N. *et al.* Stable fluorophosphines: Predicted and realized ligands for catalysis. *Angew. Chemie - Int. Ed.* **51**, 118–122 (2012).

44. Zahrt, A. F. *et al.* Prediction of higher-selectivity catalysts by computer-driven workflow and machine learning. *Science (80-.).* **363**, eaau5631 (2019).

45. Abdelfatah, K. E. *et al.* [ASAP] Prediction of Transition-State Energies of Hydrodeoxygenation Reactions on Transition-Metal Surfaces Based on Machine Learning. *J. Phys. Chem. C* **123**, 29804–29810 (2019).

46. Takahashi, K. & Miyazato, I. Rapid estimation of activation energy in heterogeneous catalytic reactions via machine learning. *J. Comput. Chem.* **39**, 2405–2408 (2018).

47. Singh, A. R., Rohr, B. A., Gauthier, J. A. & Nørskov, J. K. Predicting Chemical Reaction Barriers with a Machine Learning Model. *Catal. Letters* **149**, 2347–2354 (2019).

48. Arpe, H.-J. *Industrial Organic Chemistry*. (Wiley-VCH Verlag GmbH & Co, 2010).

49. Knudsen, R. D. *et al.* Methods of Preparation of an Olefin Oligomerization Catalyst. (2008).

50. Breuil, P.-A. R., Magna, L. & Olivier-Bourbigou, H. Role of Homogeneous Catalysis in Oligomerization of Olefins: Focus on Selected Examples Based on Grup 4 to Group 10 Transition Metal Complexes. *Catal. Letters* **145**, 173–192 (2015).

51. Britovsek, G. J. P. & McGuinness, D. S. A DFT Mechanistic Study on Ethylene Tri- and Tetramerization with Cr/PNP Catalysts: Single versus Double Insertion Pathways. *Chem. - A Eur. J.* **22**, 16891–16896 (2016).

52. Britovsek, G. J. P., McGuinness, D. S. & Tomov, A. K. Mechanistic study of ethylene tri- and tetramerisation with Cr/PNP catalysts: Effects of additional donors. *Catal. Sci. Technol.* **6**, 8234–

8241 (2016).

53. Hossain, M. A., Kim, H. S., Houk, K. N. & Cheong, M. Spin-crossover in chromium-catalyzed ethylene trimerization: Density functional theory study. *Bull. Korean Chem. Soc.* **35**, 2835–2838 (2014).

54. Yu, Z. X. & Houk, K. N. Why trimerization? Computational elucidation of the origin of selective trimerization of ethene catalyzed by [TaCl3(CH3)2] and an agostic-assisted hydride transfer mechanism. *Angew. Chemie - Int. Ed.* **42**, 808–811 (2003).

55. Gong, M. *et al.* Selective Co-Oligomerization of Ethylene and 1-Hexene by Chromium-PNP Catalysts: A DFT Study. *Organometallics* **35**, 972–981 (2016).

56. Yang, Y., Liu, Z., Cheng, R., He, X. & Liu, B. Mechanistic DFT study on ethylene trimerization of chromium catalysts supported by a versatile pyrrole ligand system. *Organometallics* **33**, 2599–2607 (2014).

57. Qi, Y. *et al.* Role of 1,2-dimethoxyethane in the transformation from ethylene polymerization to trimerization using chromium Tris(2-ethylhexanoate)-based catalyst system: A DFT study. *Organometallics* **29**, 1588–1602 (2010).

58. Budzelaar, P. H. M. Ethene trimerization at Cr1/Cr3 - A density functional theory (DFT) study1. *Can. J. Chem.* **87**, 832–837 (2009).

59. Bhaduri, S., Mukhopadhyay, S. & Kulkarni, S. A. Density functional studies on chromium catalyzed ethylene trimerization. *J. Organomet. Chem.* **694**, 1297–1307 (2009).

60. Van Rensburg, W. J. *et al.* A DFT Study toward the Mechanism of Chromium-Catalyzed Ethylene Trimerization. *Organometallics* **23**, 1207–1222 (2004).

61. Bartlett, S. A. *et al.* Activation of [CrCl3{R-SN(H)S-R}] catalysts for selective trimerization of ethene: A freeze-quench Cr K-edge XAFS study. *ACS Catal.* **4**, 4201–4204 (2014).

62. Agapie, T. Selective ethylene oligomerization: Recent advances in chromium catalysis and mechanistic investigations. *Coord. Chem. Rev.* **255**, 861–880 (2011).

63. Kayala, M. A., Azencott, C. A., Chen, J. H. & Baldi, P. Learning to predict chemical reactions. *J. Chem. Inf. Model.* **51**, 2209–2222 (2011).

64. Wei, J. N., Duvenaud, D. & Aspuru-Guzik, A. Neural networks for the prediction of organic chemistry reactions. *ACS Cent. Sci.* **2**, 725–732 (2016).

65. Coley, C. W., Green, W. H. & Jensen, K. F. Machine Learning in Computer-Aided Synthesis Planning. *Acc. Chem. Res.* **51**, 1281–1289 (2018).

66. Liu, B. *et al.* Retrosynthetic Reaction Prediction Using Neural Sequence-to-Sequence Models. *ACS Cent. Sci.* **3**, 1103–1113 (2017).

67. Coley, C. W., Barzilay, R., Jaakkola, T. S., Green, W. H. & Jensen, K. F. Prediction of Organic Reaction Outcomes Using Machine Learning. *ACS Cent. Sci.* **3**, 434–443 (2017).

68. 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 (80-.).* **360**, 186–190 (2018).

69. Sun, W. *et al.* The Use of Deep Learning to Fast Evaluate Organic Photovoltaic Materials. *Adv. Theory Simulations* **2**, 1800116 (2019).

70.   Coley, C. W. *et al.* A graph-convolutional neural network model for the prediction of chemical reactivity. *Chem. Sci.* **10**, 370–377 (2019).

71.   Tomberg, A., Johansson, M. J. & Norrby, P. O. A Predictive Tool for Electrophilic Aromatic Substitutions Using Machine Learning. *J. Org. Chem.* **84**, 4695–4703 (2019).

72.   Zahrt, A. F. *et al.* Prediction of higher-selectivity catalysts by computer-driven workflow and machine learning. *Science (80-.).* **363**, (2019).

73.   Gao, H. *et al.* Using Machine Learning to Predict Suitable Conditions for Organic Reactions. *ACS Cent. Sci.* **4**, 1465–1476 (2018).

74.   Back, S., Tran, K. & Ulissi, Z. W. Toward a Design of Active Oxygen Evolution Catalysts: Insights from Automated Density Functional Theory Calculations and Machine Learning. *ACS Catal.* **9**, 7651–7659 (2019).

75.   Chowdhury, A. J. *et al.* Prediction of Adsorption Energies for Chemical Species on Metal Catalyst Surfaces Using Machine Learning. *J. Phys. Chem. C* **122**, 28142–28150 (2018).

76.   Goldsmith, B. R., Esterhuizen, J., Liu, J. X., Bartel, C. J. & Sutton, C. Machine learning for heterogeneous catalyst design and discovery. *AIChE J.* **64**, 2311–2323 (2018).

77.   Guo, X. *et al.* Simultaneously Achieving High Activity and Selectivity toward Two-Electron O2 Electroreduction: The Power of Single-Atom Catalysts. *ACS Catal.* **9**, 11042–11054 (2019).

78.   Han, X. *et al.* Comparison of Machine Learning Algorithms in Screening Potential Additives to Ni/Al2O3 Methanation Catalysts for Improving the Anti-Coking Performance. *ChemistrySelect* **4**, 11790–11795 (2019).

79.   Jinnouchi, R. & Asahi, R. Predicting Catalytic Activity of Nanoparticles by a DFT-Aided Machine-Learning Algorithm. *J. Phys. Chem. Lett.* **8**, 4279–4283 (2017).

80.   Ohyama, J., Nishimura, S. & Takahashi, K. Data Driven Determination of Reaction Conditions in Oxidative Coupling of Methane via Machine Learning. *ChemCatChem* **11**, 4307–4313 (2019).

81.   Palkovits, R. & Palkovits, S. Using Artificial Intelligence to Forecast Water Oxidation Catalysts. *ACS Catal.* **9**, 8383–8387 (2019).

82.   Takahashi, K., Miyazato, I., Nishimura, S. & Ohyama, J. Unveiling Hidden Catalysts for the Oxidative Coupling of Methane based on Combining Machine Learning with Literature Data. *ChemCatChem* **10**, 3223–3228 (2018).

83.   Takasao, G. *et al.* Machine Learning-Aided Structure Determination for TiCl 4 -Capped MgCl 2 Nanoplate of Heterogeneous Ziegler-Natta Catalyst. *ACS Catal.* **9**, 2599–2609 (2019).

84.   Ulissi, Z. W. *et al.* Machine-learning methods enable exhaustive searches for active Bimetallic facets and reveal active site motifs for CO2 reduction. *ACS Catal.* **7**, 6600–6608 (2017).

85.   Nandy, A. *et al.* Machine Learning Accelerates the Discovery of Design Rules and Exceptions in Stable Metal-Oxo Intermediate Formation. *ACS Catal.* **9**, 8243–8255 (2019).

86.   Zhu, X. *et al.* Activity Origin and Design Principles for Oxygen Reduction on Dual- Metal-Site Catalysts: A Combined Density Functional Theory and Machine Learning Study. (2019) doi:10.1021/acs.jpclett.9b03392.

87.   Landrum, G. A., Penzotti, J. E. & Putta, S. Machine-learning models for combinatorial catalyst discovery. *Meas. Sci. Technol.* **16**, 270–277 (2005).

88.   Rizkin, B. A. & Hartman, R. L. Supervised machine learning for prediction of zirconocene-catalyzed α-olefin polymerization. *Chem. Eng. Sci.* **210**, 115224 (2019).

89.   Janet, J. P. & Kulik, H. J. Predicting electronic structure properties of transition metal complexes with neural networks. *Chem. Sci.* **8**, 5137–5152 (2017).

90.   Janet, J. P. & Kulik, H. J. Resolving Transition Metal Chemical Space: Feature Selection for Machine Learning and Structure-Property Relationships. *J. Phys. Chem. A* **121**, 8939–8954 (2017).

91.   Meyer, B., Sawatlon, B., Heinen, S., Von Lilienfeld, O. A. & Corminboeuf, C. Machine learning meets volcano plots: Computational discovery of cross-coupling catalysts. *Chem. Sci.* **9**, 7069–7077 (2018).

92.   Banerjee, S., Sreenithya, A. & Sunoj, R. B. Machine learning for predicting product distributions in catalytic regioselective reactions. *Phys. Chem. Chem. Phys.* **20**, 18311–18318 (2018).

93.   Mansouri Tehrani, A. *et al.* Machine Learning Directed Search for Ultraincompressible, Superhard Materials. *J. Am. Chem. Soc.* **140**, 9844–9853 (2018).

94.   Gu, G. X., Chen, C. T. & Buehler, M. J. De novo composite design based on machine learning algorithm. *Extrem. Mech. Lett.* **18**, 19–28 (2018).

95.   Li, Z., Ma, X. & Xin, H. Feature engineering of machine-learning chemisorption models for catalyst design. *Catal. Today* **280**, 232–238 (2017).

96.   Zhao, Y. & Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other function. *Theor. Chem. Acc.* **120**, 215–241 (2008).

97.   Marenich, A. V., Cramer, C. J. & Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **113**, 6378–6396 (2009).

98.   Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenb, D. J. Gaussian 09 Revision D.01. (2009).

99.   Jones, E., Oliphant, T. & Peterson, P. SciPy: Open source scientific tools for Python. (2001).

100.   Oliphant, T. *A Guide to NumPy*. (Trelgol Publishing, 2006).

101.   McKinney, W. Data Structures for Statistical Computing in Python. in *Proceedings of the 9th Python in Science Conference* (eds. van der Walt, S. & Millman, J.) 51–56 (2010).

102.   Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

103.   Falivene, L. *et al.* Towards the online computer-aided design of catalytic pockets. *Nat. Chem.* **11**, 872–879 (2019).

104.   Palmer, D. S., O'Boyle, N. M., Glen, R. C. & Mitchell, J. B. O. Random forest models to predict aqueous solubility. *J. Chem. Inf. Model.* **47**, 150–158 (2007).

105.   Kim, S.-K. *et al.* Bimetallic ethylene tetramerization catalysts derived from chiral DPPDME

ligands: syntheses, structural characterizations, and catalytic performance of [(DPPDME)CrCl3]2 (DPPDME = S,S- and R,R-chiraphos and mesophos). *Organometallics* **29**, 5805–5811 (2010).

106. Yang, Y., Liu, Z., Liu, B. & Duchateau, R. Selective ethylene tri-/tetramerization by in situ-formed chromium catalysts stabilized by N,P-based ancillary ligand systems. *ACS Catal.* **3**, 2353–2361 (2013).