The importance of reaction energy in predicting chemical reaction barriers with machine learning models

Nithin Lalith¹, Aayush R. Singh²*, Joseph A. Gauthier¹*

¹Department of Chemical Engineering, Texas Tech University, Lubbock, TX 79409, USA
²Dow Inc., Midland, MI 48686, USA

* corresponding author email address:
arsingh@dow.com
joe.gauthier@ttu.edu

Abstract:
Improving our fundamental understanding of complex heterocatalytic processes increasingly relies on electronic structure simulations and microkinetic models based on calculated energy differences. In particular, calculation of activation barriers, usually achieved through compute-intensive saddle point search routines, remains a serious bottleneck in understanding trends in catalytic activity for highly branched reaction networks. Although the well-known Brønsted-Evans-Polyani (BEP) scaling – a one-dimensional linear regression model – has been widely applied in such microkinetic models, they still rely on calculated reaction energies and may not generalize beyond a single facet on a single class of materials, e.g., a terrace sites on transition metals. For highly branched and energetically shallow reaction networks, such as electrochemical CO₂ reduction or waste remediation, calculating even reaction energies on many surfaces can become computationally intractable due to the combinatorial explosion of states that must be considered. Here, we investigate the feasibility of activation barrier prediction without knowledge of the reaction energy using linear and nonlinear machine learning (ML) models trained on a new database of over 500 dehydrogenation activation barriers. We and find that inclusion of the reaction energy significantly improves both classes of ML models, but complex nonlinear models can achieve performance similar to the simplest BEP scaling when predicting activation barriers on new systems. Additionally, inclusion of the reaction energy significantly improves generalizability to new systems beyond the training set. Our results suggest that the reaction energy is a critical feature to consider when building models to predict activation barriers, indicating that efforts to reliably predict reaction energies reliably through, e.g., the Open Catalyst Project and others, will be an important route to effective model development for more complex systems.
Introduction

Developing a more fundamental understanding of catalytic processes at metal surfaces, and therefore improving capabilities to discover new and improved catalysts for desired processes, increasingly relies on computer simulations such as those enabled by density functional theory (DFT).\textsuperscript{1-9} Although methods have advanced considerably over the past few decades, considerable effort is still required to bridge the so-called computation-experiment gap.\textsuperscript{10} Current practices in utilizing DFT to probe a catalytic process typically involves generation of a microkinetic model (MKM),\textsuperscript{11} which is a system rate equations derived from power-law expressions. Parameters in the models include those set by the experiment (pressures of reactants and products, temperature, etc.) and those determined by the system (molecular binding energies on the catalyst, activation barriers of elementary sub-reactions, …) and are computed using, e.g., DFT. As such, a range of calculation types are required to generate a MKM: (i) determination of energetic minima, achieved by minimizing the total DFT energy with respect to atomic positions, (ii) estimation of free-energy corrections through, e.g., ideal gas or harmonic oscillator approximations (or more complex estimations),\textsuperscript{12-16} which typically requires at least estimation of the Hessian matrix, (iii) determination of activation barriers for all elementary reaction steps in the pathway, which typically involves a saddle point search by one of several algorithms.

Activation barrier calculations in particular remain a significant challenge in theoretical investigations into a particular catalytic process, as each method for localizing a transition state (TS) comes with serious drawbacks.\textsuperscript{17, 18} Surface crawling algorithms such as the dimer method\textsuperscript{19-21} are attractive since only 2-3 DFT force calls are required for each update of the TS geometry. These methods can also quickly converge provided an excellent initial guess of both the geometry and reaction coordinate. However, they are not guaranteed to converge to a TS for the reaction of interest, and may instead converge to an undesired TS. Alternatively, interpolation schemes such as the nudged elastic band (NEB) and its variants\textsuperscript{18, 22-27} are much more likely to converge to the correct TS (if it converges), but the computational cost scales linearly with the number of images in the interpolation. Convergence typically improves as the number of images in the interpolation increases, and a typical NEB for a surface reaction may involve 5-10 images, leading to a minimum 5-10x increase in computational cost in comparison to a standard geometry optimization. In reality, interpolation methods typically are more than 5-10x as expensive, since the number of optimization steps can be significantly higher than that of a typical energetic minimum localization if the initial interpolation is poor.

As such, several groups have begun applying machine learning (ML) concepts to reduce the computational cost of TS localization. Surrogate models, which train a ML model based on DFT calculations ‘on-the-fly’ in particular have seen success in this space. For example, Garrido Torres and coworkers developed a surrogate Gaussian process regression (GPR) model to accelerate NEB calculations.\textsuperscript{28} A major advantage to a GPR as a surrogate is that model uncertainty can be used to inform the next point in the training process; however, if the approach fails to converge quickly, evaluating the GPR can become nearly as computationally intensive as the DFT itself. As another example, Ulissi and coworkers have developed a ‘FINETUNA’ package,\textsuperscript{29} which allows for tuning of vast ML potentials derived from the Open Catalyst project.\textsuperscript{30, 31} As of writing, this package is only available for geometry optimizations, with its ability to...
predict transition states an area of active work. Although not directly integrated, a tuned checkpoint from the OC2020 or OC2022 database could alternatively serve as a calculator directly within e.g. the Atomic Simulation Environment, allowing for rapid TS estimation with minimal evaluations of DFT.

With a growing emphasis on capturing the complexity of catalyst surfaces and ever-expanding reaction networks, merely reducing the number of DFT calls may still prove to be too computationally demanding. Efforts to predict activation barriers without ever localizing a transition state therefore may prove useful for investigations into complex reaction networks over a range of materials, e.g. bimetallic alloys or more complex materials design spaces. Given the combinatorial explosion of reactions that must be considered as the numbers of adsorbates + materials increases, ideally such a model would predict an activation barrier without a DFT calculation for each adsorbate/material pair, but perhaps a DFT calculation for each adsorbate and material separately – in other words, neglecting the commonly used reaction energy as a descriptor. For example, a reduced reaction network for CO hydrogenation will have on the order of 100 unique species, 200 reactions, and more than 2,000 possible pathways. Screening this reaction across e.g. 100 different bimetallic alloys and their various low surface-energy facets would quickly become intractable even for a ML model, provided it relies on the reaction energy as an input, as it would require at a minimum of two DFT calls for each combination of reaction + material.

In practice, BEP type relations have been utilized to avoid this problem for a limited catalyst design space. Rather than predicting an activation barrier directly, instead the target of the ML model is to predict binding energies of adsorbates on a given complex surface. The resulting reaction energy is then mapped to an activation barrier using a one-feature linear regression model (i.e., the BEP scaling) fit across a range of transition metals. However, for a large enough intermediate and catalyst design space, even such a BEP scaling will struggle to appropriately scale, as it will be limited by the calculation of reaction energies using DFT. In the molecular catalysis literature, some approaches have been developed to predict transition state energetics without any DFT input, instead making predictions based only on properties of the reactant and product species. Alternatively, generalized or foundational machine learning potentials can be used to minimize DFT requirements for computing the reaction energy, but the extent of error propagation (and hence reliability for catalyst discovery) in such an approach is not well understood.

In light of these challenges and aforementioned efforts in molecular catalysis, here we investigate the importance of the reaction energy in predicting activation barriers with ML approaches, with the aim of eliminating the need for (i) ML prediction of reaction energies, and (ii) transition state localization. We find that, given a robust descriptor space consisting only of quantities that do not necessitate a DFT calculation in each state involving the surface – i.e., one calculation per surface and one per molecule, but not one for each combination – achieving a similar performance to models that include reaction energy requires models with significantly more fit parameters and reduced interpretability. In particular, we find that linear regression models, when excluding reaction energy as a feature, fail to achieve the mean absolute error (MAE) of a one-feature linear regression model (i.e., BEP scaling). Excluding the reaction energy, our forward search analysis reveals that the d-band center of the metal is the most effective...
descriptor for predicting activation barriers, and does not involve a calculation for each combination of molecule + metal considered. Furthermore, we find that inclusion of the reaction energy can, in some circumstances, significantly improve model generalizability, even when using complex nonlinear models. Our results here highlight some of the challenges the field will face as the push to go beyond single component transition metals manifests: calculating the reaction energy as a means for predicting activation barriers may prove to be a difficult bar to overcome. Although our estimated activation barriers are likely not useful as predictors for the genuine barriers, they may be useful for excluding unlikely pathways for further analysis with more sophisticated approaches.

**Methods**

**Dataset development:**

Our dataset consists of 565 activation barriers for dehydrogenation reactions, which were computed using QUANTUM ESPRESSO, an open-source planewave density functional theory (DFT) code. Core electrons were described via GBRV ultrasoft pseudopotentials, while valence electrons were expanded as planewaves up to a kinetic energy cutoff of 600 eV and a charge density cutoff of 6000 eV. Exchange and correlation effects were modeled at the generalized gradient approximation level via the Bayesian Error Estimation Functional (BEEF-vdW). All geometry optimizations were performed with a force cutoff of 0.05 eV/Å using a Quasi-Newton optimizer as implemented in the Atomic Simulation Environment (ASE).32

![Figure 1: Table A shows transition metals, highlighted in yellow, used as heterogeneous catalysts in our dataset development. For each metal catalyst, between 31 and 50 dehydrogenation reactions were simulated using different molecules; a few examples are shown here.](https://doi.org/10.26434/chemrxiv-2023-t6zyj-v2)
Transition state geometries and energetics were estimated using a one-dimensional fixed bond-length scan (1D-FBL). Here, the dissociating hydrogen and the atom to which it is bound in the initial state (i.e., before dehydrogenation) is optimized across a sequence of increasing H—X bond lengths, with the bond length constrained during optimization. The highest energy image in this sequence along the reaction path is taken to be the transition state. Although this method is not guaranteed to yield a true stationary point (i.e., the vibrational modes may not have a single imaginary mode), our benchmarking reveals that the energy of the ‘transition state’ determined by the 1D-FBL agrees very closely with that of the true saddle point – see below for benchmarking. This approach was utilized because it enables rapid and accurate estimation of the activation barrier for reaction paths that are reasonably well-described by a 1D projection of the reaction coordinate, for example, dehydrogenation reactions.

The 1D-FBL method was benchmarked for accuracy by determining true saddle points for a randomly selected ~17% subset (96 barriers) of the aforementioned dataset. Here, the transition state search utilized a machine learning accelerated nudged elastic band routine. Figure 2 below illustrates the high level of parity between the methods, suggesting the one-dimensional projection is reasonable for this particular dataset.

Figure 2: Parity plot benchmarking our 1D-FBL transition state search method with the more rigorous ML accelerated NEB method.

Figure 1 also demonstrates the effectiveness of each method. While most points lie near the parity line, a handful of cases are quite far from parity. Despite generally being believed to be a cruder method of estimating activation barriers, the FBL method calculated significantly lower barriers for six reactions. Some notable outliers were the reactions for CH\textsubscript{2}NHCH\textsubscript{2} on Fe and C\textsubscript{2}H\textsubscript{2} on Au. For points below the parity line, the MLNEB method found lower activation barriers. The reaction NH\textsubscript{3} dehydrogenation on Ir
was the most extreme example of the NEB method finding a lower energy pathway. In total, of the 565 reactions present in the FBL dataset, 96 reactions converged via MLNEB and were tested for parity. Further details related to dataset generation and refinement can be found in Supporting Information (SI) Note 1.

**Machine learning:**
Using machine learning (ML) analysis and the refined FBL dataset, we created models that predict the activation barrier of dehydrogenation reactions. These models provide insight into which features of the reaction of interest most reliably predict the activation barrier. The advantage of such ML models compared to traditional methods of determining activation barriers (e.g., FBL, NEB, or MLNEB), once fit, is that they are considerably less computationally expensive. Depending on the form of the model chosen, little to no DFT calculations are needed to determine the activation barrier. The well-known BEP scaling, which utilizes the reaction energy to create a one-feature linear regression model, requires two geometry optimization calculations (i.e., the initial and final state of the reaction). In addition to the reaction energy, we considered 11 features based either on the catalyst or molecule, tabulated below in Table 1 below.

<table>
<thead>
<tr>
<th>Class</th>
<th>Feature</th>
<th>Notation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalyst</td>
<td>sp-band filling</td>
<td>$f_{sp}$</td>
</tr>
<tr>
<td></td>
<td>sp-band width</td>
<td>$w_{sp}$</td>
</tr>
<tr>
<td></td>
<td>d-band filling</td>
<td>$f_d$</td>
</tr>
<tr>
<td></td>
<td>d-band center</td>
<td>$\varepsilon_d$</td>
</tr>
<tr>
<td></td>
<td>Electronegativity</td>
<td>$\chi_{cat}$</td>
</tr>
<tr>
<td></td>
<td>Coupling Matrix</td>
<td>$V$</td>
</tr>
<tr>
<td></td>
<td>Work Function</td>
<td>$W$</td>
</tr>
<tr>
<td>Adsorbate</td>
<td>Electronegativity</td>
<td>$\chi_{het}$</td>
</tr>
<tr>
<td></td>
<td>Bond Length</td>
<td>$L_b$</td>
</tr>
<tr>
<td></td>
<td>pKa</td>
<td>$pK_a$</td>
</tr>
<tr>
<td>Dynamic</td>
<td>Electronegativity difference</td>
<td>$\Delta \chi$</td>
</tr>
<tr>
<td></td>
<td>Reaction energy</td>
<td>$\Delta E_{rxn}$</td>
</tr>
</tbody>
</table>

Of particular interest in this work is to investigate models that do not utilize reaction energy, as this feature requires at least one DFT calculations for each reactant-metal combination. Instead, the other features tabulated below rely on just one bulk metal optimization per catalyst in addition to tabulated properties of the molecules and...
metals. These features generally describe the reactivity of the adsorbate, especially when it comes to dehydrogenation. While the pKa for many of the compounds we simulated were tabulated, others had to be estimated using a similar structure; for example, we estimated the pKa of H₂CCO as 13.3, the same as a formaldehyde (H₂CO). For each metal, we considered seven different features to describe its reactivity. Features chosen included the d-band center, d-band filling, coupling matrix, among others. Finally for each dehydrogenation reaction, we included two dynamic features, or features which are exclusive to a particular reaction. These were reaction energy and the difference in electronegativity between the metal and heteroatom. We expected these dynamic features to be important to the machine learning models, particularly the reaction energy which has long been known to correlate with activation barriers. In some models, where noted, we also included binary combinations of the aforementioned features in our models. These features are the product of two other features in the reaction set (e.g., pKa*electronegativity). Binary interactions helped to improve the dynamics of our models and improve their parity. In total, for each datum in our dataset, there were a total of 78 features. A full database can be found as a supporting Microsoft Excel spreadsheet with this publication on our group Github page.

To conduct our ML analysis, we first split the FBL dataset into three subsets: training, validation, and test set, using a 60-20-20 percent split respectively. Each point in each subset consists of a list of all features (including binary combinations of features) as well as the target activation barrier determined by the FBL calculation. The training data was used to train parameters of the model, while the validation set was used to determine any model hyperparameters and in forward search analyses. Finally, we used the test set to determine the MAE of our best performing models for inter-model comparison. We report four models in our results below: (i) a multi-feature linear regression model excluding binary combinations of features, (ii) a multi-feature linear regression model including binary combinations of features, (iii) a regularized multi-feature linear regression model including binary combinations, and (iv) a random forest model including binary combinations of features. The multi-feature linear regression model is described by Eq. 1 below. Here, the output of the model y represents an activation barrier in units of eV. For each feature xᵢ there is an associated coefficient βᵢ.

\[ y = \beta_0 + \sum \beta_i x_i \]  

The final model we constructed was a random forest (RF) regression. This is a non-linear model; while it performs well, it lacks facile interpretability compared to linear regression models. Using the RandomForestRegressor class in the sci-kit learn package, we trained an RF model to predict activation barriers. For this model we trained three different hyperparameters: maximum tree depth, minimum samples per leaf and minimum samples per split. These hyperparameters describe the complexity of the regression and can contribute to underfitting or overfitting the model.

**Results and Discussion**
With the dataset created, several models were constructed to analyze the importance of the features detailed above. Ultimately, our goal is to build a model that performs at least as well as a one-feature linear regression model, i.e. a typical BEP scaling, but without including the reaction energy, since reaction energy calculations require at least two DFT geometry optimizations for each molecule—metal facet pair. As the number of metals, facets, and molecules grows, even simple BEP analysis can quickly become intractable due to the cost of computing reaction energetics. Below we begin by analyzing linear models which are attractive for their high degree of interpretability, before moving to a non-linear random forest model.

Linear regression model fitting:
Single- and multi-feature linear regression models were used to conduct a forward search, which quantifies the importance of each feature by analyzing the magnitude of the coefficient \( \beta_i \) for each feature \( x_i \). In order to create the best performing multi-feature linear regression, we initially constructed one-feature linear models fitted to the training reaction set. These trained models are then tested on the validation set; the regression model with the lowest resulting MAE is the first feature \( x_1 \) in the forward search analysis and is added to our multi-feature model. To determine the next feature to include, we compared two-feature models of each remaining feature and add the one that results the lowest validation MAE. We continued this procedure until all features are present in the model. The order in which features are added to this model and the absolute value of the \( \beta \) parameter signifies the strength of correlation between a given feature and the activation barrier. The results of our forward search analysis are shown in Figure 3.

Previous work has demonstrated that reaction energy significantly correlates with activation barrier. Using a simple linear BEP model, only utilizing the reaction energy, results in a MAE of approximately 0.28 eV on the validation set. Adding more features to the linear model, using a forward search, does help to reduce the validation error, although not significantly. Besides reaction energy, the electronegativity difference \( \Delta \chi \) and \( pK_a \) seem to have the strongest correlation with the activation barrier. Our analysis here suggests that, in this model framework, additional features have little predictive power for the activation barrier.

Given the aim to minimize total DFT requirements to predict new activation barriers, we additionally created a model that does not require determination of the reaction energy. As such, we performed a forward search without including reaction energy, with the results illustrated in Figure 3. In this case, we found that different features most reliably predict activation barriers. While the validation MAE for this is never lower than the BEP scaling, we note that the number of features at which point the validation MAE increases is significantly higher than the case of including reaction energy as a feature. This result suggests that these features (e.g., metal d-band center \( \epsilon_d \)) correlates not only the activation barrier, but also the reaction energy, as is well-known.

Including binary combinations of features in the forward search, illustrated in Figure 3 (C) we see only a modest \( \sim 0.06 \) eV reduction in the validation MAE compared to the BEP model, when including reaction energy. As in the case of excluding binary combinations of features, the forward search was also performed on a reduced feature set, not including reaction energy or any of its binary interactions, shown in Figure 3 (D).
While we once again observed improved parity over the exclusion of binary combinations of features, this reduced feature model ultimately performs worse than the BEP model, as denoted by the horizontal red line.

![Graphs showing mean absolute error for different numbers of features.](https://doi.org/10.26434/chemrxiv-2023-t6zyv2)

Figure 3: Forward search analysis for multi-feature linear regression models excluding (A,B) and including (C,D) binary interactions. In panels (A, C) the reaction energy is included as a possible feature while panels (B, D) exclude reaction energy. In both cases, the horizontal red line corresponds to the MAE of a linear regression model using the reaction energy as the only feature (i.e., a BEP-type model).

The forward search procedure used to construct our models helped us to determine the relative importance of binary interactions. For the complete feature set, we observed that reaction energy binary interactions took priority in both the linear and non-linear models. Similarly, we observed the d-band filling and d-band center binary interactions to be important in or linear and non-linear models for the reduced feature set. Both these observations reasonably track with the results of the forward search of the non-binary feature set demonstrated. Ultimately, we can conclude that binary interactions improved the parity of our models, even if the difference is slight.

We note here that we deliberately neglected to include more complex features in our analysis of linear models on this dataset, such as those that might be generated by a SISSO-type construction.\(^{45, 66}\) Although such an analysis is certainly interesting and useful for many applications, here our goal is to build a model that simultaneously minimizes DFT requirements and, ideally, also generalizes beyond dehydrogenation.
Although domain generalization in ML models is an area of active research, early results seem to suggest that more complex models tend to exhibit poorer domain generalization,\textsuperscript{67} in agreement with the well-known bias-variance tradeoff,\textsuperscript{68} and so we do not include more complex features in our linear regression models. Further discussion of regularization of linear regression models can be found in SI Note 2.

Non-linear model fitting and comparison to the BEP model:
Finally, we constructed non-linear models using random forest (RF) regression. Using a similar forward search procedure to the one used to train the $\alpha$ hyperparameter in ridge regression, we constructed the best possible RF models for this dataset, both including and excluding reaction energy and its binary interactions in the feature set. The cross-validation MAE and its dependence on number of leaves and splits is discussed in greater detail in SI Note 3. Compared to its linear BEP counterpart, the reduced-feature RF model had significantly better performance predicting the activation barriers in our validation set, with a MAE of approximately 0.25 eV. This is a significant result, since it suggests that activation barriers can be somewhat reliably predicted without any DFT calculation as an input. RF performance also noticeably improved with inclusion of the reaction energy as a feature, with our best-performing model predicting activation barriers with a MAE of 0.16 eV on the cross-validation set.

Using the lowest validation MAE, we determined the optimum hyperparameters and features to include for our linear and non-linear models. We used the test set to report their final performance and MAE; the results can be seen in Figure 4.

**Figure 4:** Parity plot showing the activation barriers in our test dataset calculated via the FBL method compared to the values calculated by each of our models. For each of the models, parameters and hyperparameters were trained using the training and validation reaction dataset.

In general, we find that the addition of more features, beyond reaction energy, only modestly improves upon BEP scaling for our linear models when compared using the test set. Similarly, while our RF model excluding reaction energy had a lower MAE than the BEP model on the validation set, we find comparable performance on the test set,
suggesting that prediction of activation barriers with minimal (or even no) DFT is possible when using a sufficiently complex model.

We note two significant caveats with these results. First, all data in this dataset take place on terraces, and so the effect of facet is not captured in this dataset. We hypothesize that, if the dataset were to be expanded to include multiple facets (e.g., steps, kinks, defects, ...), then inclusion of features such as the coordination number of the active site would lead to significant performance enhancements of the multi-feature models relative to the BEP model, as was found in our prior work. As such, the BEP performance in Figure 4 can be thought of as an upper bound on its capabilities, and we suspect that the multi-feature linear model and nonlinear models would perform better relative to the BEP model with a more complex dataset. Second, we caution readers against interpreting the test MAE values too closely. The 60-20-20 split was performed randomly, and as such there is some intrinsic uncertainty in the value of the final test MAEs reported in Figure 4. If the fitting, validation, and testing procedure were to be repeated many times in e.g., a bootstrapping protocol, confidence intervals on the test MAE would likely suggest that the BEP model, multi-feature linear model with reaction energy, and RF model without reaction energy, all have approximately similar performance. However, we again highlight the importance of such a result, as the RF model without reaction energy does not rely on a DFT calculation for each barrier – only a DFT calculation on the bulk metal to determine electronic structure parameters such as d-band center, filling, etc. Such a model, if generalizable to new chemistries, would vastly expand the chemical space accessible to microkinetic modeling efforts.

Estimating model generalizability:

To investigate the generalizability of our developed models to data outside of our dataset, we constructed specific test sets that were excluded from model training and validation sets. The results of our analysis are summarized in Table 2.

<table>
<thead>
<tr>
<th>Model</th>
<th>Random split</th>
<th>Test set: Ag</th>
<th>Test set: Mo</th>
<th>Test set: Pt</th>
<th>Test set: selected molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Valid.</td>
<td>Test</td>
<td>Valid.</td>
<td>Test</td>
<td>Valid.</td>
</tr>
<tr>
<td>1-feature lin. reg. (LR)</td>
<td>0.28</td>
<td>0.26</td>
<td>0.30</td>
<td>0.11</td>
<td>0.26</td>
</tr>
<tr>
<td>N-feature LR</td>
<td>0.23</td>
<td>0.24</td>
<td>0.22</td>
<td>0.12</td>
<td>0.21</td>
</tr>
<tr>
<td>Reduced N-feature LR</td>
<td>0.36</td>
<td>0.43</td>
<td>0.43</td>
<td>0.35</td>
<td>0.36</td>
</tr>
<tr>
<td>Random forest (RF)</td>
<td>0.18</td>
<td>0.21</td>
<td>0.22</td>
<td>0.14</td>
<td>0.21</td>
</tr>
<tr>
<td>Reduced RF</td>
<td>0.27</td>
<td>0.27</td>
<td>0.34</td>
<td>0.18</td>
<td>0.31</td>
</tr>
</tbody>
</table>

Here, we show the resulting model performance (mean absolute error, in units of eV) with a random split – i.e., identical to the performances previously discussed – alongside four specifically constructed test sets. The first three are for specific metals, which tests generalizability to catalysts outside of our dataset, but for the same set of
molecules. We find, generally, excellent generalizability to weak and moderate binding catalysts (Ag, Pt), with most models performing either similarly or slightly better than the random split case. We note here that these test sets are smaller than the test set resulting from a random split, which may explain improvement in model performance in these cases, since the models will have more data to train against before testing. In contrast, we find relatively poor generalizability to a strong binding catalyst in Mo. We hypothesize that this is because our dataset contains relatively few strong binding metals, and so removing one metal from the dataset entirely limits the model’s ability to capture this type of data. Finally, the last constructed test set corresponds to a set of selected molecules, in particular $\text{C}_2\text{H}_6$, $\text{C}_2\text{H}_6\text{NH}$, $\text{CH}_3\text{SH}$, $\text{H}_2$, $\text{PH}_3$, HF, and HCl. For this test set, we find good model performance for models that include the reaction energy, and relatively poor performance for the reduced models that exclude reaction energy. However, we note that a MAE of 0.5 eV for the reduced random forest model may still be sufficient to screen poor catalysts from a large dataset.

Conclusions
To summarize, in this work we report a new database of over 500 computed activation barriers for dehydrogenation reactions on a range of transition metal catalysts and molecules. Barriers were calculated using a one-dimensional scan of the $X$—$\text{H}$ bond length, where $X$ represents the atom from which hydrogen is abstracted across the reaction coordinate. When benchmarked with a more rigorous machine learning accelerated nudged elastic band routine, we found excellent parity across approximately 20% of the dataset, suggesting that this method is satisfactory for estimating the activation barrier – but likely not the exact transition state geometry. As such, this method may not be sufficient if e.g. entropy of activation is a desired quantity in the analysis.

Given growing interests in capturing complexity in catalysis, and the combinatorial explosion of activation barriers that must be considered as the number of metals, facets, reaction pathways, etc. increases, calculating the reaction energy with DFT for complex systems will likely prove to be computationally intractable. Using simple machine learning methods, we demonstrate the importance of the reaction energy as a feature in predicting activation barriers. Including 11 additional features and binary combinations of these features, but excluding reaction energy as a feature, our multi-feature linear regression models are unable to predict activation barriers as effectively as a single feature linear regression model including reaction energy (i.e., a typical BEP-type scaling). Our reported nonlinear models, excluding reaction energy as a feature, show similar performance compared to a BEP model in this dataset. We find that generalizability to new catalysts with the complex models is generally quite good, perhaps with an exception for strong binding catalysts, which our dataset is sparse in. However, we find that generalizability to new chemistries (e.g., $C$—$\text{N}$ bond breaking/formation) for such models is poor in comparison to a BEP type model. We note that the performance in generalizing to new chemistries may be sufficient to screen poor catalysts from a large dataset, depending on the use case of the model.

Other promising future research directions will be related to understanding the error propagation resulting from ML estimates of the reaction energy (using, e.g., potentials...
from the Open Catalyst project) in tandem with BEP-type models fit using transition metal energetics. Additionally, although (e.g.) the OC2020/2022 projects report errors for predicting energetic minima, on which the potentials are trained, much less data is available for their predictive power for activation barriers directly.

Acknowledgements
This material is based upon work supported by the National Science Foundation under Grant No. 2133576. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The authors acknowledge the High Performance Computing Center (HPCC) at Texas Tech University for providing computational resources that have contributed to the research results reported within this paper. Some of the computing for this project was performed on the Sherlock cluster. We would like to thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to these research results.

References


https://doi.org/10.26434/chemrxiv-2023-t62ryv2 ORCID: https://orcid.org/0000-0001-9542-0988 Content not peer-reviewed by ChemRxiv. License: CC BY-NC-ND 4.0


