Received: 29 September 2021 | Revised: 18 November 2021 | Accepted: 22 November 2021

DOI: 10.1002/wcms.1593

OVERVIEW

WIREs COMPUTATIONAL MOLECULAR SCIENCE WILEY

# Machine learning activation energies of chemical reactions

**Toby Lewis-Atwell**[1] | **Piers A. Townsend**[2] | **Matthew N. Grayson**[2]

[1]Department of Computer Science, Faculty of Science, University of Bath, Bath, UK

[2]Department of Chemistry, Faculty of Science, University of Bath, Bath, UK

**Correspondence**
Matthew N. Grayson, Department of Chemistry, Faculty of Science, University of Bath, Bath, UK.
Email: m.n.grayson@bath.ac.uk

**Funding information**
This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) (EP/L016354/1); U.K. Research and Innovation (UKRI) (EP/S023437/1).

**Edited by:** Peter R. Schreiner, Editor-in-Chief

[Correction added on 30 May 2022, after first online publication: Additional Funding information, U.K. Research and Innovation (UKRI), grant reference number EP/S023437/1 has been added.]

**Abstract**

Application of machine learning (ML) to the prediction of reaction activation barriers is a new and exciting field for these algorithms. The works covered here are specifically those in which ML is trained to predict the activation energies of homogeneous chemical reactions, where the activation energy is given by the energy difference between the reactants and transition state of a reaction. Particular attention is paid to works that have applied ML to directly predict reaction activation energies, the limitations that may be found in these studies, and where comparisons of different types of chemical features for ML models have been made. Also explored are models that have been able to obtain high predictive accuracies, but with reduced datasets, using the Gaussian process regression ML model. In these studies, the chemical reactions for which activation barriers are modeled include those involving small organic molecules, aromatic rings, and organometallic catalysts. Also provided are brief explanations of some of the most popular types of ML models used in chemistry, as a beginner's guide for those unfamiliar.

This article is categorized under:

Structure and Mechanism > Reaction Mechanisms and Catalysis

Computer and Information Science > Visualization

**KEYWORDS**

activation barriers, chemical reactions, data, machine learning, reactivity prediction

## 1 | INTRODUCTION

A critical component in any computational reaction modeling study is the accurate calculation of activation energies. Defined as the energy difference between the reactant states and the transition state (TS) of a reaction, it represents the minimum energy a system must attain for a reaction to proceed, and from its value, product identities, reaction rates, and selectivities can be computed.[1–5] Activation energies (or activation barriers, used interchangeably here) can be evaluated by modeling reactions with quantum mechanical (QM) calculations, particularly density functional theory (DFT).[6,7] DFT has played a major role in calculating activation energies of reactions in several valuable areas of chemistry including: mechanistic modeling of organic reactions,[8–12] drug design,[12–14] toxicology,[15,16] and catalyst design.[17–19] However, accurate QM and DFT calculations incur a significant computational expense[20–22] and their practicality for reaction modeling on a large scale is thus limited. Ideally, there could be a method that can predict in seconds or

minutes the results of QM calculations that may take hours, days, or weeks. Other methods for modeling reaction activation barriers include force field methods specially designed for the modeling of TSs.[23–27] However, these will often be limited in accuracy compared to QM methods, due to the low abundance of highly accurate data for their parameterization and the inherent approximations of force fields that make them computationally cheap.[20,22,28]

Machine learning (ML) could offer a remedy to the issue of computational cost in quantum chemistry. ML models (or ML methods, used interchangeably here) are, broadly speaking, advanced statistical methods that aim to analyze and deduce patterns and properties within large amounts of data.[29–33] Importantly, these models have extremely low computational costs compared to QM calculations. However, for a ML model to be useful, it must first be trained. This involves creating a dataset that, for chemical applications, should contain features (or descriptors) of the systems under study and the accurate values of the properties (e.g., activation energies) of those systems one wishes to predict. Chemical features may be representations or encodings of molecular geometries or atomic environments, or other properties of the systems such as atomic charges, bond orders, or orbital energies. The use of raw Cartesian coordinates as features is never advised since the values of these coordinates are not invariant upon translation or rotation. That is, a chemical system may be translated or rotated, and the internal geometry of the system will remain the same, but the absolute values of the Cartesian coordinates will be different. Therefore, the ML model will likely not be reliable since it cannot be trained to consider all possible sets of Cartesian coordinates that describe the translations or rotations of a system.

The training of ML models is carried out by passing the training data to an initial form of the model, followed by minimizing the error between the ML prediction and the reference value of the target property of interest. This is achieved by adjusting the parameters within the model's algorithm. The quality of the trained ML model is then assessed by inputting descriptors from unseen systems, with the error from this testing showing how well the model can make predictions on new data. A generic scheme for the workflow to create a ML model to predict activation energies is shown in Figure 1. It should be noted that the final accuracy of an ML model is limited to the underlying accuracy of the training data. An ML model is only able to reproduce the target values from the training data it is provided with, and therefore if that data is inaccurate, the ML will also be inaccurate (probably even more inaccurate than the training data, since for a complex system the functional fit of the ML model to the data is not likely to be perfect). A common aim for a ML model in chemistry is to make predictions of activation barriers at "chemical accuracy" which is considered to be an error of 1 kcal $mol^{-1}$.[34,35] Note, the "chemical accuracy" threshold is not applicable to all types of energetic prediction. For instance, if aiming to predict chemical selectivity, an error of 1 kcal $mol^{-1}$ could cause a substantial change in the predicted outcome. Thus, it is wise to consider whether an accuracy threshold, if used, is appropriate to the prediction being made.

Outside of chemistry, ML has been applied in many circumstances including: natural language processing[36–39]; driverless vehicles[40–44] speech recognition[45–48]; handwriting analysis[49–51]; enhancing image resolution[52–55]; robotics[56–60]; and, famously, beating the human champions of the games chess[61] and Go.[62] Within chemistry, an incomplete list of applications include: evaluating potential energy surfaces of ground[63–66] and excited states[67,68]; forming solutions to the Schrödinger equation[69,70]; modeling molecular wavefunctions[71,72]; accelerating TS optimization[73,74]; finding exchange-correlation functionals for DFT[75,76]; predicting reaction rate constants[77,78]; predicting the outcomes of organic reactions[79–84]; X-ray,[85–87] UV–Vis,[88] IR,[89–92] and NMR[93–95] spectroscopies; sequence-based biomolecular function prediction[96,97] and predictions of protein structures.[98–101] Another very recent and exciting application of ML in chemistry is the prediction of activation energies. Before discussing work where ML is used to calculate activation barriers, the following section shall provide some brief explanations of the most commonly used ML models.

## 2 | ML MODELS

### 2.1 | Linear regression

Linear regression (LR) attempts to describe the relationship between input features ($x$) and the target property ($y$) as a linear function of the input features, as in Equation (1).[29,102,103] For a single input feature, this function reduces to the "$y = mx + c$" form that most will be familiar with. When more than one input feature is considered, the model is referred to as multi-variate linear regression.

$$y = f(x) = \beta_0 + \sum_{i=1}^{n} \beta_i \cdot x_i. \quad (1)$$

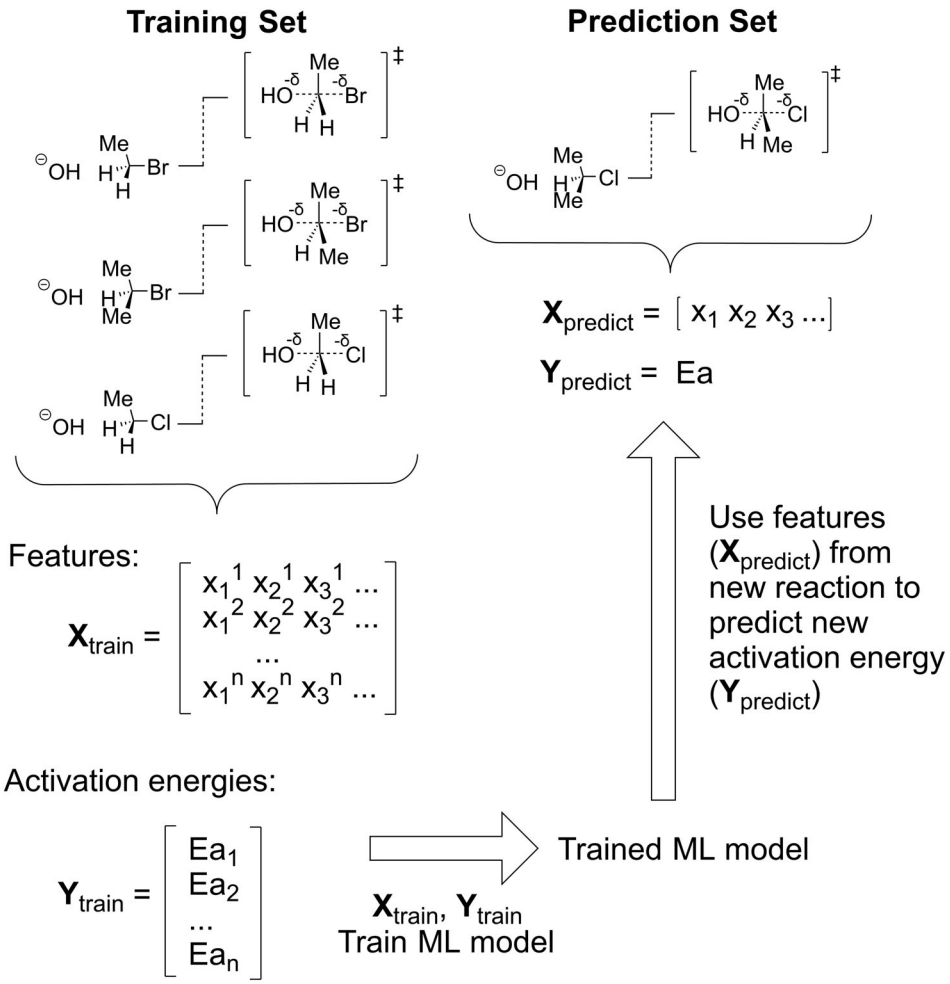


**FIGURE 1** A generic strategy for designing a ML model to predict a target chemical property, such as activation energy ($E_a$, **Y**) of $S_N2$ reactions. Encoding of both the training and unseen reacting systems gives the input features (**X**). The model is trained to reproduce the target property of interest ($\mathbf{Y}_{train}$), from the features of the training systems ($\mathbf{X}_{train}$). The trained ML model can be applied to unseen systems with features $\mathbf{X}_{predict}$, and the model's performance is assessed by how accurately it predicts the property $\mathbf{Y}_{predict}$ from the unseen system

The $\beta$ coefficients are the parameters in the model that are adjusted such that the model is fit to the data. This can be achieved by, for example, minimizing a loss function ($L$) of the mean squares error of the model with respect to the $\beta$ coefficients, given by Equation (2).

$$L = \frac{1}{n}\sum_{j=1}^{n}\left(y_j - f(x_j)\right)^2, \tag{2}$$

where $y_j$ is the true value of the target property from the training data, $f(x_j)$ is the predicted value from the linear model, $n$ is the number of datapoints in the training set, and $j$ runs over all training points with features $x_j$ and target values $y_j$.

It is also possible to introduce an extra term to the loss function that penalizes redundant features in the linear model. This is known as regularization, and two common types are called L2 and L1. L2 regularization adds the sum of the squares of the $\beta$ coefficients to the loss function, as in Equation (3).

$$L = \frac{1}{n}\sum_{j=1}^{n}\left(y_j - f(x_j)\right)^2 + \lambda\left(\sum_{i=1}^{n}\beta_i^2\right), \tag{3}$$

where $i$ runs over all of the $\beta$ coefficients for all the input features and $\lambda$ is the regularization parameter and gives the "strength" of the L2 regularization. Larger $\lambda$ values lead to an increase in the magnitude of the penalty and may lead to underfitting if too high. When L2 regularization is used in a LR model, it is known as ridge regression.[104,105]

L1 regularization adds the magnitude of the $\beta$ coefficients to the loss function, as in Equation (4).

$$L = \frac{1}{n}\sum_{j=1}^{n}\left(y_j - f(x_j)\right)^2 + \lambda\left(\sum_{i=1}^{n}|\beta_i|\right). \tag{4}$$

The use of L1 regularization with an LR model is known as least absolute shrinkage and selection operator (LASSO) regression.[106] An interesting advantage of LASSO regression is that the L1 term in the loss function causes the $\beta$ coefficients of low importance features to become zero, and it can thus be used for feature selection. When both L1 and L2 regularization terms are applied to a LR model, it is known as elastic net regression.[107] Note that regularization is a technique that is not only applicable to LR, the regularization terms can be included in the loss functions for many ML methods.

LR models are easily trained and understood by the user, however, their predictions can be less reliable than other models if the key assumption behind a LR model is not present in the data; that a linear relationship exists between the features and the target property.[32]

## 2.2 | Neural networks

A standard feedforward neural network (NN) is made up of layers of nodes (or neurons) with connections between all nodes in adjacent layers (Figure 2).[29–32,108–111] Each node operates by taking the sum of all values that are input to it and outputting the activation function value of that sum. The activation function may be, for example sigmoid (Figure 2, right), hyperbolic tangent, or rectified linear functions.[112] A sigmoid activation function means that if the total input to a node is greater than a certain threshold, the node will output a larger value. Similarly, for inputs lower than the threshold, the node will output a smaller value. This is analogous to the process which neurons in the brain undergo; neurons receive electric signals, and if the total signal passes that neuron's threshold, it "fires" and transmits its signal to further neurons.[113]

The NN's ability to learn the relationship between the input features and output variables of a dataset comes from the weight parameters between nodes. By adjusting these weights such that the error between the NN predictions and the actual values in the training data is minimized, the NN effectively learns the relationship between the input and output variables. NNs with larger numbers of layers are known as deep neural networks (DNNs) and are the algorithms behind "deep learning."[84,114–118] DNNs are powerful predictors due to the extreme flexibility of the algorithm; however, they typically require much larger datasets to be reliable[119] and all but the simplest NNs become impossible for a

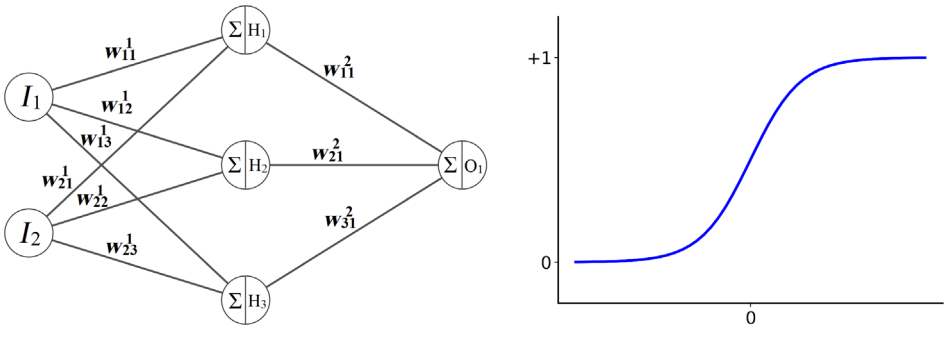


**FIGURE 2** (Left) A representation of a simple NN. Starting from the input layer, the outputs of each node are multiplied by the weights for each connection before being passed to the next layer. Each connection has a unique weight associated with it, $w_{ij}^k$, where the weight corresponds to the connection between the $i$th and $j$th nodes in the $k$th and $(k+1)$th layers respectively. An activation function is applied to the sum of the inputs to hidden layer nodes that produce the outputs of those nodes. (Right) The form of a sigmoid activation function

human mind to grasp. Hence, NNs are often referred to as "black box" methods[117,120–122] due to the difficulty in comprehending how or why the predictions of the model were made.

## 2.3 | Support vector regression

A support vector machine (SVM) is a tool for classifying data (a chemical example of which could be whether a particular compound will undergo a given reaction). This is done by constructing a hyperplane such that all datapoints from each class lie on opposite sides of the hyperplane. If the data are not linearly separable in the original feature space, a kernel function is used to map the data to a space where the categories are linearly separable by the hyperplane (Figure 3). The hyperplane is positioned so the distances between the hyperplane and its nearest datapoints (the support vectors, Figure 3) are maximized. In support vector regression (SVR), which considers continuous rather than discrete data (perhaps the percentage yield of a particular reaction), the support vectors between the hyperplane and the data are minimized, allowing the hyperplane to give the best possible description of the data. Should the data not be linearly related in the original feature space, a nonlinear kernel function can be applied to the data, and they become linearly related in the higher-dimensional space.[29,30,32,111,123–125] However, there are some considerations that need to be made when utilizing a SVM or SVR. The choice of kernel function is very rarely obvious from the data alone (since the optimal separation or relationship between the datapoints may be found in a higher-dimensional space), and potentially several different kernel functions may need to be tested to maximize the predictive performance of the model. Furthermore, from a practical perspective, the SVR algorithm must store all of the support vectors for the datapoints in memory, and therefore the computational cost of the model increases with the size of the training dataset.

## 2.4 | Kernel ridge regression

The basic idea of kernel ridge regression (KRR) is to map the input features to a higher dimensional space and perform Ridge regression in that space.[29,31,125–129] The KRR prediction for the features of a new instance, $\mathbf{x}$, is given by a weighted sum of all examples in the training set, as in Equation (5).

$$f(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i k(x_i, x), \tag{5}$$

where $k$ is the kernel function, $\alpha_i$ are the regression coefficients (cf. the $\beta$ coefficients in LR), and $x_i$ are the feature vectors for the training examples where $i$ runs over all examples in the training set. The loss function for KRR is then given by Equation (6).

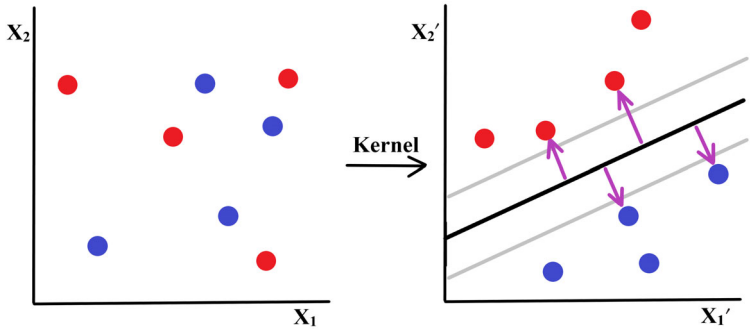


FIGURE 3 Illustration of the workings in a SVM. For example, a SVM could be attempting to classify whether a given compound will undergo a reaction (e.g., blue represents a compound that will react and red represents a compound that will not). Since the data are not linearly separable in the original feature space, a kernel function is applied that maps the data to a higher-dimensional space, where they are separable by the hyperplane. The distances between the datapoints and the hyperplane are known as the "support vectors" (purple arrows)

$$L = \frac{1}{n}\sum_{j=1}^{n}\left(y_j - f(x_j)\right)^2 + \lambda||f||^2, \tag{6}$$

where $\lambda$ is the regularization parameter, $j$ again runs over all the examples in the training set, $y_j$ is the true target value in the training set, and $||f||$ is the norm of the function $f$ from Equation (5) in the higher dimensional space. Since the square of the norm is added to the loss function, this is where the ridge regression part of KRR arises. Setting the first derivative of the loss function with respect to the $\alpha$ coefficients and solving gives the values of the $\alpha$ coefficients as in Equation (7).

$$\boldsymbol{\alpha} = (\mathbf{K} + \lambda\mathbf{I})^{-1}\mathbf{y}, \tag{7}$$

where $\boldsymbol{\alpha}$ is the vector containing the $\alpha$ coefficients, $\mathbf{I}$ is the identity matrix, $\mathbf{y}$ is the vector of the target values from the training set, and $\mathbf{K}$ is the kernel matrix with each element given by $K_{ij} = k(x_i, x_j)$, the kernel function between $x_i$ and $x_j$ which are the feature vectors for two training set examples, where $i$ and $j$ run over all training points. A critical disadvantage of KRR is that the kernel function must be computed between the new instance and every datapoint in the training set each time a prediction is to be made. This can make KRR computationally expensive for large training sets.

## 2.5 | Random forest regression and gradient boosting

Random forests consist of decision trees which contain nodes that, for a given input, split the data according to the value of that input (Figure 4). Each node is followed by two separate nodes that further split the data until the end of the tree is reached. In classification problems, categories within the data will have been separated by the end of the tree, so that any input will be placed into its proper category. In regression problems, the tree splittings formed during training minimize the root mean square error (RMSE) between the output of the decision tree and the true value corresponding to the input from the training set. Random forest regression (RFR) uses numerous decision trees and takes the average prediction of all the decision trees in the forest as its final output ("the wisdom of crowds").[29,30,32,111,124,130,131] For example, if one attempts to predict the value of a property with RFR, each tree is provided with the set of descriptors, and each provides its prediction of the property, and the average value from all trees becomes the RFR prediction (Figure 5). Random forests are generally considered to perform well for datasets of moderate size (perhaps a few hundred points) and increasing the number of decision trees within the model will typically improve model performance. However, if too many trees are used, the computational cost may become an intractable problem.

Gradient boosting (GB), or tree boosting, is also constructed from decision trees, but rather than initializing the model with a certain number of trees and training them together, GB trains one tree at a time such that a subsequent tree minimizes the error from the previous ensemble of trees.[29,32,132] GB is a very similar ML method to RFR, and it also has the issue of computational cost when very large numbers of trees are used in the model. Also, one significant difference between RFR and GB is that the decision trees in RFR can be trained in parallel, while those in GB are created

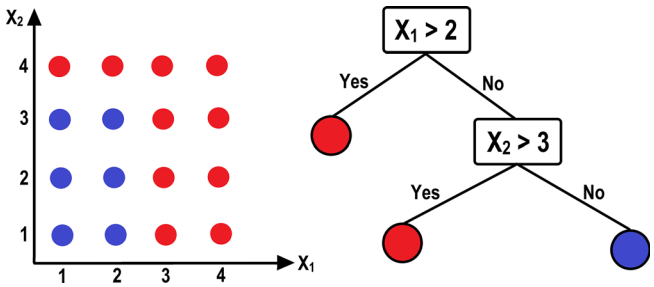


FIGURE 4 A simple decision tree (right) to classify whether a point in a dataset (left) will be blue or red based on the values of its two features

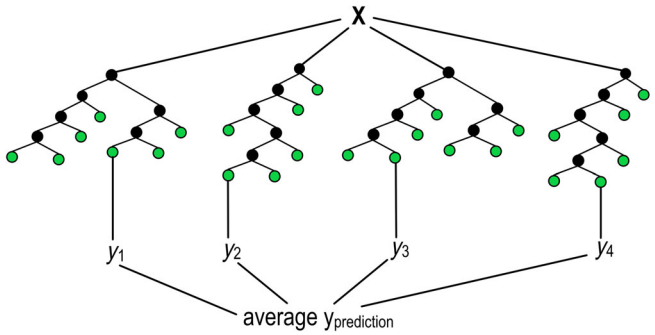


**FIGURE 5** An illustration of RFR, which is constructed from the combination of many decision trees. Each tree is supplied with the same input features (**X**) and makes its (likely rather inaccurate) prediction of the desired output property. The splittings are based on the values of the input features (e.g., atomic charges or log $K_{ow}$). The values of the features determine which path on the tree is followed and thus which prediction the single tree produces. The average value of the predictions from all trees gives the final RFR prediction ($y_{prediction}$)

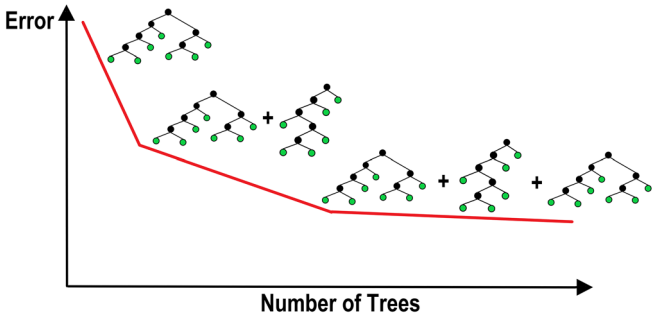


**FIGURE 6** An illustration of the GB method. Training starts from an initial tree and subsequent trees are added such that the error from all tree predictions combined is minimized

and trained in series. This means that, for the same number of trees at the same depth, GB will have a longer training time than RFR (Figure 6).

## 2.6 | Gaussian process regression

In Gaussian process regression (GPR) a prior distribution of functions for the input features is sampled, such that the mean of the prior function distribution is zero for all values of all features. The forms of the functions are defined by a kernel function, which could be for example, sinusoidal or the radial distribution function. As training data are added to the model, the function distribution is updated so that each function passes through the training points, and this is known as the posterior distribution of functions. The GPR predictions for new inputs are given by the mean of the posterior distribution of functions for those input features.[31,133–137] Empirically, GPR seems to have the advantage of not requiring a large training dataset[60,133,138] as in other models such as NNs. However, just as in KRR, it must take the entire training set into account for every prediction made, meaning its computational cost scales expensively relative to other ML models (the cost scales with the cube of the number of training points,[133,134] which presents a major bottleneck for the use of this algorithm when dealing with large datasets, as is common in problems for which ML is useful) (Figure 7).

## 2.7 | Other points for machine learning

There are a few other points with regard to ML that are worth mentioning. The first is that just as the parameters of the ML algorithm (e.g., the $\beta$ coefficients in LR, or the weights in a NN) must be optimized in order to train the model, the

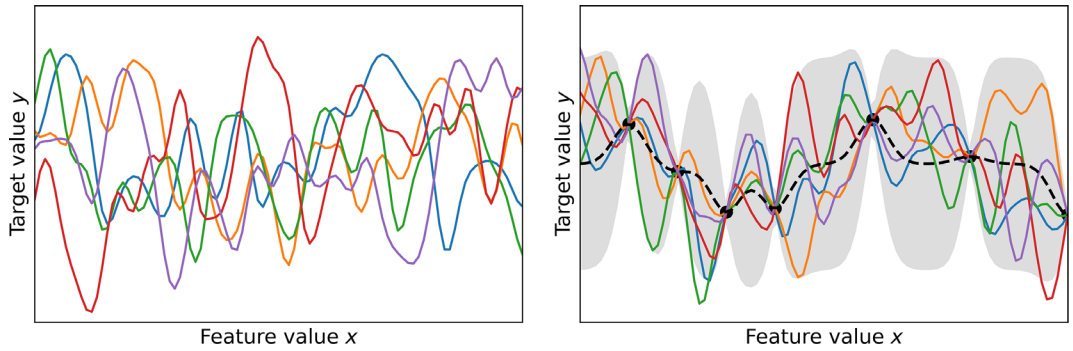


**FIGURE 7** Illustration of GPR. (Left) The prior distribution where the sample of functions has mean zero at all points. (Right) The posterior distribution after training, which is a sample of all the possible functions that pass through the training data (black dots). New GPR predictions are given by the mean of the posterior distribution (black line). The gray shading shows the standard deviation of the posterior distribution, which gives the uncertainty of the GPR prediction

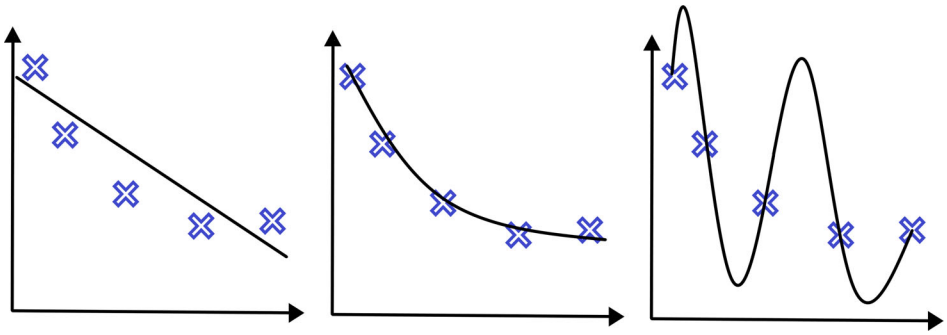

**FIGURE 8** (Left) An underfit model, where a linear relationship is not sufficient to describe the data. (Center) A model with an ideal, balanced fit. (Right) An overfit model, where the model will perfectly predict the training data, but does not capture the general underlying relationship and will likely not perform well for any data outside the training set

hyperparameters of the models may also need to be optimized to ensure that the model is best suited to the dataset at hand. These are the parameters that relate to the models themselves, rather than the ones used to fit the model to training data, for example, the number of hidden layers in a NN, or the number of decision trees in RFR.

Care must also be taken during training to avoid the occurrences of underfitting and overfitting (Figure 8). Underfitting can occur when the model is not flexible enough to account for the complexity of the data. An underfit model will show large errors between its predictions and both training and testing errors. Overfitting occurs when the training data have been passed to the model too many times, or the model is too flexible for the dataset. The model thus becomes overfit towards the training data, and it has learned the specific trend in the training data and is unable to generalize. An overfit model is characterized by a low error for the training data, but a high error for testing data. Lastly, ML methods are interpolative not extrapolative. Generally, one cannot expect good performance from ML models for inputs outside of the range of the training data. For example, if a model were trained to predict the activation energies of reactions following the E2 mechanism, this model will not likely provide reliable predictions for reactions outside its training domain, such as $S_N2$ reactions.

## 3 | ML FOR ACTIVATION ENERGIES

Given the success of ML both inside and outside of chemistry, it has been considered for the prediction of reaction activation barriers. Several groups have made use of ML in this way, and this section reviews the work concerning ML predictions of homogeneous chemical reaction activation barriers.[139–148] Note, for works discussed in this review, the terms "activation barrier" and "activation energy" refer to those of elementary reactions, rather than the apparent barrier that would be observed over an entire reaction composed of multiple elementary steps.

Figure 1 displays a general flowchart showing the very basic steps for creating a ML model to predict activation barriers. As well as obtaining a large dataset of reaction activation energies, a critical stage in creating these ML models is extracting the molecular representations, that is, the molecular features (or descriptors). Often used molecular descriptors in predicting activation barriers can be organized into different categories. "Physical organic" descriptors[139,142–144,146,149] that are molecular and atomic properties such as atomic charges, bond orders, molecular orbital energies, measurements of molecular sterics, or thermodynamic quantities (e.g., free energy or enthalpy). Also used are geometric descriptors, for example, two-dimensional molecular representations known as molecular fingerprints[139–141,147,148] and descriptors of the three-dimensional atomic environments.[142] Physical organic features seem to require expensive QM calculations to be obtained, whereas the purely geometric features may not require such expensive calculations.

First, studies that apply ML models to directly predict activation energies will be covered, along with some aspects of the use and training of NNs in the context of these studies. Following this, attention will be turned to some of the limitations in these works; either the dataset used, or their use of reaction energy (the energy difference between products and reactants) as an input feature to the models. Further, some studies that address these issues will be covered.

Works that compare different types of features are examined, followed by those that produced models with reduced data requirements, and those that attempted to interpret their models. Finally, the uses of Δ-ML (see Section 3.4) to predict reaction activation energies will be discussed, and an interesting non-direct use of activation energy ML will be briefly mentioned. There also exists a fair amount of literature concerning ML predictions of the adsorption energies of small molecules onto catalytic surfaces,[150–156] but these are beyond the scope of the discussion here and a review on ML for heterogeneous small molecule activation has been published very recently.[157] We also do not include works where ML has been used for predicting activation energies in heterogeneous systems,[158–161] only homogeneous reactions are covered here. There are also a couple of works exploring ML for activation barriers from molecular dynamics simulations,[162,163] which are also not covered here.

## 3.1 | Direct predictions of activation energies with machine learning

The first step in building a ML model for chemical property prediction is the choice of algorithm. It may not be obvious which ML method is best suited for a given problem, but given the flexibility and predictive power of NNs, they should be expected to give good predictions of reaction activation barriers. In this section, the training and use of NNs are discussed, in the context of multiple studies that used NNs (among other models) for activation barrier predictions.

In work by Choi et al. in 2018,[139] an attempt was made to assess the feasibility of using ML for predicting activation energies of organic reactions (see Figure 9 for a few example reaction types). SVR, GB, and NNs with three to six hidden layers were trained on 12,704 (6078 unique) gas-phase reaction barriers from the RMG-py dataset.[164] The models were assessed by mean absolute error (MAE), RMSE, and coefficient of determination ($R^2$) between the ML-predicted and actual values of activation energy from the dataset. The NN reached a minimum MAE (for the test data) of roughly 2.5 kcal mol$^{-1}$ at five hidden layers, indicating that overfitting was occurring with increased layers. However, none of the NNs were greater in accuracy than the GB method, which had the lowest MAE of just under 2.0 kcal mol$^{-1}$, despite the fact that one may expect the more complex and flexible NN model to show a better performance. Choi et al. state that this occurred due to an insufficient amount of data used in training the NNs. Although, all of the NNs had the

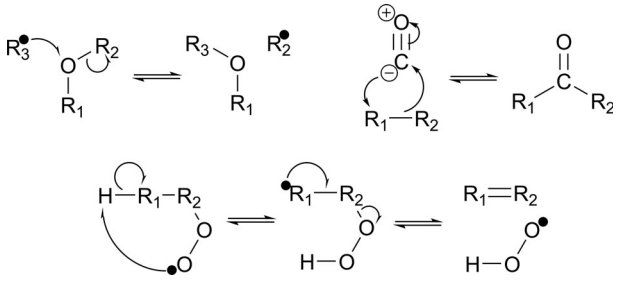

FIGURE 9 Three examples of some of the reactions in the RMG-py dataset used by Choi et al. in their ML models for predicting the activation energies of these reactions. (Top left) Radical oxygen substitution. (Top right) CO insertion. (Bottom) $HO_2$ elimination from peroxy radical. The full set of reactions available in this dataset may be found in Reference 164

same numbers of nodes per hidden layer and optimizing the numbers of hidden layer nodes could possibly have led to an improvement in NN accuracy.

Li et al.[142] used 15 ML methods (including, but not limited to, LR, SVR, RFR, KRR, GPR, GB [implemented in XGBoost], and NNs) to predict the regioselectivity of 3406 radical C–H functionalization reactions of heteroaromatic rings. Activation barriers of the reactions were determined from the differences between the Gibbs free energies of the minima and TSs, which were geometry optimized at the B3LYP/6-311+G(2d,p) level of theory, with single-point energies calculated at the M06-2X/aug-cc-pVTZ level of theory. The regioselectivity of the reactions was determined by the difference between free energy activation barriers ($\Delta\Delta G^{\ddagger}$) of the functionalization reaction for different positions on the heteroaromatic ring (Figure 10). The ML models were trained to predict the values of $\Delta\Delta G^{\ddagger}$, and several different sets of features were used as inputs (see Section 3.2). The NN attained the second lowest $R^2$ correlation coefficient of 0.967 between the ML-predicted and DFT-calculated values of $\Delta\Delta G^{\ddagger}$, very close to the best, which was GB with an $R^2$ value of 0.968. However, further analysis was only carried out with GB, not the NN. This illustrates a practical consideration when facing model choice in ML; it is likely more worthwhile to use and explore a simpler model if, for a given dataset, it has roughly equal performance with a more complex model such as a NN.

A "frustrated Lewis pair" is a system in which a Lewis acid and a Lewis base are combined, but bulky groups surround the electron accepting and donating centers, forcing the centers to remain separated and preventing the transfer of electron density between them. Frustrated Lewis pairs are used in small molecule catalysis as they can be inserted between the Lewis centers, and subsequently activated for reactions.[165–169] Migliaro and Cundari undertook a study of methane activation by frustrated Lewis pairs formed from group 13 trihalides and group 15 pentahalides with ammonia (Figure 11).[143] Note that simplified structures were used to reduce the impact of steric effects on the reaction barriers, and hence the effects of periodicity on the reaction could be analyzed directly. The group 13 elements in the Lewis acids were B, Al, Ga, In, and Tl, the group 15 elements were P, As, Sb, and Bi, and calculations were performed on the F, Cl, Br, and I halides of all nine elements. Calculations of the reactant and TS energies used the ωB97X-D density functional and the Def2-TZVPP basis set. The deprotonation of methane by the Lewis base (see Figure 11), rather than hydride abstraction by the Lewis acid, was calculated to be the more favorable mechanism both kinetically and thermodynamically. Migliaro and Cundari formulated a model for the activation energies of this mechanism. No single descriptor of

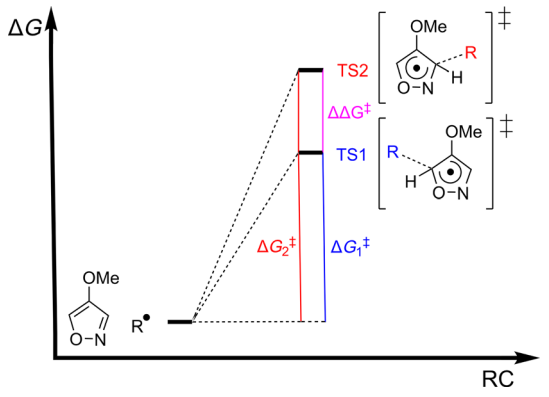


**FIGURE 10** Illustration of how Li et al. determined regioselectivities for radical (R˙) C—H functionalization reactions of heteroaromatic rings. The $\Delta G^{\ddagger}$ value for a single reaction is calculated by the difference in free energies of the reactants and the transition states. Regioselectivity is then given by the difference between $\Delta G^{\ddagger}$ values for two positions on the ring. The ML models were trained to predict these $\Delta\Delta G^{\ddagger}$ values

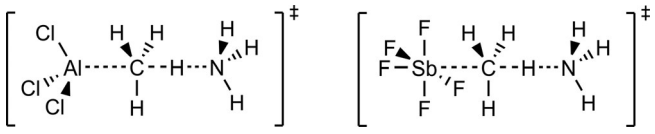


**FIGURE 11** (Left) Transition state for deprotonation of methane with a group 13 trihalide (here $AlCl_3$). (Right) Transition state for deprotonation of methane by a group 15 pentahalide (here $SbF_5$). Note that the mechanism here involves the insertion of methane between the Lewis acid–base adduct and a proton from methane is transferred to the Lewis base (ammonia)

the systems had strong linear correlation with activation energy, and hence a NN model was created. Only 35 datapoints were available (a TS for the reaction with $BiF_5$ was not found) so a very small NN with one hidden layer containing five nodes was trained to predict the activation energies for the insertion of methane into the Lewis pair. The importance of avoiding overfitting in NN training was also seen in this work. For a NN to become an accurate predictor, the training procedure must be repeated with the same data several times. The mean square error in the NN's prediction for the training, validation, and testing data decreased with each training iteration (or epoch). However, after nine iterations, overfitting towards the training data became apparent, and the errors for the validation and testing sets started to increase, while the training error continued to decrease. Thus, training was stopped at nine epochs in this case, so that the NN had its maximum possible general predictive accuracy. After training, the NN had an $R^2$ coefficient between the NN-predicted and DFT-calculated activation energies of 0.908 for all 35 datapoints, and an $R^2$ of 0.851 for the test data alone.

The works discussed above have demonstrated that NNs can be used to make strong predictions of activation energies based on input features that do not usually have strong linear correlations with activation energy. The flexibility of NNs gives rise to their strong predictive ability, but as seen in the above discussion, the user must tune the NN architecture and training process to avoid under/overfitting.

## 3.2 | Dataset limitations and the use of reaction energy as a feature

Grambow et al. highlight[141] that the results of Choi et al.[139] have some limitations. Most of the reactions in the dataset used to train and test the ML models were mostly composed of many similar reactions, which limits the general applicability of the ML models. Another unfortunate fact of the RMG-py dataset[164] used by Choi et al. is that many of the 12,704 gas-phase reaction energy barriers were different values from multiple sources, and thus, only 6078 unique reactions were actually considered. All 12,704 barriers were used to train the ML models, since there was no way of determining which value for each reaction was the more accurate. Therefore, on average, each unique reaction in the dataset would have two activation energies associated with it. It is not unreasonable to expect that activation energies for the same reaction from different sources may be similar, and the testing set would likely contain many reactions that were already included in the training set. Thus, the predictions of the ML models trained on this data would show on average, better predictive performance than if tested with reactions which the model had not already seen, albeit with data from a different source. Ideally, future workers should consider diversity of reaction datasets, such that ML models can be constructed with greater generality and lower costs of development.

Choi et al. used the reaction enthalpies and entropies (B3LYP/6-31G*) as features. This presents a practical issue and undermines the use of ML as a rapid predictor of chemical properties. For any reaction to be analyzed by this ML model, one requires thermodynamic properties of the products and reactants. These must be obtained via DFT calculations, which is not feasible on a high-throughput scale.

These same drawbacks can also be found in the study by Migliaro and Cundari.[143] As mentioned above, Migliaro and Cundari used only 35 methane deprotonation reactions, all of which follow identical mechanisms, and the only difference between structures is the periodic alteration of the halides and the central atom in the Lewis acid. However, it is not clear if this model could make accurate predictions for reactions involving more realistic frustrated Lewis pairs, with bulkier groups around the Lewis acid and base centers. The model also takes reaction energy as input, but also the dissociation energy of the acid–base adduct and the global electrophilicity index[170] (a measure of the level of Lewis acidity of a molecule), all of which had to be calculated at the expensive ωB97X-D/Def2-TZVPP level of theory.

A promising example comes from Palazzesi et al.,[145] where they created "BIreactive," which uses Extra-Tree regression[171] (similar to RFR, but with small differences in how best splittings in the trees are chosen in the training algorithm) to predict activation energies for covalent drugs, which bind to their biological target through covalent bonds rather than non-bonded interactions.[172–174] The reactions of 2083 acrylamides and 2716 2-chloroacetamides with the methanethiolate anion (used as a less-complex stand-in for the peptide glutathione) were chosen to make up the dataset (Figure 12). Conformational searches[175] were performed for each molecule, and the five lowest energy conformers of each were calculated at the ωB97XD/cc-pVDZ level of theory. Activation energies were calculated from the difference between the Boltzmann-averaged energies of the reactants and TSs. Separate BIreactive models were trained for the acrylamides and 2-chloroacetamides with 44 and 45 descriptors respectively for each system. Notably, none of the descriptors for the systems required QM calculations, and could be obtained from only the geometric information of the molecule.[176,177] The most important features to the two BIreactive models were: the number of terminal, primary

$sp^2$ carbons atoms in the acrylamides, and whether chlorine was present five bonds from the nitrogen in the 2-chloroacetamides. The testing predictions from a five-fold cross-validation were combined and, for acrylamides and 2-chloroacetamides respectively, BIreactive was able to achieve $R^2$ coefficients between the predicted and DFT-calculated activation barriers of 0.85 and 0.69; Spearman rank correlation coefficients of 0.91 and 0.84 and MAEs of 0.75 and 0.88 kcal $mol^{-1}$. The values of the Spearman coefficients[178] show how well the ML model has managed to order its predictions, that is, how well the predicted and actual values of the target activation energies would fit to a monotonic function; the closer the value is to unity, the better the overall ordering. While the accuracies of BIreactive are impressive (notably the test MAEs are substantially less than 1 kcal $mol^{-1}$), the software used to calculate the descriptors[179] is proprietary and has been discontinued at the time of writing, making use of these particular descriptors as well as a replication study rather difficult. Therefore, we would make the recommendation for the use of open-source alternatives for extracting chemical descriptors.

Choi et al.[139] suggested that, to make improvements in the quality of ML predictions of activation energies:

> "A deep learning approach with big reaction data may be a viable solution."

This is the very approach Grambow et al. took, although without using any features that require QM calculations to extract.[141] The model begins with atom-mapped representations of the products and reactants; atom-mapped means that every atom in the reactants was labeled and its final position in the products is identified by that label.[180,181] These initial representations are passed to a directed message passing neural network[182] (D-MPNN) which creates its own learned representations of the molecules. The D-MPNN representation is a series of atomic vectors that contain the bonding information for each atom (collected from the atom type), the bonds with which it is directly involved and its neighboring bonds. The D-MPNN representation was chosen as it had been found that the D-MPNN model could frequently provide better predictions of chemical properties than using more traditional molecular descriptors as inputs.[182] For each atom in the products and reactants, the D-MPNN atomic vectors are subtracted from each other to obtain a difference fingerprint of the reaction. This was done so that the atoms that undergo the greatest changes to their environments (during the reaction) have the greatest influence on the activation energy prediction, since (as Grambow et al. state) the atoms whose environments change very little have only a small effect on activation energy. The difference fingerprint is then passed to another NN that takes the difference fingerprints from each atom and collects them into an encoding of the entire reaction. Lastly, the reaction encoding is input to a final NN, which is trained to predict the activation energies of reactions from their encodings.

The model was trained with the activation energies of 33,000 elementary reactions calculated at the B97-D3/def2-mSVP level of theory, and 24,000 reactions at the ωB97X-D3/def2-TZVP level of theory. See Figure 13 for a few examples of the reactions considered. This dataset[183] included the forward and backward reactions, and structures contained up to a maximum of seven heavy atoms (carbon, nitrogen, and oxygen). It is not clear whether this model will be transferrable to larger systems, given the small sizes of the molecules in the training set. The DNNs in this model may have been able to learn the underlying principles behind the chemical reactions, and if this is the case, the model may show good performance on larger systems. Two levels of theory were chosen and a transfer learning approach was used to train the model. In this case, transfer learning[184] was performed by initially training the ML model with a large dataset of less accurate, more easily obtainable data and then the model was refined by further training with a more accurate and expensive dataset. The central idea is that the ML model can be well trained for the lower level of theory, given the larger amount of data available, and then a smaller amount of higher-level data can correct the model. This has a similar philosophy to the Δ-ML method (discussed later, Section 3.4), although in this case the low-level trained ML model is corrected by training with high level data, rather than using Δ-ML to correct lower-level calculated properties.

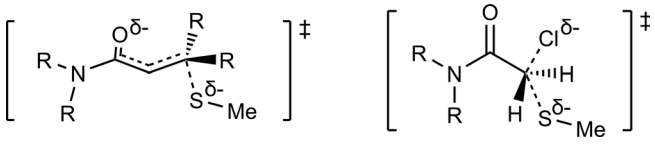


**FIGURE 12** Illustration of the TSs for the reactions between acrylamides (left) and 2-chloroacetamides (right) with the methanethiolate anion that made up the datasets for the training of the BIreactive models

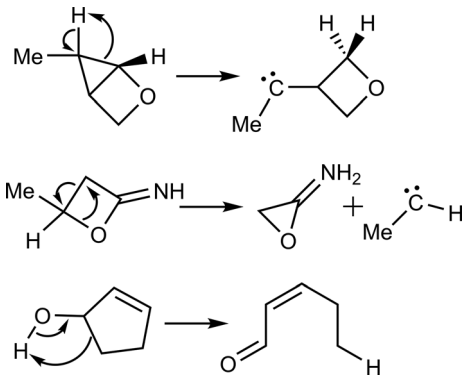


**FIGURE 13** Three examples of the elementary organic reactions used by Grambow et al. to train their DNN models for predictions of activation barriers

The model was able to achieve an impressive MAE of 1.7 kcal $mol^{-1}$ and RMSE of 3.4 kcal $mol^{-1}$ relative to the ωB97X-D3/def2-TZVP activation energies (the reactions spanned a range of 200 kcal $mol^{-1}$, and no QM-derived features were used as inputs for the model). The advantage of this model (i.e., that it can provide predictions of activation energies directly from 2-D molecular representations without the need for expensive QM calculations) also leads to a potential disadvantage. Unlike BIreactive,[145] this model does not consider the effects of conformational flexibility. Should a compound be particularly bulky, important steric effects may not be accounted for by the ML model, as they would be by QM. However, since the molecules included a maximum of seven heavy atoms, this likely did not cause a particularly significant amount of error. ML models that are able to consider the conformation of a chemical system have already been constructed for predicting potential energy surfaces.[66,118,185–187] These methods can use, for example, atom-centered symmetry functions that describe the atomic environments and are invariant with respect to translation or rotation of the system.[65,188] Barring the likely massive data requirements from sampling the potential energy surfaces for thousands of reacting systems with accurate QM methods (ensuring both minima and TSs were well represented), there is no reason why models such as these could not be applied to predicting activation barriers. However, it would seem the approach from Palazzesi et al. of conformational searching the reactants and TSs is a more efficient way to construct a dataset. If this were explored further, reliable and accurate force fields[189–191] will become vital for the conformational searching involved in ML dataset construction.

## 3.3 | Comparisons of different feature types

Along with several ML models, Li et al.[142] tested how model performances varied with different sets of input features. Three types of local features were tested, as well as the combination of each type of local feature with global features. Local features refer to those that are specific to properties or environments of individual atoms, while global features refer to those that represent properties or the structure of an entire molecule. An example of a local feature is a single atom partial charge, whereas a global feature could be the molecular electrostatic polarizability. The local features that described atomic environments based on molecular geometries were atom-centered symmetry functions[188]; smooth overlap of atomic positions (SOAPs), in which atoms have a Gaussian function associated with their positions; and the local environments are given by the integral of the sum of all other distance-weighted atomic-position Gaussians within a given cutoff radius.[135,192,193] Local physical organic (PhysOrg) features were calculated using the B3LYP/6-311+G (2d,p) level of theory and were C—H bond order, atomic charges, and buried volume[194,195] (in this case, the percentage of a sphere, centered on a specific atom, of radius 3 Å that is occupied by the heteroaromatic molecule). The global geometric features were molecular fingerprints,[196,197] where molecules are described by vectors in which each element is a bit corresponding to a substructure of the molecule, and bag of bonds,[198] where a molecule is represented as a vector containing all possible pairwise interactions of atoms. The global PhysOrg features consisted of frontier molecular orbital energies and nucleus-independent chemical shifts[199] (NMR chemical shift at the center of an aromatic ring). Following training, the importances of the local and global PhysOrg features were analyzed, and it was found that both local and global were in the top 15 most important PhysOrg features. This suggests that a sensible direction for future

chemical ML models may involve the inclusion of both molecular and atomic properties as descriptors, since features at both levels may contribute significantly to the predictions of activation barriers.

The model that achieved the greatest $R^2$ value (0.968) between the predicted and DFT-calculated $\Delta\Delta G^{\ddagger}$ values was a GB algorithm with SOAP local features and molecular fingerprints as global features. Since the SOAP/molecular fingerprint descriptors only require molecular geometries to be determined, the heteroaromatic rings were also optimized with the MMFF94 force field[200] and the PM7 semiempirical method,[201] as well as with B3LYP/6-311+G(2d,p). Using the SOAP/molecular fingerprint features, computed from the MMFF94 geometries, produced a slightly higher $R^2$ value (0.975) between the predicted and DFT-calculated $\Delta\Delta G^{\ddagger}$ values. This is rather promising, since these features could be used in a reliable ML model without the need for expensive QM calculations, as needed for the PhysOrg features. The GB model with SOAP/molecular fingerprint features and the RFR model with PhysOrg features were then applied to heteroaromatic systems with substituents that were not present in the training data. RFR was the method that gave the greatest $R^2$ for PhysOrg features.

After retraining the two models with the new systems, the PhysOrg model showed better predictions of $\Delta\Delta G^{\ddagger}$ (lower MAEs) for unseen systems from the new class of heteroaromatics, but the SOAP/molecular fingerprint model predictions did not improve. This suggests that the PhysOrg features are better able to capture the underlying influences that lead to differences in activation energies between the systems, whereas the description from the SOAP/molecular fingerprints could have reached a maximum possible accuracy. This brings up the discussion from the previous section, that is, the molecular features that tend to provide the most reliable predictions of activation energies require expensive QM calculations, which thus limits the model's feasibility for high-throughput applications. However, models trained with input features that are more easily produced may lose predictive accuracy. Perhaps there is the potential for "dual" ML models for the screening of large numbers of reactions: an initial ML model using cheaply obtainable features to provide an estimation of reaction activation energies, then a second model using QM-derived features to provide a better prediction for the reactions of interest from the first model.

Mikami[144] discovered that using so-called "interactive" quantum chemical descriptors (see definitions below) can lead to an improvement in the predictive abilities of ML models compared with using "classical" quantum chemical descriptors. In this case, the ML models were trained to predict the activation barriers of ethylene-metallocene coordination-insertion reactions (see the TS for the reaction in Figure 14). The classical descriptors were obtained from a lone metallocene (pre-reaction) and were LUMO energy, the charge on the metal (from natural population analysis[202]), the metal cation-cyclopentadiene distance, buried volume, and sterimol parameters (that provide quantitative measures of ligand size[203,204]). The "interactive" descriptors were obtained from an ethylene-metallocene complex (Figure 15) and were stabilization energy (energy difference between the starting metallocene and the intermediate ethylene-metallocene complex); and the electrical interaction, charge transfer, and core repulsion energies from natural energy decomposition analysis[207–209] (each of these three energies are a portion of the energy difference between the ethylene-metallocene complex and the unrelaxed, separated ethylene and metallocene, see Figure 15).

Calculations were carried out with 68 group IV metallocenes (Ti, Zr, and Hf) with various cyclopentadiene-based ligands with the B3LYP density functional, the SDD basis set[210–212] for the metal and Si, and the 6-31++G(d,p) basis set for all other atoms. Ten ML models were trained to predict the activation energies of the ethylene-metallocene reactions using only classical features and five were retrained with the interactive descriptors. $R^2$ coefficients between the predicted and DFT-calculated activation energies increased for the training, validation, and testing datasets for all five models. The $R^2$ values for all five models also became much closer to each other. Although, once again, the descriptors used in this work require QM calculations to be obtained. It would be interesting to assess the use of interactive descriptors for other reaction classes, particularly if the descriptors were obtained at low-cost levels of theory.

Döntgen et al. have compared atomic partial charges to bond dissociation energies for their use as descriptors for activation barriers of unimolecular reactions.[149] Reaction activation barrier heights and bond dissociation energies were

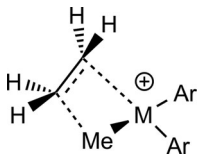


**FIGURE 14** A representation of the TS for the insertion-coordination reaction of ethylene with a cationic metallocene catalyst, studied by Mikami

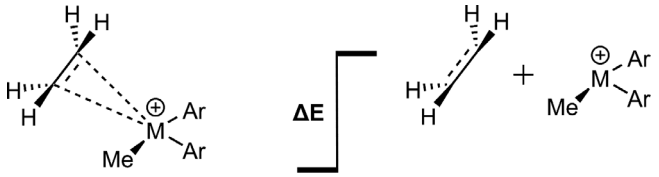


**FIGURE 15** The procedure by which electrical interaction, charge transfer and core repulsion energies are calculated from natural energy decomposition analysis. The energy change ($\Delta E$) is calculated as the energy difference between the ethylene-metallocene intermediate complex (left) and the unrelaxed and separated ethylene and metallocene (right). This energy change is then partitioned into the three energies, based on natural bond order analysis[205,206]

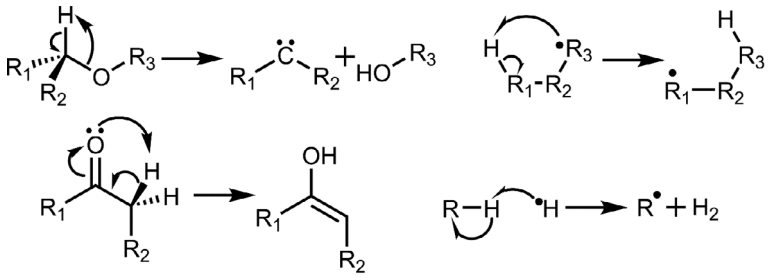


**FIGURE 16** The four reaction types studied by Döntgen et al., shown as generic mechanisms. (Top left) Alkoxy roaming reaction. (Bottom left) Keto-enol tautomerization. (Top right) Internal radical H-atom abstraction. (Bottom right) External radical H-atom abstraction

calculated at the DLPNO-CCSD(T)/CBS(cc-pVTZ, cc-pVQZ) level of theory (CBS refers to extrapolating cc-pVTZ and cc-pVQZ to the complete basis set limit[213]) for 15 alkoxy roaming reactions, 10 keto-enol tautomerizations, 17 external radical H-atom abstractions, and 14 internal radical H-atom abstractions, with geometries optimized at the B3LYP-D3BJ/def2-TZVP level of theory. See Figure 16 for illustrations of the four reactions. One atomic partial charge was used for each reaction that corresponded to the carbon atom at the "reaction center." The bond dissociation energies and partial charges were fit to the activation barrier heights data with linear and/or quadratic equations (i.e., determining the functions that best expressed the relationships between barrier height and the two descriptors for each set of reactions). For the alkoxy roaming reactions and keto-enol tautomerizations, the model relating atomic partial charges to activation barrier was a much better fit than that for bond dissociation energies, as evidenced by the lower MAEs and RMSEs for the partial charge model than the bond dissociation energy model. However, for the two sets of H-atom abstraction reactions, neither the single partial charge nor the bond dissociation energy could be reasonably fit to a polynomial equation. Despite this, for the internal H-atom abstraction reactions where a peroxy radical formed an aldehyde or ketone and an OH· radical, the product of the carbon and oxygen partial charges of the peroxide group had a strong linear correlation with the activation energy of the reactions (with $R^2$ correlation coefficient of 0.9). This indicates that the single partial charge provided insufficient information, but the two partial charges encoded enough information of the reaction mechanism that a better model for activation barrier was achieved. This perhaps indicates that a choice of features based on mechanistic understanding may provide a better description of activation energy. In this case, more than only the carbon atom in the peroxy group was significantly involved in the reaction mechanism and thus the combination of partial charges of atoms involved in the reaction was a superior descriptor.

## 3.4 | Reducing data requirements and interpreting models

Most ML models require reasonable amounts of data to ensure the accuracy of their predictions is as close as possible to the accuracy of the training data. However, highly accurate training data, whether from experiment, or high-level QM calculations, is very expensive to produce and in most instances these data will likely be of a limited scale. Therefore, a ML model that can attain a good performance with only very small amounts of training data will be very desirable indeed. GPR seems to be a good candidate for this task,[60,133,138] and in this section the papers that have used GPR for predicting activation barriers with reduced training sets are reviewed.

Another attempt to address both reducing data requirements and interpreting ML models comes from Friederich et al. with their modeling of hydrogen activation on derivatives of Vaska's complex.[140] The derivatives of Vaska's complex[214,215] that would make up the dataset were defined from the combinations of a chosen set of ligands surrounding an iridium center (Figure 17). See also Figure 18 for a representation of a TS for the hydrogen activation reaction at an example of Vaska's complex. In total, 2574 complexes were possible from the set of ligands (and their allowed positions around the complex), and the ML models were to be trained to predict activation energies of the reaction between the complexes and molecular hydrogen. However, the TSs for only 2197 complexes could be optimized using the PBE density functional,[216] the def2-SVP basis set,[217] and Grimme's D3 dispersion correction.[218]

The first part of the study was to train NNs to predict the activation energies of the reactions. The NNs used autocorrelation and deltametric functions as features. These combine the values of an atomic property for all atoms within a certain number of bonds from the iridium center, through pairwise products (for autocorrelation) or differences (for deltametric). Thus, the functions give unique encodings of the environments within a complex as the number of bonds from the iridium center varies. The atomic properties used were electronegativities, atomic numbers, the number of atoms at a given number of bonds from the center, coordination numbers, and covalent radii. Thus, for example, one feature could be the deltametric function for electronegativities at a distance of two bonds, which would be calculated by the sum of the pairwise differences between the electronegativities of all atoms two bonds from the center. Note that none of these features require QM calculations to be determined, only a two-dimensional representation of the complex is required. The best NN model had a MAE for the test data of 1.12 kcal mol$^{-1}$ when trained on 80% of the training data available; an impressive result. Ideally however, the ML could be able to make predictions with this accuracy (or better) with a reduced set of training data. Hence the second part of their study involved using GPR and feature selection, to ascertain if an accurate ML model could be constructed on a reduced training set. As well as the autocorrelation/deltametric function features, Morgan fingerprints[197,219] (which provide string-based representations of atomic environments taking into account all atoms within a given circular cutoff) and molecular fingerprints from RDKit[220] were also included as potential features for the GPR model. A GB model was used to select which features had the most substantial contributions to activation energy. The GPR model trained on 80% of the data had a MAE of 0.59 kcal mol$^{-1}$, but perhaps more significantly, the GPR was able to reach a MAE below 1 kcal mol$^{-1}$ with only 20% of the training data. This is a significant result; not only did the GPR model show excellent performance with very limited amounts of training data, but it did so with features that were not derived from expensive QM calculations.

After the ML models were trained, the GB method was used to analyze which input features had the greatest importance to the GPR and NN models, thereby allowing interpretation of the factors that lead the models to their predictions for any given complex. For the NN and GPR, the autocorrelation feature corresponding to the atomic numbers of atoms

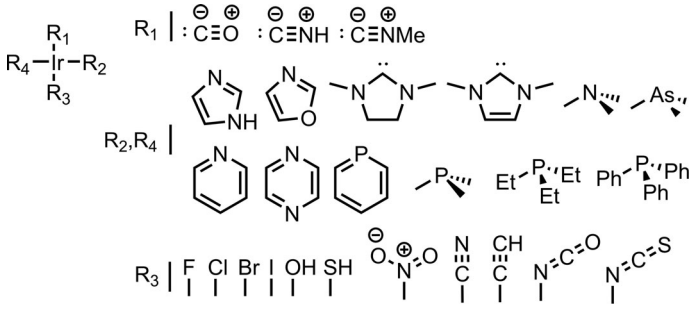


FIGURE 17 The ligands composing the chemical space of the derivatives of Vaska's complex considered by Friederich et al. The R1 ligands are σ-donors and π-acceptors, the R2 and R4 ligands are σ-donors, and the R3 ligands are σ-donors and π-donors

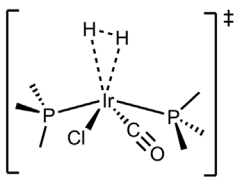


FIGURE 18 An illustration of an example TS for hydrogen activation on the [Ir(PMe$_3$)$_2$(CO)Cl] complex

at a distance of two bonds from the center was the most important. In the NN, autocorrelation functions for atoms at other distances, and the deltametric functions for electronegativities, also had high importances. For GPR, the Morgan and RDKit fingerprints were the other high-importance features. Taken together, the analysis showed that reactivity around the complexes was more largely dependent on electronic effects, rather than steric effects from the ligands. Further in-depth analyses using the fingerprints from the GPR model allowed for interpretation of the effects of changing ligands on activation energy. For example, when a fluoride ligand was replaced with a cyanide, the predicted activation energy decreased by 14.4 kcal $mol^{-1}$ which was due to the electron withdrawing character of the cyanide, leading to less repulsion between the iridium center and the approaching hydrogen. Friederich et al. note that this type of interpretation of ML models can be used for the screening and design of new potential catalysts.[140]

Jorner et al. also utilized GPR in low data circumstances[146] when they created a "hybrid" model, in which a ML model was trained on experimental data, and DFT was used to calculate the features for the model. The rate constants for 443 nucleophilic aromatic substitution ($S_NAr$) reactions were obtained from the literature, see Figure 19 for an illustration of a selection of the reactions. GPR models were trained to reproduce these rate constants and thus, from the predictions, activation barriers could be calculated. Reactants and TSs for each reaction were conformationally searched with the GFN2-xTB semiempirical method,[221] and the lowest energy conformers from the searches were optimized at the ωB97X-D/6-31+G(d) level of theory, with single-point energies calculated at the ωB97X-D/6-311+G(d,p) level of theory. Electronic properties of the reactants and TSs were calculated with the B3LYP/6-31+G(d) level of theory and used as features for the models. Reactant features were ionization energies; electron attachment energies[222]; the surface electrostatic potential and electrostatic potentials at reactive nuclei[223,224]; local and global descriptors of nucleophilicity and electrophilicity[225]; atomic charges; bond orders; solvent accessible surface areas[226]; and a descriptor for dispersion.[227] TS features were the activation barrier of the reaction as calculated with the ωB97X-D/6-311+G(d,p) level of theory; TS atomic charges; TS bond orders; and TS nuclear electrostatic potentials. Separate GPR models were trained using different sets of the electronic (DFT-calculated) features. One set used all the electronic features from reactants and TSs, another without any TS features, and another using a small subset of the features, including the DFT calculated activation energy. GPR models were also trained with Morgan fingerprints, as well as "In Silico Design and Data Analysis" descriptors[228] (which encode molecules and reactants based on the fragments that make up the systems). The final feature set was a deep learning encoding of the reactions using a Bidirectional Encoder Representations from Transformers (BERT) classifier.[229] The DNN in the BERT classifier is trained to create fingerprint representations of the reactions, and these learned fingerprint representations were used as input features for the barrier ML models. Overall, GPR trained on the full set of electronic features performed best, with a MAE of 0.77 kcal $mol^{-1}$.

Since the hybrid model proposed by Jorner et al. uses experimental data, and most experimental datasets are likely to be of reduced size compared to those constructed from DFT, it will be crucial for the ML method in this type of model to be able to provide reliable results with restricted amounts of data. As also seen in the study by Friederich et al.,[140] GPR seems to fill this limited-data role very well; it was able to reach a MAE less than 1 kcal $mol^{-1}$ with less

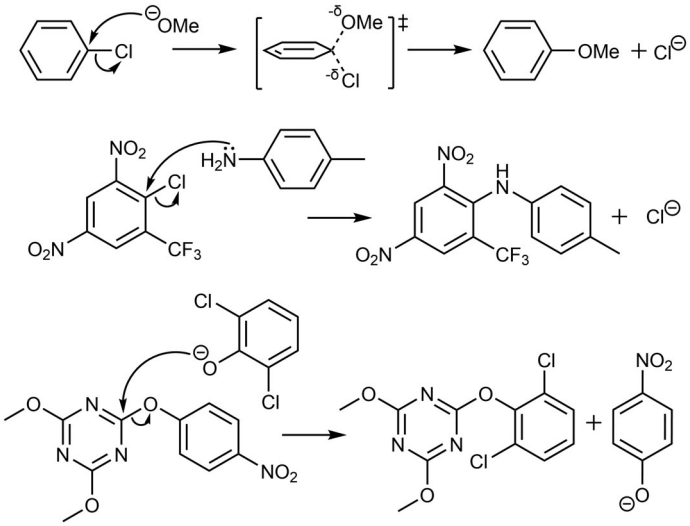


FIGURE 19 Three of the $S_NAr$ reactions modeled by Jorner et al. from their experimental dataset

than 120 training samples and was also found to outperform LR, RFR, and SVR models trained with the same experimental rate constants with the full set of electronic features. The major advantage of this hybrid method is that the model is not confined to the accuracy of DFT; the limit of the model's accuracy is experiment itself. It also represents a useful method in the case when a limited amount of training data is available for ML (e.g., when using experimental data). In this case, the additional cost of DFT calculations for input features is justified by the greatly improved accuracy of the ML model.

Jorner et al. interpreted the model by calculating the Spearman coefficients between the model predictions and the features. This found, as may be expected, that the DFT calculated activation barrier was the most important feature to the predictions. However, the GPR model that did not include any features from the TSs still had a very respectable MAE of 0.86 kcal $mol^{-1}$, and the MAE was still lower than 1 kcal $mol^{-1}$ with under 200 training samples. Thus, not including TS features did not lead to massively lowered accuracy for the $S_NAr$ reactions. The model was still able to give chemical accuracy with only a modest amount of extra training data. The model based on the deep learning BERT features was also able to reach chemical accuracy, although with just over 350 training samples. In addition, the BERT features only require the connectivity of the systems in the reactions to be created, with no QM calculations needed. With a rather small number of datapoints, the risk of overfitting is apparent, although given the very strong performance of this model this may not be an issue. However, it is still worth guarding against overfitting when creating ML models, since this will deteriorate the general performance of the model, and one should always aim to use the largest dataset that is feasible to obtain.

There has recently been an increasing amount of attention paid to the interpretation of ML models, especially to avoid so-called "black box" approaches (especially DNNs), where the ML methods can make useful predictions, but their complexity means that it is not possible for humans to grasp exactly why the model came to its conclusions.[230] An argument can be made[231] that a better approach is to construct explainable, more easily understood models (such as LR), rather than to interpret feature importances after a more complex model has been trained. However, at least in the context of predicting activation barriers, LR models do not typically match the performance of more advanced algorithms such as NNs, GPR, and KRR.[140,142,144,146,158,159] This is not to say that LR models are irrelevant; their accuracies can be reasonable when trained with suitable features, but future workers attempting to predict activation barriers with ML should consider that a potential trade-off between model performance and interpretability may occur.

## 3.5 | Δ-ML for activation energy prediction

Δ-ML is a technique developed by Ramakrishnan et al.[232] where an ML model is trained to predict a correction for a property calculated at a low level of theory (known as the baseline). Rather than training the model to directly predict the high-level (targetline) value of the property, Δ-ML models are trained on the difference between the baseline and targetline values. The Δ-ML model is thus able to correct a property calculated at a low level of theory closer to its value at a higher level of theory (represented graphically in Figure 20).

Hammett's equation is notable for its simultaneous simplicity and predictive ability.[233–237] However, Hammett's original method is saddled with some significant disadvantages. The values of the substituent constants are highly dependent on the choice of reference reaction and may not be reliable for other reactions.[238–240] Hammett's method also easily overfits towards the reference reaction, meaning its predictions will be good for the reference reaction, but poorer for others. Thus, it was the work of Bragato et al.[148] to form a new Hammett method such that the substituent constants could become more transferable and the overall model be made more reliable for different reactions. This was achieved by calculating all reaction constants at the same time and calculating each substituent constant from its average over all reactions. This means the substituent constants give, generally speaking, better predictions for all reactions in the dataset rather than being biased towards the reference.

For a dataset of approximately 2400 $S_N2$ reactions (Figure 21), with MP2 calculated activation energies, Bragato et al. created Hammett models based on both the old and new methods as well as a ML and a Δ-ML model, to predict the activation energies of the $S_N2$ reactions. Note that the Hammett model is still applicable to activation energies since, from TS theory, activation energy is directly proportional to the negative logarithm of rate constant.

Their chosen ML model was KRR, and it was trained on inputs of one-hot strings (zeros and ones only) encoded to represent the functional groups and their positions, with reaction activation energy as output. As may be expected, KRR was able to outperform both Hammett models. When plotted, the MAEs between the predicted and the MP2

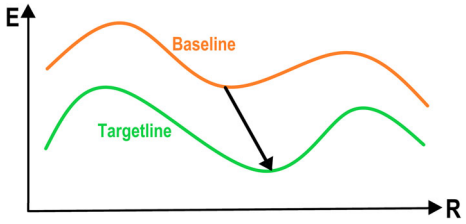


**FIGURE 20** Illustration of how Δ-ML can make predictions of chemical properties (E) in chemical feature space (R). Δ-ML is trained to predict the difference between the baseline (orange) and targetline (green) values of the property of interest. This predicted difference can then be used to correct the baseline value closer to the accuracy of the targetline (black arrow)

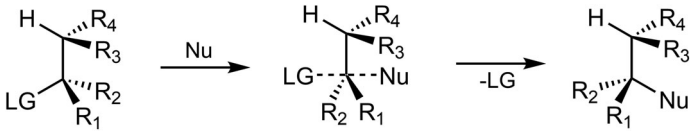


**FIGURE 21** $S_N2$ reactions with activation energies calculated at the MP2 level of theory used by Bragato et al. for training of the Hammett and ML methods. Nucleophiles and leaving groups can be any of H, F, Cl and Br. R-groups (substituents) can be any of H, $NO_2$, CN, $NH_3$ and $CH_3$

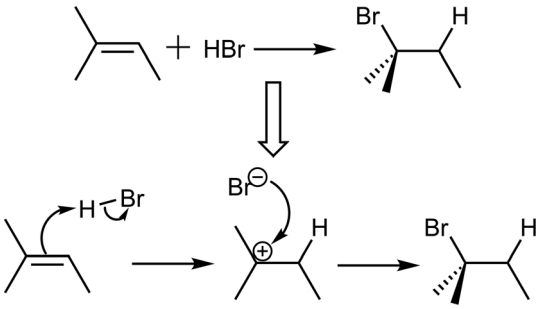


**FIGURE 22** Illustration of how ReactionPredictor breaks down an overall reaction (top) into a sequence of elementary steps (bottom). For a given reaction, the most favorable elementary steps are selected by the DNN and the entire reaction is modeled from start to completion with DFT

activation energies were consistently lower for KRR than the new Hammett model and much lower than the original as the number of training points was increased. The greater performance of KRR is due to its greater flexibility; it can account for deviations from linearity in the data that the linear Hammett models cannot. However, KRR does have the disadvantage of being much more complex and requiring more data compared to the Hammett model. For the Hammett equation, only one parameter is needed per substituent, per possible position, but for KRR a parameter must be determined for every training point.

The Δ-ML also used KRR and was trained to predict the difference between the activation energies from the new Hammett method as a baseline, and the MP2 activation energies as the targetline. The "normal" ML and Δ-ML models eventually reached the same error (MAEs just greater than 1.5 kcal $mol^{-1}$) with enough training points, but Δ-ML consistently had lower MAEs for all numbers of training points. This has been noted as an advantage of the Δ-ML approach.[232,241] The function that describes the error from a lower level of quantum chemical theory, relative to a higher level is apparently smoother[232] (and hence easier for ML algorithms to interpolate its form), compared to the function describing the relationship between the chemical system and its properties.

Δ-ML makes for an effective combination of QM and ML. It is able to reach the same level of accuracy as a ML model trained with molecular descriptors, but with reduced data requirements due to the smoother and more easily

**TABLE 1** Summary of the works discussed in this review using ML to predict reaction activation barriers

| Reference | Authors | Reaction types | Dataset details | Barrier source | Best model | Best test accuracy | Important features |
|---|---|---|---|---|---|---|---|
| 139 | Choi et al. | Simple gas-phase organic reactions (Figure 9) | Total: 12,704, train: 80%, test: 20% | Computation: B3LYP/6-31G* | GB | MAE: 1.95 kcal $mol^{-1}$; RMSE: 4.49 kcal $mol^{-1}$; $R^2$: 0.89 | Molecular fingerprints, topological indices, thermodynamic quantities. |
| 135 | Li et al. | Radical C—H bond functionalization of heteroarenes (Figure 10) | Total: 3406 | Computation: M06-2X/aug-cc-pVTZ //B3LYP/6-311+G(2d,p) | GB | $R^2$: 0.97 | Physical organic features (atomic charges, molecular orbital energies, bond orders, percent buried volume, chemical shifts), SOAPs. |
| 143 | Migliaro and Cundari | Hydride/proton abstraction with frustrated Lewis pairs (Figure 11) | Total: 35, train: 70%, validation: 15%, test: 15% | Computation: ωB97X-D/def2-TZVPP/SMD-MeCN | NN | $R^2$: 0.85 | Global electrophilicity index, bond dissociation energies, reaction energies. |
| 141 | Grambow et al. | Intramolecular organic reactions (Figure 13) | Total: 33,000 (lower level) and 24,000 (higher level), train: 85%, validation: 5%, test: 10% | Computation: B97-D3/def2-mSVP and ωB97X-D3/def2-TZVP | DNN | MAE: 1.7 kcal $mol^{-1}$; RMSE: 3.4 kcal $mol^{-1}$ | Atom, bond and molecule features from RDKit.[220] |
| 145 | Palazzesi et al. | Acrylamides and 2-chloroacetamides reacting with methanthiolate anion (Figure 12) | Total: 2083 acrylamides and 2716 2-chloroacetamides, train: 80%, test: 20% | Computation: ωB97XD/cc-pVDZ | Extra-tree regression | MAE: 0.75 kcal $mol^{-1}$ and 0.88 kcal $mol^{-1}$; $R^2$: 0.85 and 0.69; Spearman: 0.91 and 0.84 | Dragon[179] descriptors. |
| 144 | Mikami | Insertion-coordination reactions of ethylene with metallocene catalysts (Figure 14) | Total: 68, train: 66%, validation: 24%, test: 20% | Computation: B3LYP/6-31++G(d,p) and B3LYP/SDD | GB | $R^2$: 0.84 | "Interactive-quantum-chemical-descriptors" extracted from natural energy decomposition analysis. Calculations. |
| 149 | Döntgen et al. | Alkoxy roaming reactions, keto-enol tautomerisation, radical H-atom abstractions (Figure 16) | Total: 15 alkoxy roaming, 10 keto-enol, 17 external H-atom abstractions and 14 internal H-atom abstractions | Computation: DLPNO-CCSD(T)/CBS(cc-pVTZ, cc-pVQZ) | Polynomial equation | MAE: 2.3 kcal $mol^{-1}$, 0.2 kcal $mol^{-1}$, 1.1 kcal $mol^{-1}$; RMSD: 1.0 kcal $mol^{-1}$, 0.1 kcal $mol^{-1}$, 1.1 kcal $mol^{-1}$ | Atomic partial charges. |

**TABLE 1** (Continued)

| Reference | Authors | Reaction types | Dataset details | Barrier source | Best model | Best test accuracy | Important features |
|---|---|---|---|---|---|---|---|
| 140 | Friederich et al. | Hydrogen activation on Vaska's complex (Figure 17) | Total: 2197, train: 80%, validation: 10%, test: 10% | Computation: PBE-D3/def2-SVP | GPR | MAE: 0.59 kcal $mol^{-1}$; $R^2$: 0.95 | Electronegativity, atomic number, distance from metal center, atomic covalent radii, molecular fingerprints. |
| 146 | Jorner et al. | Nucleophilic aromatic substitutions (Figure 19) | Total: 443, train: 80%, test: 20% | Experiment: rate constants from the literature | GPR | MAE: 0.77; RMSE: 1.01; $R^2$: 0.93 | DFT-calculated activation barrier, atomic charges, bond orders, nucleophilicity, electrophilicity, solvent accessible surface area, dispersion. |
| 148 | Bragato et al. | $S_N2$ reactions (Figure 21) | Total: 2400 | Computation: MP2 | KRR (Δ-ML) | MAE: 1.6 kcal $mol^{-1}$ | Molecular fingerprints. |
| 147 | Sadowski et al. | Elementary organic reactions (Figure 22) | Total: 6000, train: 90%, test: 10% | Computation: DFT | DNN | — | Molecular fingerprints. |

*Note*: The "reaction types" column briefly describes the chemical reactions for which barriers were predicted by each of the references. The "dataset details" column provides the total number of datapoints used as well as the proportions of those data points used for training, validation and testing (if provided). The source from which the activation barriers were extracted are given in the "barrier source" column, and the best ML model and its performance metrics (if provided) are in the columns "best model" and "best test accuracy." For each study, the most relevant molecular descriptors used in the best predictions of activation barriers are listed in the "important features" column.

learnt relationship between the baseline and targetline properties. This allows for QM reaction modeling studies to be performed at increased speed and scale.

### 3.6 | QM-ML synergy

An interesting example of a somewhat indirect use of reaction activation energies with ML came from Sadowski et al. and their software ReactionPredictor.[147] The program was designed to be able to predict the likely outcomes of chemical reactions through a synergy of QM and ML. The procedure for this task was as follows: when given a set of reactants, electron sources and sinks are identified[242,243] and a DNN ranks all possible electron transfers between source–sink pairs (corresponding to an elementary step in a reaction, see also Figure 22) in order of their favorability. After ranking by the DNN, the reactants, products, and TSs for the top few reaction pathways are optimized with DFT and the activation energies for the elementary reactions are calculated. Hence the overall mechanism of a reaction can be determined based on which elementary steps are most favorable, as ranked by the DNN.

In this approach, the synergy between ML and QM occurs when the DNN has incorrectly ranked the reactions, as will be discovered after QM calculations. In that case the DNN is retrained with the correct ordering for these reactions, and thus its predictions are brought closer in line with QM. The ReactionPredictor system was later tested by Fooshee et al.[244] on a new dataset of reactions and achieved an accuracy of 80% for correctly predicting the products of an unseen benchmark dataset of multistep reactions.

The task of ReactionPredictor requires a great deal of flexibility from its ML model. Not only does the model have to interpret the input for a given reaction, but it also needs to be able to learn the underlying principles behind how organic reaction mechanisms proceed, based only on elementary reaction steps. DNNs are optimally suited to this type of problem. DNNs appear to gain some understanding of the underlying chemistry of the systems in their training sets. This is also seen when DNNs are used to predict potential energies of chemical systems[66,185,245]; when trained on data from smaller systems, they are still able to make accurate predictions for larger systems.

## 4 | SUMMARY AND CONCLUSION

This review has examined the work using ML to predict activation energy barriers of homogeneous chemical reactions. Using ML models for activation barrier prediction is a relatively new, rapidly evolving area of study with all studies presented having being published within 5 years of the time of writing. Please see Table 1 for a summary of the key information from these studies. NNs have been seen to be flexible and effective ML techniques for this task, especially DNNs, and their use in the context of these works was discussed. A couple of potential limitations in some studies were noted. First, models using training data that consisted mostly of similar reactions would likely display very limited ability for generalization, but even more prominently was the use of input features derived from QM calculations. Although the ML models mean that difficult TS structures need not be found with QM to make predictions, the purpose of using ML in chemistry is to reduce the need for expensive QM calculations. However, also seen are models that exclusively use geometrical features as inputs, and still give very promising results. On the other hand, these models will not be able to offer the same level of mechanistic insight that could be obtained by analyzing feature importances in a model using QM-derived features.

As the use of ML in computational chemistry increases, the need for large datasets of chemical properties generated by QM calculations will also increase. It will therefore be a worthy endeavor to create tools that are capable of performing high-throughput QM calculations for a wide variety of different reaction types.[246] Attention has also been drawn to methods that have a reduced requirement for the amount of training data needed, and those that make attempts to interpret ML models. It is debatable whether it is better to use simpler models that can be easily interpreted, or to construct less comprehensible models that (at least in the context of activation energies) give more reliable results and interpret them by computing feature importances from the fully trained models. Δ-ML has also been seen to be a very effective technique for the acceleration of activation energy evaluation. Given the very recent influx of work using ML for reaction barrier modeling, there is little doubt that there is much work to be done towards the goal of general, rapid and chemically accurate activation energies.

### CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

## AUTHOR CONTRIBUTIONS

**Toby Lewis-Atwell:** Conceptualization (equal); investigation (lead); methodology (equal); writing – original draft (lead). **Piers A. Townsend:** Conceptualization (equal); investigation (supporting); project administration (supporting); supervision (equal); writing – review and editing (equal). **Matthew N. Grayson:** Conceptualization (lead); investigation (supporting); methodology (equal); project administration (lead); supervision (lead); writing – review and editing (equal).

## DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

## ORCID

*Toby Lewis-Atwell* https://orcid.org/0000-0002-0918-8754
*Piers A. Townsend* https://orcid.org/0000-0002-7164-7958
*Matthew N. Grayson* https://orcid.org/0000-0003-2116-7929

## RELATED WIREs ARTICLES

Machine learning methods in chemoinformatics
Making machine learning a useful tool in the accelerated discovery of transition metal complexes
Learning molecular potentials with neural networks

## REFERENCES

1. Peng Q, Duarte F, Paton RS. Computing organic stereoselectivity from concepts to quantitative calculations and predictions. Chem Soc Rev. 2016;45(22):6093–107. https://doi.org/10.1039/c6cs00573j
2. Engkvist O, Norrby P-O, Selmi N, Lam Y-H, Peng Z, Sherer EC, et al. Computational prediction of chemical reactions: current status and outlook. Drug Discov Today. 2018 Jun;23(6):1203–18. https://doi.org/10.1016/j.drudis.2018.02.014
3. Y-hong L, Abramov Y, Ananthula RS, Elward JM, Hilden LR, SON L, et al. Applications of quantum chemistry in pharmaceutical process development: current state and opportunities. Org Process Res Dev. 2020 Jul;24(8):1496–507. https://doi.org/10.1021/acs.oprd.0c00222
4. Battin-Leclerc F, Blurock E, Bounaceur R, Fournet R, Glaude P-A, Herbinet O, et al. Towards cleaner combustion engines through groundbreaking detailed chemical kinetic models. Chem Soc Rev. 2011;40(9):4762–82. https://doi.org/10.1039/c0cs00207k
5. Vogel P, Houk KN. Organic chemistry: theory, reactivity and mechanisms in modern synthesis. Weinheim, Germany: Wiley-VCH; 2019.
6. Mardirossian N, Head-Gordon M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. Mol Phys. 2017 Jun;115(19):2315–72. https://doi.org/10.1080/00268976.2017.1333644
7. Houk K, Paddon-Row M, Rondan N, Wu Y, Brown F, Spellmeyer D, et al. Theory and modeling of stereoselective organic reactions. Science. 1986 Mar;231(4742):1108–17. https://doi.org/10.1126/science.3945819
8. Lam Y-h, Grayson MN, Holland MC, Simon A, Houk KN. Theory and modeling of asymmetric catalytic reactions. Acc Chem Res. 2016 Mar;49(4):750–62. https://doi.org/10.1021/acs.accounts.6b00006
9. Domingo L, Ríos-Gutiérrez M, Pérez P. Applications of the conceptual density functional theory indices to organic chemistry reactivity. Molecules. 2016 Jun;21(6):748. https://doi.org/10.3390/molecules21060748
10. Falcone BN, Grayson MN, Rodriguez JB. Mechanistic insights into a chiral phosphoric acid-catalyzed asymmetric pinacol rearrangement. J Org Chem. 2018 Nov;83(23):14683–7. https://doi.org/10.1021/acs.joc.8b02812
11. Momo PB, Leveille AN, Farrar EHE, Grayson MN, Mattson AE, Burtoloso ACB. Enantioselective S–H insertion reactions of α-carbonyl sulfoxonium ylides. Angew Chem. 2020 May;132(36):15684–9. https://doi.org/10.1002/ange.202005563
12. Zhou T, Huang D, Caflisch A. Quantum mechanical methods for drug design. Curr Top Med Chem. 2010 Jan;10(1):33–45. https://doi.org/10.2174/156802610790232242
13. LaPointe S, Weaver D. A review of density functional theory quantum mechanics as applied to pharmaceutically relevant systems. Curr Comput Aided Drug Des. 2007 Dec;3(4):290–6. https://doi.org/10.2174/157340907782799390
14. Rydberg P, Jørgensen FS, Olsen L. Use of density functional theory in drug metabolism studies. Expert Opin Drug Metab Toxicol. 2013 Dec;10(2):215–27. https://doi.org/10.1517/17425255.2014.864278
15. Townsend PA, Grayson MN. Reactivity prediction in aza-Michael additions without transition state calculations: the Ames test for mutagenicity. Chem Commun. 2020;56(88):13661–4. https://doi.org/10.1039/d0cc05681b
16. Townsend PA, Grayson MN. Density functional theory transition-state modeling for the prediction of Ames mutagenicity in 1,4 Michael acceptors. J Chem Inf Model. 2019 Nov;59(12):5099–103. https://doi.org/10.1021/acs.jcim.9b00966
17. Lim S, Choi J-W, Suh DJ, Song KH, Ham HC, Ha J-M. Combined experimental and density functional theory (DFT) studies on the catalyst design for the oxidative coupling of methane. J Catal. 2019 Jul;375:478–92. https://doi.org/10.1016/j.jcat.2019.04.008

18. Rosen AS, Notestein JM, Snurr RQ. Identifying promising metalorganic frameworks for heterogeneous catalysis via high-throughput periodic density functional theory. J Comput Chem. 2019 Feb;40(12):1305–18. https://doi.org/10.1002/jcc.25787
19. Uhe A, Hölscher M, Leitner W. Carboxylation of arene C–H bonds with $CO_2$: a DFT-based approach to catalyst design. Chem Eur J. 2011 Dec;18(1):170–7. https://doi.org/10.1002/chem.201102785
20. Jensen F. Introduction to computational chemistry. Chichester: Wiley; 2017.
21. Harvey J. Computational chemistry. Oxford: Oxford University Press; 2018.
22. Lewars EG. Computational chemistry: introduction to the theory and applications of molecular and quantum mechanics. Cham, Switzerland: Springer; 2016.
23. Eksterowicz JE, Houk KN. Transition-state modeling with empirical force fields. Chem Rev. 1993 Nov;93(7):2439–61. https://doi.org/10.1021/cr00023a006
24. Jensen F. Locating minima on seams of intersecting potential energy surfaces. An application to transition structure modeling. J Am Chem Soc. 1992 Feb;114(5):1596–603. https://doi.org/10.1021/ja00031a009
25. Åqvist J, Warshel A. Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches. Chem Rev. 1993 Nov;93(7):2523–44. https://doi.org/10.1021/cr00023a010
26. Hansen E, Rosales AR, Tutkowski B, Norrby P-O, Wiest O. Prediction of stereochemistry using Q2MM. Acc Chem Res. 2016 Apr;49(5):996–1005. https://doi.org/10.1021/acs.accounts.6b00037
27. van Duin AC, Dasgupta S, Lorant F, Goddard WA. ReaxFF: a reactive force field for hydrocarbons. J Phys Chem A. 2001 Sep;104(41):9396–409. https://doi.org/10.1021/jp004368u
28. Lewis-Atwell T, Townsend PA, Grayson MN. Comparisons of different force fields in conformational analysis and searching of organic molecules: a review. Tetrahedron. 2021 Jan;79:131865. https://doi.org/10.1016/j.tet.2020.131865
29. Hastie T, Tibshirani R, Freidman J. The elements of statistical learning: data mining, inference, and prediction. New York: Springer; 2009.
30. Trappenberg TP. Fundamentals of machine learning. New York: Oxford University Press; 2020.
31. Rupp M. Machine learning for quantum mechanics in a nutshell. Int J Quantum Chem. 2015 Jul;115(16):1058–73. https://doi.org/10.1002/qua.24954
32. Strieth-Kalthoff F, Sandfort F, Segler MHS, Glorius F. Machine learning the ropes: principles applications and directions in synthetic chemistry. Chem Soc Rev. 2020;49(17):6154–68. https://doi.org/10.1039/c9cs00786e
33. Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. Nature. 2018 Jul;559(7715):547–55. https://doi.org/10.1038/s41586-018-0337-2
34. Zaspel P, Huang B, Harbrecht H, von Lilienfeld OA. Boosting quantum machine learning models with a multilevel combination technique: Pople diagrams revisited. J Chem Theory Comput. 2018 Dec;15(3):1546–59. https://doi.org/10.1021/acs.jctc.8b00832
35. Jorner K, Tomberg A, Bauer C, Sköld C, Norrby P-O. Organic reactivity from mechanism to machine learning. Nat Rev Chem. 2021 Mar;5(4):240–55. https://doi.org/10.1038/s41570-021-00260-x
36. WDV H. Evaluation of machine learning methods for natural language processing tasks. LREC 2002: third international conference on language resources and evaluation. Paris, France: European Language Resources Association (ELRA); 2002.
37. Collier R. An historical overview of natural language processing systems that learn. Artif Intell Rev. 1994;8(1):17–54. https://doi.org/10.1007/bf00851349
38. Daelemans W, Hoste V, Meulder FD, Naudts B. Combined optimization of feature selection and algorithm parameters in machine learning of language. Machine learning: ECML 2003. Berlin Heidelberg: Springer; 2003. p. 84–95. https://doi.org/10.1007/978-3-540-39857-8_10
39. Sidorov G, Velasquez F, Stamatatos E, Gelbukh A, Chanona-Hernández L. Syntactic N-grams as machine learning features for natural language processing. Expert Syst Appl. 2014 Feb;41(3):853–60. https://doi.org/10.1016/j.eswa.2013.08.015
40. Kabzan J, Hewing L, Liniger A, Zeilinger MN. Learning-based model predictive control for autonomous racing. IEEE Robot Automat Lett. 2019 Oct;4(4):3363–70. https://doi.org/10.1109/lra.2019.2926677
41. Hodges C, An S, Rahmani H, Bennamoun M. Deep learning for driverless vehicles. Handbook of deep learning applications. Springer, Cham: Springer International Publishing; 2019. p. 83–99. https://doi.org/10.1007/978-3-030-11479-4_4
42. Vishnukumar HJ, Butting B, Muller C, Sax E. Machine learning and deep neural network artificial intelligence core for lab and real-world test and validation for ADAS and autonomous vehicles: AI for efficient and quality test and validation. 2017 intelligent systems conference (IntelliSys). London: IEEE; 2017. https://doi.org/10.1109/intellisys.2017.8324372
43. Gang W. Safety evaluation model for smart driverless car using support vector machine. J Intell Fuzzy Syst. 2019 Jul;37(1):433–40. https://doi.org/10.3233/jifs-179098
44. Karthikeyan M, Sathiamoorthy S, Vasudevan M. Lane keep assist system for an autonomous vehicle using support vector machine learning algorithm. Innovative data communication technologies and application. Cham, Switzerland: Springer International Publishing; 2020. p. 101–8. https://doi.org/10.1007/978-3-030-38040-3_11
45. Deng L, Li X. Machine learning paradigms for speech recognition: An overview. IEEE Trans Audio Speech Lang Process. 2013 May;21(5):1060–89. https://doi.org/10.1109/tasl.2013.2244083
46. Padmanabhan J, Premkumar MJJ. Machine learning in automatic speech recognition: a survey. IETE Tech Rev. 2015 Feb;32(4):240–51. https://doi.org/10.1080/02564602.2015.1010611
47. Nassif AB, Shahin I, Attili I, Azzeh M, Shaalan K. Speech recognition using deep neural networks: a systematic review. IEEE Access. 2019;7:19143–65. https://doi.org/10.1109/access.2019.2896880

48. Ganapathiraju A, Hamaker J, Picone J. Applications of support vector machines to speech recognition. IEEE Trans Signal Process. 2004 Aug;52(8):2348–55. https://doi.org/10.1109/tsp.2004.831018
49. Adankon MM, Cheriet M. Model selection for the LS-SVM. Application to handwriting recognition. Pattern Recogn. 2009 Dec;42(12): 3264–70. https://doi.org/10.1016/j.patcog.2008.10.023
50. Pham V, Bluche T, Kermorvant C, Louradour J. Dropout improves recurrent neural networks for handwriting recognition. 2014 14th international conference on frontiers in handwriting recognition. Hersonissos, Greece: IEEE; 2014. https://doi.org/10.1109/icfhr.2014.55
51. Joshi P, Agarwal A, Dhavale A, Suryavanshi R, Kodolikar S. Handwriting analysis for detection of personality traits using machine learning approach. Int J Comput Appl. 2015 Nov;130(15):40–5. https://doi.org/10.5120/ijca2015907189
52. An L, Bhanu B. Image super-resolution by extreme learning machine. 2012 19th IEEE international conference on image processing. Orlando, FL: IEEE; 2012. https://doi.org/10.1109/icip.2012.6467333
53. Tang Y, Yan P, Yuan Y, Li X. Single-image super-resolution via local learning. Int J Mach Learn Cybern. 2011 Feb;2(1):15–23. https://doi.org/10.1007/s13042-011-0011-6
54. Yang W, Zhang X, Tian Y, Wang W, Xue J-H, Liao Q. Deep learning for single image super-resolution: a brief review. IEEE Trans Multimedia. 2019 Dec;21(12):3106–21. https://doi.org/10.1109/tmm.2019.2919431
55. Wang Z, Chen J, Hoi SCH. Deep learning for image super-resolution: a survey. IEEE Trans Pattern Anal Mach Intell. 2021 Oct;43(10): 3365–87. https://doi.org/10.1109/tpami.2020.2982166
56. Mosavi A, Varkonyi A. Learning in robotics. Int J Comput Appl. 2017 Jan;157(1):8–11. https://doi.org/10.5120/ijca2017911661
57. Sünderhauf N, Brock O, Scheirer W, Hadsell R, Fox D, Leitner J, et al. The limits and potentials of deep learning for robotics. Int J Robot Res. 2018 Apr;37(4–5):405–20. https://doi.org/10.1177/0278364918770733
58. Dorigo M, Schnepf U. Genetics-based machine learning and behavior-based robotics: a new synthesis. IEEE Trans Syst Man Cybern. 1993;23(1):141–54. https://doi.org/10.1109/21.214773
59. Kassahun Y, Yu B, Tibebu AT, Stoyanov D, Giannarou S, Metzen JH, et al. Surgical robotics beyond enhanced dexterity instrumentation: a survey of machine learning techniques and their role in intelligent and autonomous surgical actions. Int J Comput Assist Radiol Surg. 2015 Oct;11(4):553–68. https://doi.org/10.1007/s11548-015-1305-z
60. Deisenroth MP, Fox D, Rasmussen CE. Gaussian processes for data-efficient learning in robotics and control. IEEE Trans Pattern Anal Mach Intell. 2015 Feb;37(2):408–23. https://doi.org/10.1109/tpami.2013.218
61. Campbell M, Hoane AJ, Hsu F-h. Deep blue. Artif Intell. 2002 Jan;134(1–2):57–83. https://doi.org/10.1016/s0004-3702%2801%2900129-1
62. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, et al. Mastering the game of Go with deep neural networks and tree search. Nature. 2016 Jan;529(7587):484–9. https://doi.org/10.1038/nature16961
63. Ho T-S, Rabitz H. A general method for constructing multidimensional molecular potential energy surfaces from ab initio calculations. J Chem Phys. 1996 Feb;104(7):2584–97. https://doi.org/10.1063/1.470984
64. Lorenz S, Groß A, Scheffler M. Representing high-dimensional potential-energy surfaces for reactions at surfaces by neural networks. Chem Phys Lett. 2004 Sep;395(4–6):210–5. https://doi.org/10.1016/j.cplett.2004.07.076
65. Behler J, Parrinello M. Generalized neural-network representation of high-dimensional potential-energy surfaces. Phys Rev Lett. 2007 Apr;98(14):146401. https://doi.org/10.1103/physrevlett.98.146401
66. Smith JS, Isayev O, Roitberg AE. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. Chem Sci. 2017;8(4):3192–203. https://doi.org/10.1039/c6sc05720a
67. Westermayr J, Gastegger M, Marquetand P. Combining SchNet and SHARC: the SchNarc machine learning approach for excited-state dynamics. J Phys Chem Lett. 2020 Apr;11(10):3828–34. https://doi.org/10.1021/acs.jpclett.0c00527
68. Dral PO, Barbatti M, Thiel W. Nonadiabatic excited-state dynamics with machine learning. J Phys Chem Lett. 2018 Sep;9(19):5660–3. https://doi.org/10.1021/acs.jpclett.8b02469
69. Han J, Zhang L, E W. Solving many-electron Schrödinger equation using deep neural networks. J Comput Phys. 2019 Dec;399:108929. https://doi.org/10.1016/j.jcp.2019.108929
70. Hermann J, Schätzle Z, Noé F. Deep-neural-network solution of the electronic Schrödinger equation. Nat Chem. 2020 Sep;12(10): 891–7. https://doi.org/10.1038/s41557-020-0544-y
71. Schütt KT, Gastegger M, Tkatchenko A, Müller K-R, Maurer RJ. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. Nat Commun. 2019 Nov;10(1):5024. https://doi.org/10.1038/s41467-019-12875-2
72. Gastegger M, McSloy A, Luya M, Schütt KT, Maurer RJ. A deep neural network for molecular wave functions in quasi-atomic minimal basis representation. J Chem Phys. 2020 Jul;153(4):044123. https://doi.org/10.1063/5.0012911
73. Pozun ZD, Hansen K, Sheppard D, Rupp M, Müller K-R, Henkelman G. Optimizing transition states via kernel-based machine learning. J Chem Phys. 2012 May;136(17):174101. https://doi.org/10.1063/1.4707167
74. Peterson AA. Acceleration of saddle-point searches with machine learning. J Chem Phys. 2016 Aug;145(7):074106. https://doi.org/10.1063/1.4960708
75. Zhou Y, Wu J, Chen S, Chen GH. Toward the exact exchange correlation potential: a three-dimensional convolutional neural network construct. J Phys Chem Lett. 2019 Nov;10(22):7264–9. https://doi.org/10.1021/acs.jpclett.9b02838
76. Dick S, Fernandez-Serra M. Machine learning accurate exchange and correlation functionals of the electronic density. Nat Commun. 2020 Jul;11(1):3509. https://doi.org/10.1038/s41467-020-17265-7

77. Houston PL, Nandi A, Bowman JM. A machine learning approach for prediction of rate constants. J Phys Chem Lett. 2019 Aug;10(17):5250–8. https://doi.org/10.1021/acs.jpclett.9b01810
78. Komp E, Valleau S. Machine learning quantum reaction rate constants. J Phys Chem A. 2020 Sep;124(41):8607–13. https://doi.org/10.1021/acs.jpca.0c05992
79. Kayala MA, Azencott C-A, Chen JH, Baldi P. Learning to predict chemical reactions. J Chem Inf Model. 2011 Sep;51(9):2209–22. https://doi.org/10.1021/ci200207y
80. Coley CW, Jin W, Rogers L, Jamison TF, Jaakkola TS, Green WH, et al. A graph-convolutional neural network model for the prediction of chemical reactivity. Chem Sci. 2019;10(2):370–7. https://doi.org/10.1039/c8sc04228d
81. Wei JN, Duvenaud D, Aspuru-Guzik A. Neural networks for the prediction of organic chemistry reactions. ACS Cent Sci. 2016 Oct;2(10):725–32. https://doi.org/10.1021/acscentsci.6b00219
82. Liu X, Zhang T, Yang T, Liu X, Song X, Yang Y, et al. Solving chemistry problems via an end-to-end approach: a proof of concept. J Phys Chem A. 2020 Oct;124(42):8866–73. https://doi.org/10.1021/acs.jpca.0c06319
83. Almeida AF de, Moreira R, Rodrigues T. Synthetic organic chemistry driven by artificial intelligence. Nat Rev Chem. 2019 Aug;3(10):589–604. https://doi.org/10.1038/s41570-019-0124-0
84. Fooshee D, Mood A, Gutman E, Tavakoli M, Urban G, Liu F, et al. Deep learning for chemical reaction prediction. Mol Syst Des Eng. 2018;3(3):442–52. https://doi.org/10.1039/c7me00107j
85. Rankine CD, Madkhali MMM, Penfold TJ. A deep neural network for the rapid prediction of X-ray absorption spectra. J Phys Chem A. 2020 May;124(21):4263–70. https://doi.org/10.1021/acs.jpca.0c03723
86. Madkhali MMM, Rankine CD, Penfold TJ. The role of structural representation in the performance of a deep neural network for X-ray spectroscopy. Molecules. 2020 Jun;25(11):2715. https://doi.org/10.3390/molecules25112715
87. Liu Y, Marcella N, Timoshenko J, Halder A, Yang B, Kolipaka L, et al. Mapping XANES spectra on structural descriptors of copper oxide clusters using supervised machine learning. J Chem Phys. 2019 Oct;151(16):164201. https://doi.org/10.1063/1.5126597
88. Karlik B, Yilmaz MF, Ozdemir M, Yavuz CT, Danisman Y. A hybrid machine learning model to study UV–Vis spectra of gold nanospheres. Plasmonics. 2020 Sep;16(1):147–55. https://doi.org/10.1007/s11468-020-01267-8
89. Fu W, Hopkins WS. Applying machine learning to vibrational spectroscopy. J Phys Chem A. 2017 Dec;122(1):167–71. https://doi.org/10.1021/acs.jpca.7b10303
90. Kananenka AA, Yao K, Corcelli SA, Skinner JL. Machine learning for vibrational spectroscopic maps. J Chem Theory Comput. 2019 Oct;15(12):6850–8. https://doi.org/10.1021/acs.jctc.9b00698
91. Acquarelli J, Laarhoven T van, Gerretzen J, Tran TN, Buydens LMC, Marchiori E. Convolutional neural networks for vibrational spectroscopic data analysis. Anal Chim Acta. 2017 Feb;954:22–31. https://doi.org/10.1016/j.aca.2016.12.010
92. Goodacre R. Explanatory analysis of spectroscopic data using machine learning of simple interpretable rules. Vib Spectrosc. 2003 Aug;32(1):33–45. https://doi.org/10.1016/S0924-2031(03)00045-6
93. Cobas C. NMR signal processing prediction, and structure verification with machine learning techniques. Magn Reson Chem. 2020 Jan;58(6):512–9. https://doi.org/10.1002/mrc.4989
94. Gao P, Zhang J, Peng Q, Zhang J, Glezakou V-A. General protocol for the accurate prediction of molecular $^{13}C/^{1}H$ NMR chemical shifts via machine learning augmented DFT. J Chem Inf Model. 2020 Jun;60(8):3746–54. https://doi.org/10.1021/acs.jcim.0c00388
95. Rupp M, Ramakrishnan R, Lilienfeld OA von. Machine learning for quantum mechanical properties of atoms in molecules. J Phys Chem Lett. 2015 Aug;6(16):3309–13. https://doi.org/10.1021/acs.jpclett.5b01456
96. Alley EC, Khimulya G, Biswas S, AlQuraishi M, Church GM. Unified rational protein engineering with sequence-based deep representation learning. Nat Methods. 2019 Oct;16:1315–22. https://doi.org/10.1038/s41592-019-0598-1
97. Wang H, Chen X, Li C, Liu Y, Yang F, Wang C. Sequence-based prediction of cysteine reactivity using machine learning. 2017 Oct;57(4):451–60. https://doi.org/10.1021/acs.biochem.7b00897
98. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. 2021 Jul;596:583–9. https://doi.org/10.1038/s41586-021-03819-2
99. Wang T, Qiao Y, Ding W, Mao W, Zhou Y, Gong H. Improved fragment sampling for ab initio protein structure prediction using deep neural networks. Nat Mach Intell. 2019 Jul;1:347–55. https://doi.org/10.1038/s42256-019-0075-7
100. AlQuraishi M. Machine learning in protein structure prediction. Curr Opin Chem Biol. 2021 Dec;65:1–8. https://doi.org/10.1016/j.cbpa.2021.04.005
101. Torrisi M, Pollastri G, Le Q. Deep learning methods in protein structure prediction. Comput Struct Biotechnol J. 2020 Dec;18:1301–10. https://doi.org/10.1016/j.csbj.2019.12.011
102. Montgomery DC, Peck EA, Vining GG. Introduction to linear regression analysis. Hoboken, NJ: John Wiley & Sons; 2012.
103. Santiago CB, Guo J-Y, Sigman MS. Predictive and mechanistic multivariate linear regression models for reaction development. Chem Sci. 2018;9(9):2398–412. https://doi.org/10.1039/c7sc04679k
104. McDonald GC. Ridge regression. WIREs Comput Stat. 2009 Jul;1(1):93–100. https://doi.org/10.1002/wics.14
105. Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. Technometrics. 2000 Feb;42(1):80–6. https://doi.org/10.1080/00401706.2000.10485983
106. Tibshirani R. Regression shrinkage and selection via the Lasso. J R Stat Soc B Methodol. 1996 Jan;58(1):267–88. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

107. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Series B Stat Methodol. 2005 Apr;67(2):301–20. https://doi.org/10.1111/j.1467-9868.2005.00503.x
108. Behler J. Representing potential energy surfaces by high-dimensional neural network potentials. J Phys Condens Matter. 2014 Apr; 26(18):183001. https://doi.org/10.1088/0953-8984/26/18/183001
109. Behler J. Constructing high-dimensional neural network potentials: a tutorial review. Int J Quantum Chem. 2015 Mar;115(16):1032–50. https://doi.org/10.1002/qua.24890
110. Gurney K. An introduction to neural networks. London: CRC Press; 1997.
111. Mitchell JBO. Machine learning methods in chemoinformatics. WIREs Comput Mol Sci. 2014 Feb;4(5):468–81. https://doi.org/10.1002/wcms.1183
112. Eckle K, Schmidt-Hieber J. A comparison of deep networks with ReLU activation function and linear spline-type methods. Neural Netw. 2019 Feb;110:232–42. https://doi.org/10.1016/j.neunet.2018.11.005
113. Cartwright H. Applications of artificial intelligence in chemistry. Oxford: Oxford University Press; 1993.
114. Schmidhuber J. Deep learning in neural networks: An overview. Neural Netw. 2015 Jan;61:85–117. https://doi.org/10.1016/j.neunet.2014.09.003
115. Shinde PP, Shah S. A review of machine learning and deep learning applications. 2018 fourth international conference on computing communication control and automation (ICCUBEA). Pune, India: IEEE; 2018. https://doi.org/10.1109/iccubea.2018.8697857
116. Islam MT, Siddique BMNK, Rahman S, Jabid T. Image recognition with deep learning. 2018 international conference on intelligent informatics and biomedical sciences (ICIIBMS). Bangkok, Thailand: IEEE; 2018. https://doi.org/10.1109/iciibms.2018.8549986
117. Goh GB, Hodas NO, Vishnu A. Deep learning for computational chemistry. J Comput Chem. 2017 Mar;38(16):1291–307. https://doi.org/10.1002/jcc.24764
118. Schütt KT, Sauceda HE, Kindermans P-J, Tkatchenko A, Müller K-R. SchNet – a deep learning architecture for molecules and materials. J Chem Phys. 2018 Jun;148(24):241722. https://doi.org/10.1063/1.5019779
119. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015 May;521(7553):436–44. https://doi.org/10.1038/nature14539
120. Behler J. Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations. Phys Chem Chem Phys. 2011; 13(40):17930–55. https://doi.org/10.1039/c1cp21668f
121. Tzeng F-Y, Ma K-L. Opening the black box – data driven visualization of neural networks. VIS 05 IEEE visualization. Minneapolis, MN: IEEE; 2005. https://doi.org/10.1109/visual.2005.1532820
122. Dayhoff JE, DeLeo JM. Artificial neural networks: opening the black box. Cancer. 2001;91:1615–35.
123. Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods. New York: Cambridge University Press; 2000.
124. Skyner RE, McDonagh JL, Groom CR, Mourik T van, Mitchell JBO. A review of methods for the calculation of solution free energies and the modelling of systems in solution. Phys Chem Chem Phys. 2015;17(9):6174–91. https://doi.org/10.1039/c5cp00288e
125. Schölkopf B, Smola AJ. Learning with kernels: support vector machines, regularization, optimization, and beyond. Cambridge, MA: The MIT Press; 2018.
126. Vovk V. Empirical inference. Heidelberg, Germany: Springer; 2013.
127. Vu K, Snyder JC, Li L, Rupp M, Chen BF, Khelif T, et al. Understanding kernel ridge regression: common behaviors from simple functions to density functionals. Int J Quantum Chem. 2015 May;115(16):1115–28. https://doi.org/10.1002/qua.24939
128. Muller K-R, Mika S, Ratsch G, Tsuda K, Scholkopf B. An introduction to kernel-based learning algorithms. IEEE Trans Neural Netw. 2001 Mar;12(2):181–201. https://doi.org/10.1109/72.914517
129. Faber FA, Hutchison L, Huang B, Gilmer J, Schoenholz SS, Dahl GE, et al. Prediction errors of molecular machine learning models lower than hybrid DFT error. J Chem Theory Comput. 2017 Oct;13(11):5255–64. https://doi.org/10.1021/acs.jctc.7b00577
130. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32. https://doi.org/10.1023/a%3A1010933404324
131. Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP. Random forest: a classification and regression tool for compound classification and QSAR modeling. J Chem Inf Comput Sci. 2003 Nov;43(6):1947–58. https://doi.org/10.1021/ci034160g
132. Natekin A, Knoll A. Gradient boosting machines a tutorial. Front Neurorobot. 2013;7:21. https://doi.org/10.3389/fnbot.2013.00021
133. Rasmussen CE. Gaussian processes in machine learning. Advanced lectures on machine learning. Berlin Heidelberg: Springer; 2004. p. 63–71. https://doi.org/10.1007/978-3-540-28650-9_4
134. Rasmussen CE, Williams CKI. Gaussian processes for machine learning. Cambridge, MA: The MIT Press; 2006.
135. Bartók AP, Csányi G. Gaussian approximation potentials: a brief tutorial introduction. Int J Quantum Chem. 2015 Apr;115(16):1051–7. https://doi.org/10.1002/qua.24927
136. Simm GN, Reiher M. Error-controlled exploration of chemical reaction networks with Gaussian processes. J Chem Theory Comput. 2018 Sep;14(10):5238–48. https://doi.org/10.1021/acs.jctc.8b00504
137. Saatçi Y. Scalable inference for structured Gaussian process models [PhD thesis]. University of Cambridge; 2011.
138. Fang W, Zhang S, Huang H, Dang S, Huang Z, Li W, et al. Learn to make decision with small data for autonomous driving: deep Gaussian process and feedback control. J Adv Transp. 2020 Aug;2020:1–11. https://doi.org/10.1155/2020/8495264
139. Choi S, Kim Y, Kim JW, Kim Z, Kim WY. Feasibility of activation energy prediction of gas-phase reactions by machine learning. Chem Eur J. 2018 Apr;24(47):12354–8. https://doi.org/10.1002/chem.201800345
140. Friederich P, dos Passos Gomes G, Bin RD, Aspuru-Guzik A, Balcells D. Machine learning dihydrogen activation in the chemical space surrounding Vaska's complex. Chem Sci. 2020;11(18):4584–601. https://doi.org/10.1039/d0sc00445f

141. Grambow CA, Pattanaik L, Green WH. Deep learning of activation energies. J Phys Chem Lett. 2020 Mar;11(8):2992–7. https://doi.org/10.1021/acs.jpclett.0c00500
142. Li X, Zhang S-Q, Xu L-C, Hong X. Predicting regioselectivity in radical C–H functionalization of heterocycles through machine learning. Angew Chem. 2020 May;132(32):13355–61. https://doi.org/10.1002/ange.202000959
143. Migliaro I, Cundari TR. Density functional study of methane activation by frustrated Lewis pairs with group 13 trihalides and group 15 pentahalides and a machine learning analysis of their barrier heights. J Chem Inf Model. 2020 Sep;60(10):4958–66. https://doi.org/10.1021/acs.jcim.0c00862
144. Mikami K. Interactive-quantum-chemical-descriptors enabling accurate prediction of an activation energy through machine learning. Polymer. 2020 Aug;203:122738. https://doi.org/10.1016/j.polymer.2020.122738
145. Palazzesi F, Hermann MR, Grundl MA, Pautsch A, Seeliger D, Tautermann CS, et al. BIreactive: a machine-learning model to estimate covalent warhead reactivity. J Chem Inf Model. 2020 Apr;60(6):2915–23. https://doi.org/10.1021/acs.jcim.9b01058
146. Jorner K, Brinck T, Norrby P-O, Buttar D. Machine learning meets mechanistic modelling for accurate prediction of experimental activation energies. Chem Sci. 2021;12(3):1163–75. https://doi.org/10.1039/d0sc04896h
147. Sadowski P, Fooshee D, Subrahmanya N, Baldi P. Synergies between quantum mechanics and machine learning in reaction prediction. J Chem Inf Model. 2016 Oct;56(11):2125–8. https://doi.org/10.1021/acs.jcim.6b00351
148. Bragato M, von Rudorff GF, von Lilienfeld OA. Data enhanced Hammett-equation: reaction barriers in chemical space. Chem Sci. 2020;11(43):11859–68. https://doi.org/10.1039/d0sc04235h
149. Döntgen M, Fenard Y, Heufer KA. Atomic partial charges as descriptors for barrier heights. J Chem Inf Model. 2020 Oct;60(12):5928–31. https://doi.org/10.1021/acs.jcim.0c00787
150. Ulissi ZW, Tang MT, Xiao J, Liu X, Torelli DA, Karamad M, et al. Machine-learning methods enable exhaustive searches for active bimetallic facets and reveal active site motifs for $CO_2$ reduction. ACS Catal. 2017 Aug;7(10):6600–8. https://doi.org/10.1021/acscatal.7b01648
151. Toyao T, Suzuki K, Kikuchi S, Takakusagi S, K-ichi S, Takigawa I. Toward effective utilization of methane: machine learning prediction of adsorption energies on metal alloys. J Phys Chem C. 2018 Mar;122(15):8315–26. https://doi.org/10.1021/acs.jpcc.7b12670
152. Hoyt RA, Montemore MM, Fampiou I, Chen W, Tritsaris G, Kaxiras E. Machine learning prediction of H adsorption energies on Ag alloys. J Chem Inf Model. 2019 Mar;59(4):1357–65. https://doi.org/10.1021/acs.jcim.8b00657
153. Xu J, Cao X-M, Hu P. Improved prediction for the methane activation mechanism on rutile metal oxides by a machine learning model with geometrical descriptors. J Phys Chem C. 2019 Nov;123(47):28802–10. https://doi.org/10.1021/acs.jpcc.9b08939
154. Praveen CS, Comas-Vives A. Design of an accurate machine learning algorithm to predict the binding energies of several adsorbates on multiple sites of metal surfaces. ChemCatChem. 2020 Jul;12(18):4611–7. https://doi.org/10.1002/cctc.202000517
155. Zhu X, Tsang DCW, Wang L, Su Z, Hou D, Li L, et al. Machine learning exploration of the critical factors for $CO_2$ adsorption capacity on porous carbon materials at different pressures. J Clean Prod. 2020 Nov;273:122915. https://doi.org/10.1016/j.jclepro.2020.122915
156. Lu Z, Chen ZW, Singh CV. Neural network-assisted development of high-entropy alloy catalysts: decoupling ligand and coordination effects. Matter. 2020 Oct;3(4):1318–33. https://doi.org/10.1016/j.matt.2020.07.029
157. Gu GH, Choi C, Lee Y, Situmorang AB, Noh J, Kim Y-H, et al. Progress in computational and machine-learning methods for heterogeneous small-molecule activation. Adv Mater. 2020 Mar;32(35):1907865. https://doi.org/10.1002/adma.201907865
158. Singh AR, Rohr BA, Gauthier JA, Nørskov JK. Predicting chemical reaction barriers with a machine learning model. Catal Lett. 2019 Mar;149(9):2347–54. https://doi.org/10.1007/s10562-019-02705-x
159. Takahashi K, Miyazato I. Rapid estimation of activation energy in heterogeneous catalytic reactions via machine learning. J Comput Chem. 2018 Oct;39(28):2405–8. https://doi.org/10.1002/jcc.25567
160. Artrith N, Lin Z, Chen JG. Predicting the activity and selectivity of bimetallic metal catalysts for ethanol reforming using machine learning. ACS Catal. 2020 Aug;10(16):9438–44. https://doi.org/10.1021/acscatal.0c02089
161. Bučko T, Gešvandtnerová M, Rocca D. Ab initio calculations of free energy of activation at multiple electronic structure levels made affordable: An effective combination of perturbation theory and machine learning. J Chem Theory Comput. 2020 Jul;16(10):6049–60. https://doi.org/10.1021/acs.jctc.0c00486
162. Gastegger M, Marquetand P. High-dimensional neural network potentials for organic reactions and an improved training algorithm. J Chem Theory Comput. 2015 Apr;11(5):2187–98. https://doi.org/10.1021/acs.jctc.5b00211
163. Esch B von der, Dietschreit JCB, Peters LDM, Ochsenfeld C. Finding reactive configurations: a machine learning approach for estimating energy barriers applied to Sirtuin 5. J Chem Theory Comput. 2019 Nov;15(12):6660–7. https://doi.org/10.1021/acs.jctc.9b00876
164. Gao CW, Allen JW, Green WH, West RH. Reaction mechanism generator: automatic construction of chemical kinetic mechanisms. Comput Phys Commun. 2016 Jun;203:212–25. https://doi.org/10.1016/j.cpc.2016.02.013
165. Stephan DW. Frustrated Lewis pairs. J Am Chem Soc. 2015 Aug;137(32):10018–32. https://doi.org/10.1021/jacs.5b06794
166. Stephan DW. Frustrated Lewis pairs: from concept to catalysis. Acc Chem Res. 2014 Dec;48(2):306–16. https://doi.org/10.1021/ar500375j
167. Welch GC, Juan RRS, Masuda JD, Stephan DW. Reversible metal-free hydrogen activation. Science. 2006 Nov;314(5802):1124–6. https://doi.org/10.1126/science.1134230
168. Stephan DW. The broadening reach of frustrated Lewis pair chemistry. Science. 2016 Dec;354(6317):aaf7229. https://doi.org/10.1126/science.aaf7229
169. Villegas-Escobar N, Toro-Labbé A, Becerra M, Real-Enriquez M, Mora JR, Rincon L. A DFT study of hydrogen and methane activation by B(C6F5)3/P(t-Bu)3 and Al(C6F5)3/P(t-Bu)3 frustrated Lewis pairs. J Mol Model. 2017 Jul;23(8):234. https://doi.org/10.1007/s00894-017-3404-y

170. Jupp AR, Johnstone TC, Stephan DW. Improving the global electrophilicity index (GEI) as a measure of Lewis acidity. Inorg Chem. 2018 Nov;57(23):14764–71. https://doi.org/10.1021/acs.inorgchem.8b02517
171. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learn. 2006 Mar;63(1):3–42. https://doi.org/10.1007/s10994-006-6226-1
172. Singh J, Petter RC, Baillie TA, Whitty A. The resurgence of covalent drugs. Nat Rev Drug Discov. 2011 Apr;10(4):307–17. https://doi.org/10.1038/nrd3410
173. Bauer RA. Covalent inhibitors in drug discovery: from accidental discoveries to avoided liabilities and designed therapies. Drug Discov Today. 2015 Sep;20(9):1061–73. https://doi.org/10.1016/j.drudis.2015.05.005
174. Gehringer M, Laufer SA. Emerging and re-emerging warheads for targeted covalent inhibitors: applications in medicinal chemistry and chemical biology. J Med Chem. 2018 Dec;62(12):5673–724. https://doi.org/10.1021/acs.jmedchem.8b01153
175. Hawkins PCD, Skillman AG, Warren GL, Ellingson BA, Stahl MT. Conformer generation with OMEGA: algorithm and validation using high quality structures from the protein databank and Cambridge structural database. J Chem Inf Model. 2010 Mar;50(4):572–84. https://doi.org/10.1021/ci100031x
176. Todeschini R, Consonni V. Molecular descriptors for chemoinformatics: volume I: alphabetical listing/volume II: appendices, references, volume 41. Weinheim, Germany: John Wiley & Sons; 2009.
177. Todeschini R, Consonni V. Handbook of molecular descriptors. Weinheim, Germany: John Wiley & Sons; 2008.
178. Dodge Y. The concise encyclopedia of statistics. New York: Springer; 2008.
179. DRAGON. 7.0 – Kode – Chemoinformatics. Available from: https://chm.kode-solutions.net/pf/dragon-7-0/ [accessed 2021 Sep].
180. Fooshee D, Andronico A, Baldi P. ReactionMap: An efficient atom-mapping algorithm for chemical reactions. J Chem Inf Model. 2013 Nov;53(11):2812–9. https://doi.org/10.1021/ci400326p
181. Jaworski W, Szymkuć S, Mikulak-Klucznik B, Piecuch K, Klucznik T, Kaźmierowski M, et al. Automatic mapping of atoms across both simple and complex chemical reactions. Nat Commun. 2019 Mar;10(1):1434. https://doi.org/10.1038/s41467-019-09440-2
182. Yang K, Swanson K, Jin W, Coley C, Eiden P, Gao H, et al. Analyzing learned molecular representations for property prediction. J Chem Inf Model. 2019 Jul;59(8):3370–88. https://doi.org/10.1021/acs.jcim.9b00237
183. Grambow CA, Pattanaik L, Green WH. Reactants products, and transition states of elementary chemical reactions based on quantum chemistry. Sci Data. 2020 May;7(1):137. https://doi.org/10.1038/s41597-020-0460-4
184. Grambow CA, Li Y-P, Green WH. Accurate thermochemistry with small data sets: a bond additivity correction and transfer learning approach. J Phys Chem A. 2019 Jun;123(27):5826–35. https://doi.org/10.1021/acs.jpca.9b04195
185. Schütt KT, Arbabzadah F, Chmiela S, Müller KR, Tkatchenko A. Quantum-chemical insights from deep tensor neural networks. Nat Commun. 2017 Jan;8(1):13890. https://doi.org/10.1038/ncomms13890
186. Unke OT, Meuwly M. PhysNet: a neural network for predicting energies forces, dipole moments, and partial charges. J Chem Theory Comput. 2019 May;15(6):3678–93. https://doi.org/10.1021/acs.jctc.9b00181
187. Lee K, Yoo D, Jeong W, Han S. SIMPLE-NN: an efficient package for training and executing neural-network interatomic potentials. Comput Phys Commun. 2019 Sep;242:95–103. https://doi.org/10.1016/j.cpc.2019.04.014
188. Behler J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. J Chem Phys. 2011 Feb;134(7):074106. https://doi.org/10.1063/1.3553717
189. Spicher S, Grimme S. Robust atomistic modeling of materials organometallic, and biochemical systems. Angew Chem. 2020 May;132(36):15795–803. https://doi.org/10.1002/ange.202004239
190. Roos K, Wu C, Damm W, Reboul M, Stevenson JM, Lu C, et al. OPLS3e: extending force field coverage for drug-like small molecules. J Chem Theory Comput. 2019 Feb;15(3):1863–74. https://doi.org/10.1021/acs.jctc.8b01026
191. Kanal IY, Keith JA, Hutchison GR. A sobering assessment of small-molecule force field methods for low energy conformer predictions. Int J Quantum Chem. 2017 Oct;118(5):e25512. https://doi.org/10.1002/qua.25512
192. Bartók AP, Kondor R, Csányi G. On representing chemical environments. Phys Rev B. 2013 May;87(18):184115. https://doi.org/10.1103/physrevb.87.184115
193. De S, Bartók AP, Csányi G, Ceriotti M. Comparing molecules and solids across structural and alchemical space. Phys Chem Chem Phys. 2016;18(20):13754–69. https://doi.org/10.1039/c6cp00415f
194. Clavier H, Nolan SP. Percent buried volume for phosphine and N-heterocyclic carbene ligands: steric properties in organometallic chemistry. Chem Commun. 2010;46(6):841–61. https://doi.org/10.1039/b922984a
195. Maley SM, Kwon D-H, Rollins N, Stanley JC, Sydora OL, Bischof SM, et al. Quantum-mechanical transition-state model combined with machine learning provides catalyst design features for selective Cr olefin oligomerization. Chem Sci. 2020;11(35):9665–74. https://doi.org/10.1039/d0sc03552a
196. Durant JL, Leland BA, Henry DR, Nourse JG. Reoptimization of MDL keys for use in drug discovery. J Chem Inf Comput Sci. 2002 Nov;42(6):1273–80. https://doi.org/10.1021/ci010132r
197. Rogers D, Hahn M. Extended-connectivity fingerprints. J Chem Inf Model. 2010 Apr;50(5):742–54. https://doi.org/10.1021/ci100050t
198. Hansen K, Biegler F, Ramakrishnan R, Pronobis W, von Lilienfeld OA, Müller K-R, et al. Machine learning predictions of molecular properties: accurate many-body potentials and nonlocality in chemical space. J Phys Chem Lett. 2015 Jun;6(12):2326–31. https://doi.org/10.1021/acs.jpclett.5b00831
199. von Ragué Schleyer P, Maerker C, Dransfeld A, Jiao H, van Eikema Hommes NJR. Nucleus-independent chemical shifts: a simple and efficient aromaticity probe. J Am Chem Soc. 1996 Jan;118(26):6317–8. https://doi.org/10.1021/ja960582d

200. Halgren TA. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. J Comput Chem. 1996; 17:490–519.
201. Stewart JJP. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. J Mol Model. 2012 Nov;19(1):1–32. https://doi.org/10.1007/s00894-012-1667-x
202. Reed AE, Weinstock RB, Weinhold F. Natural population analysis. J Chem Phys. 1985 Jul;83(2):735–46. https://doi.org/10.1063/1.449486
203. Brethomé AV, Paton RS, Fletcher SP. Retooling asymmetric conjugate additions for sterically demanding substrates with an iterative data-driven approach. ACS Catal. 2019 Jul;9(8):7179–87. https://doi.org/10.1021/acscatal.9b01814
204. Harper KC, Bess EN, Sigman MS. Multidimensional steric parameters in the analysis of asymmetric catalytic reactions. Nat Chem. 2012 Mar;4(5):366–74. https://doi.org/10.1038/nchem.1297
205. Foster JP, Weinhold F. Natural hybrid orbitals. J Am Chem Soc. 1980 Nov;102(24):7211–8. https://doi.org/10.1021/ja00544a007
206. Reed AE, Curtiss LA, Weinhold F. Intermolecular interactions from a natural bond orbital donor-acceptor viewpoint. Chem Rev. 1988 Sep;88(6):899–926. https://doi.org/10.1021/cr00088a005
207. Glendening ED, Streitwieser A. Natural energy decomposition analysis: an energy partitioning procedure for molecular interactions with application to weak hydrogen bonding strong ionic, and moderate donor acceptor interactions. J Chem Phys. 1994 Feb;100(4): 2900–9. https://doi.org/10.1063/1.466432
208. Glendening ED. Natural energy decomposition analysis: explicit evaluation of electrostatic and polarization effects with application to aqueous clusters of alkali metal cations and neutrals. J Am Chem Soc. 1996 Jan;118(10):2473–82. https://doi.org/10.1021/ja951834y
209. Glendening ED. Natural energy decomposition analysis: extension to density functional methods and analysis of cooperative effects in water clusters. J Phys Chem A. 2005 Nov;109(51):11936–40. https://doi.org/10.1021/jp058209s
210. Dolg M, Wedig U, Stoll H, Preuss H. Energy-adjusted ab initio pseudopotentials for the first row transition elements. J Chem Phys. 1987 Jan;86(2):866–72. https://doi.org/10.1063/1.452288
211. Igel-Mann G, Stoll H, Preuss H. Pseudopotentials for main group elements (IIIa through VIIa). Mol Phys. 1988 Dec;65(6):1321–8. https://doi.org/10.1080/00268978800101811
212. Andrae D, Häußermann U, Dolg M, Stoll H, Preuß H. Energy-adjusted ab initio pseudopotentials for the second and third row transition elements. Theor Chim Acta. 1990;77(2):123–41. https://doi.org/10.1007/bf01114537
213. Halkier A, Helgaker T, Jørgensen P, Klopper W, Koch H, Olsen J, et al. Basis-set convergence in correlated calculations on Ne, $N_2$, and $H_2O$. Chem Phys Lett. 1998 Apr;286(3–4):243–52. https://doi.org/10.1016/s0009-2614(98)00111-0
214. Vaska L, DiLuzio JW. Carbonyl and hydrido-carbonyl complexes of iridium by reaction with alcohols. Hydrido complexes by reaction with acid. J Am Chem Soc. 1961 Jun;83(12):2784–5. https://doi.org/10.1021/ja01473a054
215. Vaska L, DiLuzio JW. Activation of hydrogen by a transition metal complex at normal conditions leading to a stable molecular dihydride. J Am Chem Soc. 1962 Feb;84(4):679–80. https://doi.org/10.1021/ja00863a040
216. Perdew JP, Burke K, Ernzerhof M. Generalized gradient approximation made simple. Phys Rev Lett. 1996 Oct;77(18):3865–8. https://doi.org/10.1103/physrevlett.77.3865
217. Schäfer A, Horn H, Ahlrichs R. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. J Chem Phys. 1992 Aug;97(4): 2571–7. https://doi.org/10.1063/1.463096
218. Grimme S, Antony J, Ehrlich S, Krieg H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. J Chem Phys. 2010 Apr;132(15):154104. https://doi.org/10.1063/1.3382344
219. Morgan HL. The generation of a unique machine description for chemical structures – a technique developed at chemical abstracts service. J Chem Doc. 1965 May;5(2):107–13. https://doi.org/10.1021/c160017a018
220. RDKit: Open-source cheminformatics software. Available from: https://rdkit.org/ [accessed 2021 Sep].
221. Bannwarth C, Ehlert S, Grimme S. GFN2-xTB – an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. J Chem Theory Comput. 2019 Feb;15(3): 1652–71. https://doi.org/10.1021/acs.jctc.8b01176
222. Stenlid JH, Brinck T. Nucleophilic aromatic substitution reactions described by the local electron attachment energy. J Org Chem. 2017 Feb;82(6):3072–83. https://doi.org/10.1021/acs.joc.7b00059
223. Murray JS, Politzer P. Molecular electrostatic potentials and noncovalent interactions. WIREs Comput Mol Sci. 2017 Jul;7(6):e1326. https://doi.org/10.1002/wcms.1326
224. Murray JS, Politzer P. The electrostatic potential: an overview. WIREs Comput Mol Sci. 2011 Feb;1(2):153–63. https://doi.org/10.1002/wcms.19
225. Oller J, Pérez P, Ayers PW, Vöhringer-Martinez E. Global and local reactivity descriptors based on quadratic and linear energy models for α,β-unsaturated organic compounds. Int J Quantum Chem. 2018 Oct;118(20):e25706. https://doi.org/10.1002/qua.25706
226. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. J Mol Biol. 1971 Feb;55(3):379. https://doi.org/10.1016/0022-2836(71)90324-x
227. Pollice R, Chen P. A universal quantitative descriptor of the dispersion interaction potential. Angew Chem Int Ed. 2019 Jul;58(29): 9758–69. https://doi.org/10.1002/anie.201905439
228. Varnek A, Fourches D, Hoonakker F, Solov'ev VP. Substructural fragments: an universal language to encode reactions molecular and supramolecular structures. J Comput Aided Mol Des. 2005 Sep;19(9–10):693–703. https://doi.org/10.1007/s10822-005-9008-0

229. Schwaller P, Probst D, Vaucher AC, Nair VH, Kreutter D, Laino T, et al. Mapping the space of chemical reactions using attention-based neural networks. Nat Mach Intell. 2021 Jan;3(2):144–52. https://doi.org/10.1038/s42256-020-00284-w
230. Loyola-Gonzalez O. Black-box vs. white-box: understanding their advantages and weaknesses from a practical point of view. IEEE Access. 2019;7:154096–113. https://doi.org/10.1109/access.2019.2949286
231. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell. 2019 May;1(5):206–15. https://doi.org/10.1038/s42256-019-0048-x
232. Ramakrishnan R, Dral PO, Rupp M, Lilienfeld OA von. Big data meets quantum chemistry approximations: the Δ-machine learning approach. J Chem Theory Comput. 2015 Apr;11(5):2087–96. https://doi.org/10.1021/acs.jctc.5b00099
233. Jaffé HH. A reëxamination of the Hammett equation. Chem Rev. 1953 Oct;53(2):191–261. https://doi.org/10.1021/cr60165a003
234. Hammett LP. Linear free energy relationships in rate and equilibrium phenomena. Trans Faraday Soc. 1938;34:156. https://doi.org/10.1039/tf9383400156
235. Ayoubi-Chianeh M, Kassaee MZ. Toward triplet disilavinylidenes: a Hammett electronic survey for substituent effects on singlet-triplet energy gaps of silylenes by DFT. J Phys Org Chem. 2019 Jun;32(10):e3988. https://doi.org/10.1002/poc.3988
236. Hammett LP. Some relations between reaction rates and equilibrium constants. Chem Rev. 1935 Aug;17(1):125–36. https://doi.org/10.1021/cr60056a010
237. Hammett LP. The effect of structure upon the reactions of organic compounds. Benzene derivatives. J Am Chem Soc. 1937 Jan;59(1):96–103. https://doi.org/10.1021/ja01280a022
238. Liveris M, Lutz PG, Miller J. The SN mechanism in aromatic compounds. Part XIX. J Am Chem Soc. 1956 Jul;78(14):3375–8. https://doi.org/10.1021/ja01595a031
239. Pearson DE, Baxter JF, Martin JC. Hammetts sigma constants in certain electrophilic reactions. J Org Chem. 1952 Nov;17(11):1511–8. https://doi.org/10.1021/jo50011a019
240. Hine J. Polar effects on rates and equilibria. J Am Chem Soc. 1959 Mar;81(5):1126–9. https://doi.org/10.1021/ja01514a028
241. Bogojeski M, Vogt-Maranto L, Tuckerman ME, Müller K-R, Burke K. Quantum chemical accuracy from density functional approximations via machine learning. Nat Commun. 2020 Oct;11(1):5223. https://doi.org/10.1038/s41467-020-19093-1
242. Grossman RB. The art of writing reasonable organic reaction mechanisms. Cham, Switzerland: Springer; 2019.
243. Kayala MA, Baldi P. ReactionPredictor: prediction of complex chemical reactions at the mechanistic level using machine learning. J Chem Inf Model. 2012 Oct;52(10):2526–40. https://doi.org/10.1021/ci3003039
244. Fooshee D, Mood A, Gutman E, Tavakoli M, Urban G, Liu F, et al. Deep learning for chemical reaction prediction. Mol Syst Des Eng. 2017 Dec;3:442–52. https://doi.org/10.1039/c7me00107j
245. Yao K, Herr JE, Toth DW, Mckintyre R, Parkhill J. The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics. Chem Sci. 2018;9(8):2261–9. https://doi.org/10.1039/c7sc04934j
246. Guan Y, Ingman VM, Rooks BJ, Wheeler SE. AARON: An automated reaction optimizer for new catalysts. J Chem Theory Comput. 2018 Aug;14(10):5249–61. https://doi.org/10.1021/acs.jctc.8b00578


**How to cite this article:** Lewis-Atwell T, Townsend PA, Grayson MN. Machine learning activation energies of chemical reactions. WIREs Comput Mol Sci. 2022;12:e1593. https://doi.org/10.1002/wcms.1593