Fast Predictions of Reaction Barrier Heights: Toward Coupled-Cluster Accuracy

Fast Predictions of Reaction Barrier Heights: Towards Coupled-Cluster Accuracy

Kevin A. Spiekermann, Lagnajit Pattanaik, and William H. Green*

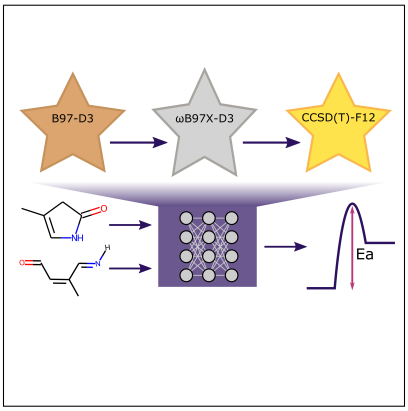
Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, United States

E-mail: whgreen@mit.edu

Abstract

Quantitative estimates of reaction barriers are essential for developing kinetic mechanisms and predicting reaction outcomes. However, the lack of experimental data and the steep scaling of accurate quantum calculations often hinder the ability to obtain reliable kinetic values. Here, we train a directed message passing neural network on nearly 24,000 diverse gas-phase reactions calculated at CCSD(T)-F12a/cc-pVDZ-F12//ωB97X-D3/def2-TZVP. Our model uses 75% fewer parameters from previous studies, an improved reaction representation, and proper data splits to accurately estimate performance on unseen reactions. Using information from only the reactant and product, our model quickly predicts barrier heights with a testing MAE of 2.6 kcal mol^{-1} relative to the coupled-cluster data, making it more accurate than a good density functional theory calculation. Further, our results show that future modeling efforts to estimate reaction properties would significantly benefit from fine-tuning calibration using a transfer learning technique. We anticipate this model will accelerate and improve kinetic predictions for small molecule chemistry.

TOC Graphic



Keywords

barrier height, deep learning

Introduction

Accurately predicting the time evolution of reacting chemical systems and the yields of various products and side products has been one of the main projects of physical chemistry for more than 135 years.[1,2] Some important systems have been very heavily studied, leading to chemical kinetic model predictions so compelling that they drive governmental and business decisions, and have even led to major international treaties (e.g. the Montreal Protocol).[3,4] If a large number of data are available, one can make useful predictions, e.g. regarding which organic synthesis routes are likely to succeed at specified reaction conditions.[5,6] However, data relevant to a particular system of interest are usually scarce, and it is often impractical to do enough experiments to develop a predictive model with the desired accuracy or generalizability.

Over the last ∼25 years, it has become possible to accurately compute the rate coefficients of individual reactions using quantum chemistry and rate theory.[7,8] This suggests it might be possible to accurately predict the behavior of a reacting system even before any experimental data are available on that system, with many ramifications.[9] Indeed, recently several research groups have constructed models to quantitatively predict the time-evolution of diverse chemical systems, based largely on parameters derived from quantum chemistry rather than experiment. For example, earlier this year our group has published models for combustion,[10] pyrolysis,[11,12] and the degradation of pharmaceutical compounds.[13] However, because quantum chemistry calculations of rate coefficients are slow, only a small fraction of the numerous rate coefficients in these models have been computed accurately, which can make the predictions somewhat erratic. The vision of reliable predictive chemical kinetics can only be realized if accurate rate coefficients can be calculated much more rapidly than is possible today.

The traditional workflow to obtain kinetic parameters, shown in Figure 1, is quite involved.[7,8] The process starts by constructing 3D structures for the reactant(s) and product(s), which are optimized using quantum chemical methods. The search for a good 3D structure

for the transition state (TS) often consumes the most human time, since automated saddle point finders are not yet sufficiently reliable.[14,15] After the most promising TS structure is identified, it is typically optimized using a more accurate method, and often an intrinsic reaction coordinate (IRC) calculation is run to confirm it connects the reactant(s) and product(s). Then the energy of reactant(s) and TS is re-computed at a high-level of theory (e.g. coupled-cluster) with zero-point energy corrections to provide an accurate reaction barrier. If the studied structures are flexible with many rotatable bonds or other large-amplitude motions, conformational effects must also be considered. This is usually done by approximating each hindered internal rotor as independent,[16–18] but often this approximation is not accurate, requiring some treatment of the coupling between different rotors.[19] Alternative methods which involve multiple conformational minima for reactants and transition states and couple conformational effects with low frequency anharmonic torsional modes–such as the multistructure methods developed by Truhlar and coworkers–are becoming popular, but it can be challenging to find all the conformers.[20–23] The workflow culminates in calculating the partition functions, which canonical transition state theory (TST) uses to estimate the high-pressure limit rate coefficient $k_\infty(T)$, including effects of symmetry/reaction path degeneracy, tunneling, and other corrections.[8,24–26]

Given the importance and impact of kinetic models, several efforts have been made to accelerate portions of the traditional workflow in Figure 1, either by accelerating individual steps, or by skipping over steps, illustrated as arrows in the interior of the figure. Many programs have been published for rapid conformer generation of stable species.[23,27–30] Several methods also exist for generating good TS guesses, such as using hand-made templates[31–33] or deep learning.[34–36]

Despite all these advances, calculations of rate coefficients are still limited by the accurate yet expensive quantum methods used to find the saddle point geometry, and the poor scaling of the even more expensive methods used to compute its energy. For example, coupled-cluster CCSD(T)-F12 calculations are commonly considered the gold-standard in quantum

chemistry due to their reliable single point energies and reaction energies,[37–42] yet they scale as $O(N^7)$, where N is the number of orbitals.[43] For some reactions, even CCSD(T) is not accurate enough, and more expensive calculations are required.[8]

Since so many computer resources and human efforts are required to obtain each reliable value, especially for reactions of large molecules, it would be advantageous to directly estimate kinetic parameters instead. Indeed, established methods for directly estimating barrier heights or $\log(k_\infty(T))$ from reactant and product identities include simple models such as Evans-Polanyi relationships[44] and the Hammett correlations.[45] In his textbook, Benson presented simple methods to predict Arrhenius A-factors.[46] Several authors have extended Benson's popular thermochemical group additivity approach[46–48] to estimate activation energies,[49,50] A-factors,[51] and rates.[52] Unfortunately, linear models are an inherently limited representation, so the simple Benson-type groups that work so well for predicting the thermochemistry of many organics are not sufficient for predicting rate coefficients over a broad range of reactions. Much larger supergroups[53] or decision trees[54–56] are needed to reach acceptable accuracy. Still, the supergroups and decision trees are often manually defined for each reaction type, which is tedious and error-prone. Thus, there has been interest in automating tree construction[57,58] to facilitate the incorporation of new training reactions, but the process is still cumbersome.

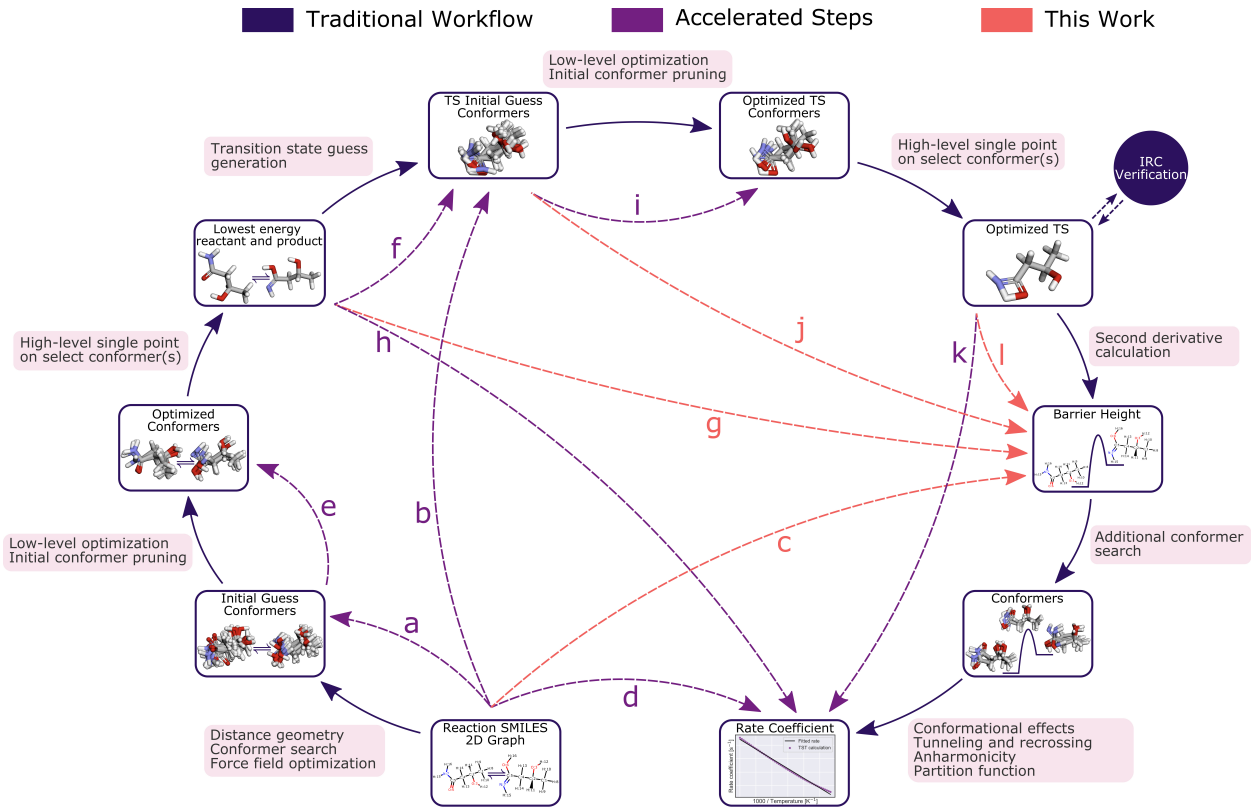


Figure 1: The conventional quantum chemistry and TST workflow for predicting a high-pressure limit rate coefficient is shown by the dark arrows. Several methods have been proposed in literature (and in the present work, shown by orange arrows) to accelerate steps in the process: a) Initial conformer generation ref 23,27–30,59–61 b) TS guess generation ref 32 c) Estimating barrier heights from 2D representation ref 44,49,55,57,62–65 d) Estimating rate coefficient from 2D representation ref 51,52,66–68 e) Semi-empirical optimization ref 69 f) TS guess generation ref 33–36,70–73 g) Estimating barrier heights from 3D representation ref 74,75 h) Estimating rate coefficient from 3D representation ref 76,77 i) Accelerating TS optimization ref 78–80 j) Estimating barrier height from un-optimized TS guess k) Estimating rate coefficient from optimized TS ref 81 l) Estimating barrier height from optimized TS ref 82,83

Recent work has focused on other nonlinear methods to predict kinetic parameters. For example, Heinen et al.[65] trained a kernel ridge regression (KRR) model to predict activation energies. The model was fit to their previously generated dataset[82] of S_N2 and E2 reactions with single point energies computed at DF-LCCSD/cc-TZVP. Stuyver et al.[63] trained a graph neural network (GNN), augmented with quantum descriptors, on the same dataset. In both cases, random splits gave a testing MAE of $\sim$2.5 kcal mol^{-1}. The relatively low

error is expected since the test set had a similar composition as the training set i.e. it is a measure of interpolation. To assess generalizability, Stuyver et al. also restricted training to three of the four nucleophiles in the dataset, reserving the fourth only for the test set. Although their augmented GNN extrapolated better than KRR to the unseen nucleophile, both models showed significantly higher testing errors relative to random splitting. Further, since all reactions start from a substituted ethyl-based scaffold, it is unclear how well these models would generalize to other reactants.

Estimators for activation energy have also been applied to catalysis. Takahashi et al.[75] trained several types of models (linear, random forest, and support vector regression) to predict the activation energy for a small dataset of heterogeneous catalytic reactions. When using a random split, testing errors were $\sim$0.90 eV (>20 kcal mol^{-1}) for all models, much too large to be useful. Similarly, Singh et al.[74] trained a feed forward neural network (FFN) model to predict reaction barriers for a small dataset containing dehydrogenation reactions as well as N_2 and O_2 dissociations. They report a testing error of 0.22 eV when using a random split.

Komp and Valleau[84] trained two different FFNs to predict the natural logarithm of the "partition function" at a given temperature. The first model takes the molecular geometry and inverse temperature as input to predict the "partition function" of a molecule. The second model uses the geometry and "partition function" of the reactant and product as well as the inverse temperature to predict a "partition function" of the TS. Unfortunately, the quantities used by those authors for training and testing their model were not the true partition functions that appear in rate theory, but instead rigid-rotor harmonic-oscillator (RRHO) partition functions of a single conformer of the TS, reactant or product. However, the great majority of the species in their data set have rotatable bonds and a large number of low-energy conformers, and many of the "product" structures and vibrational frequencies they used were actually properties of van der Waals complexes of two or three product molecules. Partition functions are drastically affected by conformers and rotors,[19–23] which

is why the refined dataset from ref 42 published high-pressure limit TST rate coefficients for only a small subset of the reactions–those where the TS and the reactant(s) were rigid so there was only one low-lying conformer. Further, in the Komp and Valleau paper, no analysis was provided for how sensitive the model predictions are to the input 3D geometry; as shown in our results section below, this could be a significant problem.

Another focus is using machine learning (ML) methods to directly estimate $k_\infty(T)$. Houston et al.[76] used Gaussian process regression (GPR) on 13 reactions to predict bimolecular rate constants over a large temperature range. As a follow up, Nandi et al.[77] clustered the input data, retrained the GPR model on each cluster, and then predicted rate coefficients for the O(^{3}P) + HCl reaction. In other work, Komp et al.[81] trained a FFN on ∼1.5 million data points to predict the product of $k_\infty(T)$ with the reactant partition function for 1D barrier problems. FFNs have also been fit to the rate coefficients from small datasets of hydroxyl radical reactions[66,67] and ionic liquids.[68] In all these cases, the data sets employed did not cover much reaction space.

Still another approach is to use Δ-ML, a technique developed by Ramakrishnan et al.[85] in which a model is fit to the residuals between a high- and low-level of theory. Thus, rather than directly predicting a high-quality value for the parameter of interest, the Δ-ML model predicts the correction to the low-level value, which should be relatively inexpensive to calculate. In the context of kinetics, Bragato et al.[83] showed that using KRR for Δ-ML outperformed direct ML approaches for predicting barrier heights from von Rudorff et al.'s[82] S_N2 dataset.

In the remainder of this paper, we present a new model for estimating the zero-point energy corrected barrier height from either 2D or 3D structures of the reactants and products. Many of the ML models summarized here, as well as in recent reviews,[86–88] are trained on relatively small datasets from a specific reaction family, which severely hinders their generalizability. In contrast, our goal is to develop a deep learning model that quickly predicts accurate barrier heights for a diverse set of reactions. Our model uses information from only

the reactant and product so future workflows could avoid the often delicate and challenging task of finding the TS geometry. This model would be useful for directly estimating barrier heights during the automated generation of kinetic models;[55] it would also offer substantial speedup when refining existing kinetic models by avoiding the need for a TS geometry. Lastly, this model would allow for quantitative ranking of potential reactions produced from an automated enumeration.[14,15,89,90]

Our work brings several advances over a recent study, which trained a graph neural network to predict barrier heights on a similar dataset calculated at a lower level of theory.[91] First, our model is trained on a refined dataset that allows barrier height predictions to approach coupled-cluster accuracy. As shown below, our new model's predictions are about 1 kcal mol^{-1} more accurate than a good density functional theory (DFT) calculation and offer a $\sim 10^5$ factor speedup. Our model also uses an improved reaction representation, which substantially improves the accuracy, and it uses 75% fewer parameters than models from previously published works,[62,92] making our lightweight model more practical to integrate into prediction workflows. Importantly, our training procedure uses proper data splits to accurately estimate the model's performance on unseen reactions. Finally, our work also compares performance of models using 2D and 3D information, highlighting opportunities for future innovation.

Methods

Dataset

To train our model, we leverage a recently refined gas-phase dataset of elementary reactions with atom-mapped SMILES.[42,93] These data span a diverse set of reactions, whose neutral molecules involve carbon, hydrogen, nitrogen, and oxygen and contain up to seven heavy atoms. Reactions are available at three levels of theory: B97-D3/def2-mSVP, ωB97X-D3/def2-TZVP, and CCSD(T)-F12a/cc-pVDZ-F12//ωB97X-D3/def2-TZVP. Data from all

levels of theory are used in a transfer learning approach, which first trains the model on a larger dataset with lower accuracy and then fine-tunes the model with a slightly smaller dataset with higher accuracy calculations. Similar to in ref. 62 and 94, the model is initialized with the final weights from the previous run and then uses a smaller learning rate. The final model is evaluated against the coupled-cluster values. At each level, the zero-point energies from the harmonic vibrational analysis were added to the reactant, product, and TS energies. These energies used scaled vibrational frequencies to account for anharmonic effects.[95,96] The barrier heights were calculated by subtracting the resulting TS and reactant energies. Reaction enthalpies were calculated by subtracting the resulting product and reactant energies, which include bond additivity corrections as described in ref 42. These reactions largely overlap with those used to train the model in ref. 62. However, ref. 42 refined the energies of each species using coupled-cluster calculations. All of the earlier works fitting these reactions[62,84,92] only had access to DFT energies. Suspecting convergence errors in the quantum chemistry, we removed any reaction in which the coupled-cluster reaction enthalpy is over 10 kcal mol^{-1} above the barrier height.

All reactions contain one reactant and one to three products. All the transition states and unimolecular reactants have an even number of electrons, and were computed as singlets (S=0). Almost all of the products also have an even number of electrons. However, there are 49 reactions unique to the larger B97-D3 dataset whose product was a pair of radicals (each product had an odd number electrons). We modeled these odd-electron species as doublets (S=$\frac{1}{2}$). We augment our data by including the reverse reactions, which results in approximately 33,000 reactions at B97-D3 and 24,000 reactions at both the ωB97X-D3 and CCSD(T)-F12a levels. When creating training, validation, and testing sets, we use five folds, each with an 85:5:10 split. To create these sets, we use a scaffold split on the reactant SMILES, which partitions the data based on the Bemis-Murcko scaffold[97] as calculated by RDKit.[98] Scaffold splits are a better measure of generalizability compared to random splits.[92,99–103] When assigning reactions to each set, we place each pair of forward and reverse

reactions in the same set. Otherwise, if we separate forward and reverse reactions between sets, the model's testing error will be polluted by data leakage–the same transition state would appear in both the training and test set–and so will not reflect the true performance of the model when evaluating a new reaction.[92]

2D D-MPNN Model

When developing data-driven methods to predict barrier heights, graph neural networks (GNNs) are a natural choice; here, molecules are abstracted as graphs where atoms are graph nodes and bonds are graph edges.[104] GNNs operate by updating representations of nodes or edges with the information from neighboring nodes or edges, propagating information throughout the graph. For our work, we adapt Chemprop,[99] a directed message passing neural network (D-MPNN), which is a type of GNN that passes messages across directed bonds. Like other GNNs, the D-MPNN architecture builds a learned molecular representation by aggregating atomic representations after the message passing phase and feeding this representation through a dense layer to predict the molecular property of interest.

Extending the D-MPNN architecture to reaction properties is an intriguing way to model barrier heights. Grambow et al.[62] altered Chemprop to predict reaction properties by embedding the reactant and product with the same D-MPNN, subtracting the learned atomic representations, and passing the aggregated molecular representation through a FFN to obtain the final prediction. Heid and Green[92] showed that using the established condensed graph of reaction (CGR) representation[105] improved Chemprop's performance on barrier height prediction. A key aspect of the CGR representation is that it removes disjoint graphs present in bimolecular reactions and allows message passing between all atoms, overcoming a limitation of the Grambow et al. method. Since CGR is a superposition of the reactant and product graphs, only one graph is input to the model, leaving the rest of Chemprop's architecture unchanged.

Here, we extend the CGR version of Chemprop by incorporating additional atom and

bond features. We also concatenate additional features before the dense layer readout, such as reaction enthalpy, to improve prediction performance. Table 1 shows the improvement from each modification, and the Supporting Information contains more detail about the network, training procedure, and hyperparameter optimization. Our modified code and final weights are freely available on GitHub.[106]

3D DimeReaction Model

Since models based on molecular geometries have worked well for property prediction,[107] neural network potentials,[108] and excited-state dynamics,[109] we also explored 3D networks for predicting the reaction barrier. We start with DimeNet++,[110] which is a faster and more accurate version of the original DimeNet model.[111] Both are directional message passing networks that operate on 3D coordinates and have shown promising results for property prediction of individual molecules, such as on the popular QM9 dataset.[112] Here, we extend the architecture so that it can predict reaction properties. We follow a similar approach from Grambow et al.[62] by passing the reactant and product through the same DimeNet++ model, subtracting the learned molecular representation, and passing the result to a dense layer to predict the regression target. Our modified code and final weights are freely available on GitHub.[113]

Our DimeReaction model is trained on the unimolecular reactions from the dataset. Since the model operates on molecular coordinates, reactions with multiple products should not be arbitrarily aligned in 3D space. However, properly translating and rotating those molecules to form a multi-product complex is beyond the scope of this work. Further, the unimolecular reactions account for about 70% of the data, giving nearly 17,000 reactions after augmenting with the reverse reactions which are also unimolecular. As before, a scaffold split is used to create the training, validation, and testing sets with an 85:5:10 split, and each pair of forward and reverse reactions is placed in the same set. The model is trained on only one fold of the data since preliminary results indicated sub-par performance relative to the 2D-MPNN

model. Additional detail can be found in the Supporting Information.

Results and Discussion

2D D-MPNN Model

The testing mean absolute error (MAE) is 2.89 ± 0.16 kcal mol^{-1} and the root-mean-square error (RMSE) is 4.90 ± 0.16 kcal mol^{-1}, such that the bounds correspond to one standard deviation calculated across five folds. The parity plot in Figure 2a shows the model's predictive power across the entire range of data; accuracy is maintained even in regions where the data are sparser. The residuals are centered around zero, indicating no systematic over- or under-prediction, which is further supported by the error histogram in Figure 2b. 95% of the reaction barriers are predicted within 6 kcal mol^{-1} of the calculated coupled-cluster value.

The overall testing performance is improved by using ensembles of models, which is consistent with previous literature.[99,114] For a given fold, different initializations are used when training a model on that data split. The predictions are then averaged from each ensemble and compared to the target values from the respective test set. Here, five different initializations are used for each of the five data splits, resulting in a total of 25 models. The weights for all 25 models are published on GitHub.[106] Ensembling lowers the testing MAE (RMSE) to 2.57 ± 0.13 (4.58 ± 0.16) kcal mol^{-1}.

It is also important to examine model performance specifically on low barrier reactions since these are the most feasible reaction pathways included in kinetic models. For example, filtering the testing reactions to those with $\Delta E_0 < 50$ kcal mol^{-1} gives a testing MAE (RMSE) of 2.07 ± 0.18 (3.76 ± 0.33) kcal mol^{-1}, which is a slight improvement compared to the test errors for all reactions. The subsequent analysis and all figures and tables use weights from the first initial seed within each fold since the purpose is simply to explain interesting trends rather than obtain the best predictions for each intermediate case.

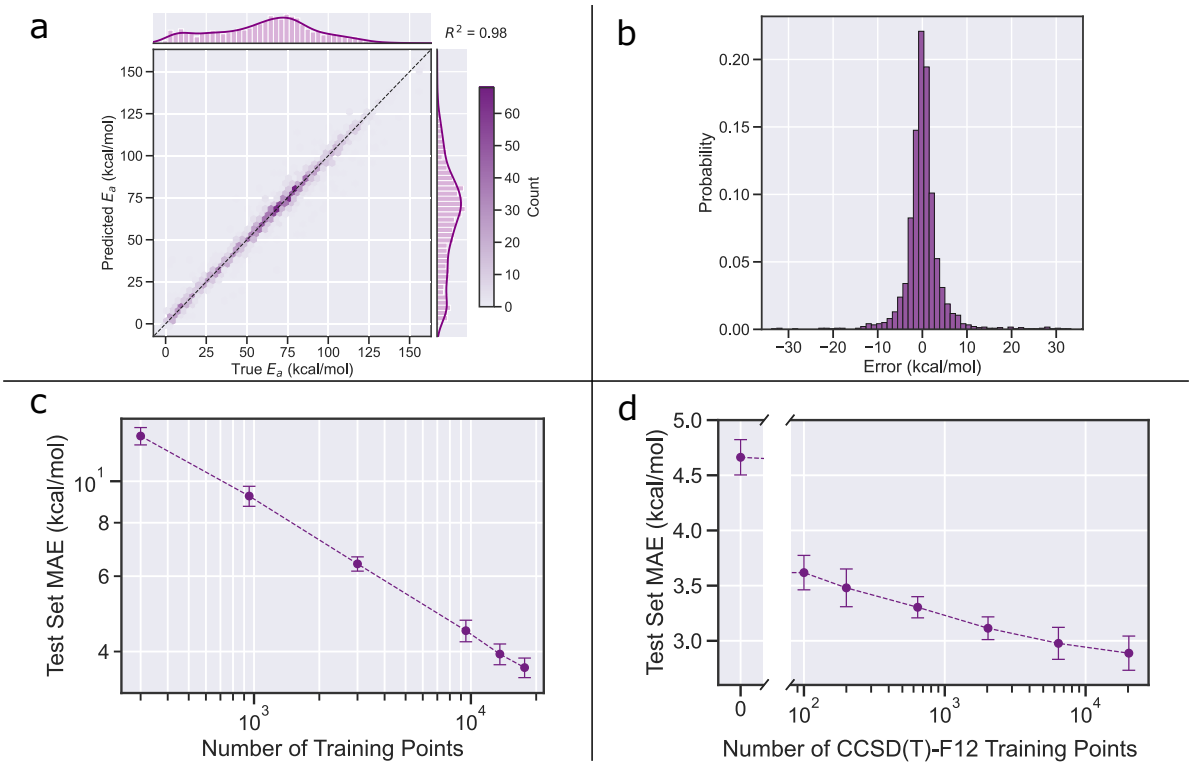


Figure 2: Deep learning model results for predicting reaction barriers. Error bars indicate one standard deviation calculated across the five folds. (a) Parity plot of model predictions vs "true" (CCSD(T)-F12a) barriers ΔE_0 for the first fold. (b) Histogram of testing errors (predicted minus "true") for the first fold. (c) Testing MAE vs. the number of data points at each stage of training. These reactions are present in all three level of theory datasets. (d) Testing MAE vs. number of coupled-cluster training reactions. Each point shows a model pretrained on all B97-D3 and ωB97X-D3 reactions and then fine-tuned with increasing amounts of high-accuracy data.

Figure 2c shows that the model strongly benefits from additional training data. The learning curve (how the model improves as more training data is added) has not yet shown asymptotic behavior, which implies further improvement may be possible with more data. The slope of the learning curve is similar to those we have observed when developing models for other chemical properties using Chemprop.[115] To ensure that the only variable changing was dataset size, each plotted point uses the exact same reactions and data splits for all three quantum chemistry datasets (one for each level of theory) for the transfer learning scheme.

As described in ref 42, a barrier height calculated at ωB97X-D3 has an MAE of about 3.5 kcal mol^{-1} from the coupled-cluster value. Thus, this is the ideal prediction error that

one could expect when training a model only on DFT data. Indeed, the previous model from Grambow et al.,[62] which was trained to predict ωB97X-D3 values, has an MAE of 4.74 ± 0.10 kcal mol^{-1} relative to the coupled-cluster values (approximately twice as large as the model error from this work). Thus, Figure 2d is particularly exciting since it shows that even a small amount of fine-tuning gives a meaningful improvement to our model's predictions. Each plotted point shows a model pretrained on both the B97-D3 and ωB97X-D3 datasets and then fine-tuned with increasing amounts of high accuracy data. The same coupled-cluster validation and test set is used during each run. Although the model certainly benefits from additional coupled-cluster values, performing such calculations, which scale as $O(N^7)$, may not always be practical for large datasets or for large molecules. However, we find that using these high-accuracy calculations for even just 50 reactions (augmented with the reverse to yield 100 total training reactions), lowers the testing MAE by over 1 kcal mol^{-1}. Similar behavior was observed in some of our prior transfer learning studies, where models built on large DFT data sets were fine-tuned using a small number of more accurate data.[94,115] We believe this works because the model trained from the DFT data on a large number of reactions has learned how a wide variety of chemical structures affect the barrier height. But the DFT numbers are rough and benefit from calibration using a few high-accuracy numbers. We suggest this type of fine-tuning may be a good strategy whenever DFT gives reasonable predictions of a chemical property, one can afford to generate a large DFT data set, and one has a smaller high-accuracy data set.

Table 1 demonstrates how various factors contribute to improving model performance. The biggest improvement comes from optimizing the hyperparameters (see Supporting Information) and allowing enough degrees of freedom for the model to capture the diverse chemistry in this dataset. Adding ring size to both the atom and bond features further improves the model. We also use tried using RDKit[98] to generate molecular feature vectors[116] for each SMILES and pass the difference of the product and reactant vector as input to the dense layer. However, this does not improve results. Some recent work has shown

that using quantum mechanical descriptors can improve model performance for reaction prediction. Similar to both Stuyver and Coley[63] and Guan et al.,[101] we use the D-MPNN network developed by Guan et al.[101] without modification to predict atomic descriptors that are concatenated with the learned atomic representations before the aggregated molecular representation is passed to the dense layer. However, we observe no effect of using these descriptors. Finally, using reaction enthalpy as an input to the dense layer, rather than co-training on both the barrier height and enthalpy as in ref 62, further improves the performance. This approach requires quantum calculations for the reactant and product to obtain the reaction enthalpy; however, this is often easier than obtaining a TS, so our model still offers substantial time savings compared to the traditional workflow outlined in Figure 1. Alternatively, the predicted enthalpy from our co-trained model could serve as input to the final model, which offers further time savings and only minimal increases in testing error (see Supporting Information).

Table 1: Sequential optimization study showing the improvement in testing error (kcal mol^{-1}). Each row includes all changes from the previous row. Results are from the first fold of cross-validation.

	B97-D3		**ωB97X-D3**		**CCSD(T)-F12a**	
Description	MAE	RMSE	MAE	RMSE	MAE	RMSE
Default	6.74	10.22	5.30	8.17	5.07	8.18
+Optimize hyperparameters	5.70	9.29	3.61	6.08	3.44	5.86
+Ring features	5.28	8.79	3.33	5.77	3.24	5.66
+Input dH	**4.81**	**7.88**	**3.06**	**5.15**	**2.98**	**4.96**

As mentioned above, creating proper data splits is essential to evaluating model performance. We use a scaffold split on the reactants from the forward reactions to create training, validation, and testing sets. The reverse reaction is then added to the corresponding set, which ensures that each set is independent. In contrast, the previous work from Grambow et al.[62] performed a scaffold split on the reactants from the forward and reverse reactions. However, since the reactant and product of the same reaction can have different scaffolds,

the forward reaction could be placed in the training set while the reverse reaction could be placed in the test set. Indeed, ∼80% of the testing reactions from Grambow et al. had their forward or reverse counterpart in the training set. This data leakage resulted in underestimating the true testing error since the regression task had accidentally been reframed as an easier problem.[92] Rather than asking the model to produce a barrier height for a new reaction, the data leakage meant that the model was mostly predicting a barrier height for a reaction it has already seen in either the forward or reverse direction. The dramatic impact on model performance is shown in Table 2.

Table 2: Importance of independent data splits on E_a testing errors (kcal mol^{-1}). Results are from the first fold of cross-validation.

		Proper	**B97-D3**		**ωB97X-D3**		**CCSD(T)-F12a**	
Model	**Hyperparameters**	**Splits**	MAE	RMSE	MAE	RMSE	MAE	RMSE
Grambow et al.	Grambow et al.	No	2.99	5.87	1.91	3.35	1.87	3.29
Grambow et al.	Grambow et al.	Yes	8.15	11.46	5.94	8.55	5.50	7.97
Grambow et al.	This work	Yes	7.16	10.85	5.13	7.71	4.31	6.93
This work	**This work**	**Yes**	**4.81**	**7.88**	**3.06**	**5.15**	**2.98**	**4.96**

3D DimeReaction Model

One hypothesis is that using 3D information may improve the predictive capabilities relative to using a 2D attributed graph. Since obtaining a TS geometry is difficult, our ideal DimeReaction model would receive an optimized structure for the reactant and product and then predict a barrier height. We find that model performance is better when only training on the coupled-cluster data, rather than performing transfer learning as described earlier. Still, results from this approach were quite poor with a barrier height testing MAE (RMSE) of 6.20 (10.10) kcal mol^{-1}. Although this large error is surprising, additional tests indicate that the DimeNet++ model functions as intended. For example, training a multi-task model that predicts both E_a and dH yields excellent performance on reaction enthalpy with a testing MAE of 1.78 kcal mol^{-1}. Thus, when the regression target is a function of the input geometries, such as using reactant and product structures to predict reaction enthalpy

rather than the barrier height, the model performs well. Additional results and analysis are provided in the Supporting Information.

Another potential workflow is using the difference between the learned representations of the reactant and TS geometries (rather than the reactant and product) since these directly correspond to a barrier height. To test this idea, we first considered the ideal scenario of using optimized TS structures. As expected, this resulted in a very low testing error—lower than our best 2D model—shown as the first row in Table 3. While it is chemically satisfying to see the reactant and TS encode more information about the barrier height than the reactant and product, using an optimized TS geometry here is not sensible, at least not for molecules in this size range. Once one has the optimal TS geometry, it is not expensive to perform a single-point coupled-cluster calculation and then proceed with canonical transition state theory to compute $k_\infty(T)$.

A more practical workflow involves quickly generating a TS guess and training a model to operate on this non-optimized structure. Several methods exist for generating TS guesses, from using hand-made templates[31–33] to using deep learning,[34–36] so it should be quite practical for future users to quickly obtain a TS guess. To first get an idea of how much a non-optimized TS structure would impact the results, we add increasing amounts of Gaussian noise to the optimized structures that were then used as input to the model. As seen in Table 3, the model is very sensitive to the TS structure, as testing performance quickly decreases when adding noise. We next use one TS guess predictor to quickly generate TS guesses. The graph network from Pattanaik et al.[34] had achieved a testing RMSD of 0.28 Å. Using this network to generate TS guesses that were used as input to our DimeReaction model, we obtain a testing MAE (RMSE) of 6.19 (9.26) kcal mol^{-1}, which is inline with the trend from Table 3.

Table 3: Impact of noisy TS inputs on DimeReaction's testing errors (kcal mol^{-1}). The first row corresponds to using an optimized TS structure with no added noise.

σ	RMSD (Å)	MAE	RMSE
0	0	2.25	3.91
0.05	0.086	3.62	5.66
0.10	0.173	4.81	7.16
0.15	0.260	6.28	8.83
0.20	0.346	7.42	10.42

The sensitivity of the 3D reaction model renders it impractical compared to the 2D workflow. However, it highlights an area of further research; future studies should investigate methods to quickly and accurately produce 3D TS guesses, which can subsequently be used in predicting kinetic parameters. Currently, available estimators achieve geometric errors on TS structure generation that are much too high for such modeling efforts. As another idea, the sensitivity could make DimeNet++ quite useful for screening which initial guess geometries are worth the computational expense for optimization and frequency calculation.

Conclusion

To accurately predict the time evolution of a reacting chemical system, one needs a quick way to estimate rate coefficients to decide which of the many conceivable reactions are important enough to include in the kinetic model. Due to the high computational cost, it is usually impractical to directly compute $k_\infty(T)$ for all of the reactions in a system of interest. Methods are needed to accelerate or bypass some of the computational bottlenecks in the conventional TST workflow based on high-accuracy quantum chemistry. Even rough estimates can be helpful in deciding which reactions require expensive calculations. The need for fast estimates of rate coefficients has been recognized for many decades, and excellent estimators have been developed for certain reaction families. However, data scarcity has made it difficult to develop models that generalize well over a broad range of reactions.

With recent advances in compute power, we can now routinely generate large datasets with DFT and sometimes higher-level quantum chemistry methods, opening up the possibility of constructing accurate estimators with much broader scope using machine learning techniques.

For instance, the generation of new high-quality kinetics datasets[42,91] allowed the 2D D-MPNN model presented here to quickly predict reliable values for barrier heights across a diverse set of thousands of gas-phase reactions. The model gives more accurate barrier heights on average than good (e.g. ωB97X-D3/def2-TZVP) DFT calculations for vastly lower computational cost. By directly estimating barrier heights, we anticipate the model will be useful during the automated generation of kinetic models by reducing the amount of computational effort devoted to reactions which cannot be important. The new model can also quickly rank the most feasible (lower barrier) reactions produced from an automated enumeration of possible reactions, making the process of discovering new low-barrier reactions much more efficient. Combining this barrier height estimator with other methods for computing or estimating Arrhenius A-factors would allow rapid estimation of reasonable rate coefficients.

Going forward, existing kinetics datasets should be expanded. As shown in Figure 2c, additional data would be beneficial for reducing the errors in predicted barrier heights. Although the dataset from Spiekermann et al.[42] is the largest coupled-cluster reaction barrier dataset to our knowledge, it is still relatively limited; its species contain at most only seven heavy atoms, they only include the elements H, C, N, and O, all transition states are closed shell singlets, no ions are included, and there are only gas-phase reactions. The dataset used in this work also contains only a single conformer for each species, so some of the prediction error may be due to not using the lowest energy conformers when calculating the barrier height. Future dataset generation should perform thorough conformer searches. This would also facilitate computation of $k_\infty(T)$ for reactions where the reactant(s) and/or TS are non-rigid e.g. because they contain internal rotors.

When it comes to training models, using proper data splits is essential to evaluating

model performance. As shown by Table 2, each pair of forward and reverse reactions must be placed in the same set to ensure that the training, validation, and testing sets are independent. Otherwise, the model's testing error will appear unrealistically low since it has been polluted by data leakage. Scaffold splitting is also considered a better measure of generalizability.[92,99–103] In contrast, random splitting creates an easier prediction task that is less useful at measuring extrapolation capabilities. Finally, we emphasize the importance of fine-tuning models using whatever high-accuracy data is available. As shown in Figure 2d, even small amounts of high-accuracy data give meaningful improvements to model predictions.

Acknowledgement

We acknowledge financial support from the Gas Phase Chemical Physics Program of the U.S. Department of Energy (DOE), Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences under Award number DE-SC0014901. LP is funded by the Machine Learning for Pharmaceutical Discovery and Synthesis (MLPDS) consortium and the MIT-Takeda Fellowship. The authors acknowledge the MIT SuperCloud and Lincoln Laboratory Supercomputing Center for providing computing resources that have contributed to the research results reported within this paper. The authors also thank Dr. Esther Heid for useful discussions about data splitting.

Supporting Information Available

Detailed description of network architecture, training procedure, and hyperparameter optimization for both our modified Chemprop and DimeReaction models; comparison of CGR representations; results from using RDKit molecular features and atomic quantum descriptors; results from using a predicted reaction enthalpy; model performance on different quantum methods.

Biographies

Kevin Spiekermann was born in California and attended the University of California San Diego where he majored in Chemical Engineering. He earned a M.S. in Chemical Engineering Practice from MIT and is currently pursuing a Ph.D. in Chemical Engineering, also at MIT. His research includes improving the automated generation and refinement of kinetic mechanisms, part of which includes training models that can quickly estimate barrier heights.

Lagnajit (Lucky) Pattanaik is a 5th year graduate student co-advised in the groups of Prof. William H. Green and Klavs F. Jensen. He earned his B.S. in Chemical Engineering at The Ohio State University, where he performed experimental research developing heterogeneous catalysts for biomass conversion. He currently attends MIT, where he has earned a M.S. in Chemical Engineering Practice and is pursuing a Ph.D. in Chemical Engineering. His research primarily focuses on developing 3D graph neural models for applications in reaction kinetics and drug discovery.

Professor William H. Green received his B.A. in Chemistry (with Highest Honors) from Swarthmore College in 1983 and earned his Ph.D. in Physical Chemistry from the University of California, Berkeley in 1988. After postdoctoral research at Cambridge University and the University of Pennsylvania, he was a principal investigator at Exxon's Corporate Research Laboratory before joining the MIT faculty in 1997. Professor Green leads a research group focusing on reaction kinetics and related computational methods, including machine learning. He holds MIT's Hoyt C. Hottel Chair, and is a Fellow of the AAAS and of the Combustion Institute. He has earned many awards, including the ACS Glenn Award in Fuel Chemistry and the AIChE Wilhelm Award in Reaction Engineering.

References

(1) Van't Hoff, J. H. *Etudes De Dynamique Chimique*; Muller, 1884; Vol. 1.

(2) Arrhenius, S. Über Die Reaktionsgeschwindigkeit Bei Der Inversion Von Rohrzucker Durch SÄUren. *Zeitschrift für physikalische Chemie* **1889**, *4*, 226–248, a translation of the four pages in this paper that deal with temperature dependence is included in

Back, M. H., and Laidler, K. J., "Selected Readings in Chemical Kinetics," Pergamon Press, Oxford, 1967, pp. 31-35.

(3) Montreal Protocol on Substances That Deplete the Ozone Layer. *Washington, DC: US Government Printing Office* **1987**, *26*, 128–136.

(4) Velders, G. J.; Andersen, S. O.; Daniel, J. S.; Fahey, D. W.; McFarland, M. The Importance of the Montreal Protocol in Protecting Climate. *Proc. Natl. Acad. Sci.* **2007**, *104*, 4814–4819.

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

(6) Gao, H.; Struble, T. J.; Coley, C. W.; Wang, Y.; Green, W. H.; Jensen, K. F. Using Machine Learning to Predict Suitable Conditions for Organic Reactions. *ACS Cent. Sci.* **2018**, *4*, 1465–1476.

(7) Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. Current Status of Transition-State Theory. *J. Phys. Chem.* **1996**, *100*, 12771–12800.

(8) Klippenstein, S. J. From Theoretical Reaction Dynamics to Chemical Modeling of Combustion. *Proc. Combust. Inst.* **2017**, *36*, 77–111.

(9) Green, W. H. Moving from Postdictive to Predictive Kinetics in Reaction Engineering. *AIChE Journal* **2020**, *66*, e17059.

(10) Pio, G.; Dong, X.; Salzano, E.; Green, W. H. Automatically Generated Model for Light Alkene Combustion. *Combust. Flame* **2022**, *241*, 112080.

(11) Payne, A. M.‡; Spiekermann, K. A.‡; Green, W. H., Detailed Reaction Mechanism for 350-400°C Pyrolysis of an Alkane, Aromatic, and Long-Chain Alkylaromatic Mixture. *Energy Fuels* **2022**, *36*, 1635–1646.

(12) Vermeire, F. H.; Aravindakshan, S. U.; Jocher, A.; Liu, M.; Chu, T.-C.; Hawtof, R. E.; Van de Vijver, R.; Prendergast, M. B.; Van Geem, K. M.; Green, W. H. Detailed Kinetic Modeling for the Pyrolysis of a Jet A Surrogate. *Energy Fuels* **2022**, *36*, 1304–1315.

(13) Wu, H.; Grinberg Dana, A.; Ranasinghe, D. S.; Pickard IV, F. C.; Wood, G. P.; Zelesky, T.; Sluggett, G. W.; Mustakis, J.; Green, W. H. Kinetic Modeling of API Oxidation: 2. Imipramine Stress Testing. *Mol. Pharmaceutics* **2022**, *19*, 1526–1539.

(14) Grambow, C. A.; Jamal, A.; Li, Y.-P.; Green, W. H.; Zador, J.; Suleimanov, Y. V. Unimolecular Reaction Pathways of a γ-Ketohydroperoxide from Combined Application of Automated Reaction Discovery Methods. *J. Am. Chem. Soc.* **2018**, *140*, 1035–1048.

(15) Maeda, S.; Harabuchi, Y. On Benchmarking of Automated Methods for Performing Exhaustive Reaction Path Search. *J. Chem. Theory Comput.* **2019**, *15*, 2111–2115.

(16) Pitzer, K. S.; Gwinn, W. D. Energy Levels and Thermodynamic Functions for Molecules with Internal Rotation I. Rigid Frame with Attached Tops. *J. Chem. Phys.* **1942**, *10*, 428–440.

(17) Truhlar, D. G. A Simple Approximation for the Vibrational Partition Function of a Hindered Internal Rotation. *J. Comput. Chem.* **1991**, *12*, 266–270.

(18) Pfaendtner, J.; Yu, X.; Broadbelt, L. J. The 1-D Hindered Rotor Approximation. *Theor. Chem. Acc.* **2007**, *118*, 881–898.

(19) Sharma, S.; Raman, S.; Green, W. H. Intramolecular Hydrogen Migration in Alkylperoxy and Hydroperoxyalkylperoxy Radicals: Accurate Treatment of Hindered Rotors. *J. Phys. Chem. A* **2010**, *114*, 5689–5701.

(20) Zheng, J.; Yu, T.; Papajak, E.; Alecu, I.; Mielke, S. L.; Truhlar, D. G. Practical Methods for Including Torsional Anharmonicity in Thermochemical Calculations on Complex Molecules: The Internal-Coordinate Multi-Structural Approximation. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10885–10907.

(21) Yu, T.; Zheng, J.; Truhlar, D. G. Multi-Structural Variational Transition State Theory. Kinetics of the 1, 4-Hydrogen Shift Isomerization of the Pentyl Radical with Torsional Anharmonicity. *Chem. Sci.* **2011**, *2*, 2199–2213.

(22) Zheng, J.; Truhlar, D. G. Quantum Thermochemistry: Multistructural Method with Torsional Anharmonicity Based on a Coupled Torsional Potential. *J. Chem. Theory Comput.* **2013**, *9*, 1356–1367.

(23) Zheng, J.; Meana-Pañeda, R.; Truhlar, D. G. MSTor Version 2013: A New Version of the Computer Code for the Multi-Structural Torsional Anharmonicity, Now With a Coupled Torsional Potential. *Comput. Phys. Commun.* **2013**, *184*, 2032–2033.

(24) Gonzalez-Lavado, E.; Corchado, J. C.; Suleimanov, Y. V.; Green, W. H.; Espinosa-Garcia, J. Theoretical Kinetics Study of the $O(^3P)$+ CH_4/CD_4 Hydrogen Abstraction Reaction: the Role of Anharmonicity, Recrossing Effects, and Quantum Mechanical Tunneling. *J. Phys. Chem. A* **2014**, *118*, 3243–3252.

(25) Zheng, J.; Bao, J. L.; Meana-Pañeda, R.; Zhang, S.; Lynch, B. J.; Corchado, J. C.; Chuang, Y.-Y.; Fast, P. L.; Hu, W.-P.; Liu, Y.-P. et al. *Polyrate*-version 2017-C. University of Minnesota: Minneapolis, 2017.

(26) Goldman, M. J.; Ono, S.; Green, W. H. Correct Symmetry Treatment for X+ X Reactions Prevents Large Errors in Predicted Isotope Enrichment. *J. Phys. Chem. A* **2019**, *123*, 2320–2324.

(27) Hawkins, P. C.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Struc-

tures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572–584.

(28) Riniker, S.; Landrum, G. A. Better Informed Distance Geometry: Using What We Know to Improve Conformation Generation. *J. Chem. Inf. Model.* **2015**, *55*, 2562–2574.

(29) Pracht, P.; Bohle, F.; Grimme, S. Automated Exploration of the Low-Energy Chemical Space with Fast Quantum Chemical Methods. *Phys. Chem. Chem. Phys.* **2020**, *22*, 7169–7192.

(30) Ganea, O.; Pattanaik, L.; Coley, C.; Barzilay, R.; Jensen, K.; Green, W.; Jaakkola, T. GeoMol: Torsional Geometric Generation of Molecular 3D Conformer Ensembles. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 13757–13769.

(31) Bhoorasingh, P. L.; West, R. H. Transition State Geometry Prediction Using Molecular Group Contributions. *Phys. Chem. Chem. Phys.* **2015**, *17*, 32173–32182.

(32) Harms, N.; Underkoffler, C.; West, R. Advances in Automated Transition State Theory Calculations: Improvements on the AutoTST Framework. *10.26434/chemrxiv* **2020**, *13277870.v2*, Accessed 2022-03-08.

(33) Van de Vijver, R.; Zádor, J. KinBot: Automated Stationary Point Search on Potential Energy Surfaces. *Comput. Phys. Commun.* **2020**, *248*, 106947.

(34) Pattanaik, L.; Ingraham, J. B.; Grambow, C. A.; Green, W. H. Generating Transition States of Isomerization Reactions with Deep Learning. *Phys. Chem. Chem. Phys.* **2020**, *22*, 23618–23626.

(35) Jackson, R.; Zhang, W.; Pearson, J. TSNet: Predicting Transition State Structures with Tensor Field Networks and Transfer Learning. *Chem. Sci.* **2021**, *12*, 10022–10040.

(36) Makoś, M. Z.; Verma, N.; Larson, E. C.; Freindorf, M.; Kraka, E. Generative Adversarial Networks for Transition State Geometry Prediction. *J. Chem. Phys.* **2021**, *155*, 024116.

(37) Bischoff, F. A.; Wolfsegger, S.; Tew, D. P.; Klopper, W. Assessment of Basis Sets for F12 Explicitly-Correlated Molecular Electronic-Structure Methods. *Mol. Phys.* **2009**, *107*, 963–975.

(38) Knizia, G.; Adler, T. B.; Werner, H.-J. Simplified CCSD(T)-F12 Methods: Theory and Benchmarks. *J. Chem. Phys.* **2009**, *130*, 054104.

(39) Adler, T. B.; Knizia, G.; Werner, H.-J. A Simple and Efficient CCSD(T)-F12 Approximation. *J. Chem. Phys.* **2007**, *127*, 221106.

(40) Zheng, J.; Zhao, Y.; Truhlar, D. G. The DBH24/08 Database and Its Use to Assess Electronic Structure Model Chemistries for Chemical Reaction Barrier Heights. *J. Chem. Theory Comput.* **2009**, *5*, 808–821.

(41) Pfeiffer, F.; Rauhut, G.; Feller, D.; Peterson, K. A. Anharmonic Zero Point Vibrational Energies: Tipping the Scales in Accurate Thermochemistry Calculations? *J. Chem. Phys.* **2013**, *138*, 044311.

(42) Spiekermann, K. A.; Pattanaik, L.; Green, W. H. High Accuracy Barrier Heights, Enthalpies, and Rate Coefficients for Chemical Reactions. *Sci. Data* **2022**, Submitted.

(43) Feller, D.; Peterson, K. A.; Dixon, D. A. A Survey of Factors Contributing to Accurate Theoretical Predictions of Atomization Energies and Molecular Structures. *J. Chem. Phys.* **2008**, *129*, 204105.

(44) Evans, M.; Polanyi, M. Inertia and Driving Force of Chemical Reactions. *Trans. Faraday Soc.* **1938**, *34*, 11–24.

(45) Hammett, L. P. The Effect of Structure Upon the Reactions of Organic Compounds. Benzene Derivatives. *J. Am. Chem. Soc.* **1937**, *59*, 96–103.

(46) Benson, S. W. *Thermochemical Kinetics*; Wiley, 1976.

(47) Benson, S. W.; Buss, J. H. Additivity Rules for the Estimation of Molecular Properties. Thermodynamic Properties. *J. Chem. Phys.* **1958**, *29*, 546–572.

(48) Cohen, N.; Benson, S. Estimation of Heats of Formation of Organic Compounds by Additivity Methods. *Chem. Rev.* **1993**, *93*, 2419–2438.

(49) Saeys, M.; Reyniers, M.-F.; Marin, G. B.; Van Speybroeck, V.; Waroquier, M. Ab Initio Group Contribution Method for Activation Energies for Radical Additions. *AIChE Journal* **2004**, *50*, 426–444.

(50) Saeys, M.; Reyniers, M.-F.; Van Speybroeck, V.; Waroquier, M.; Marin, G. B. Ab Initio Group Contribution Method for Activation Energies of Hydrogen Abstraction Reactions. *ChemPhysChem* **2006**, *7*, 188–199.

(51) Adamczyk, A. J.; Reyniers, M.-F.; Marin, G. B.; Broadbelt, L. J. Exploring 1, 2-Hydrogen Shift in Silicon Nanoparticles: Reaction Kinetics from Quantum Chemical Calculations and Derivation of Transition State Group Additivity Database. *J. Phys. Chem. A* **2009**, *113*, 10933–10946.

(52) Sumathi, R.; Carstensen, H.-H.; Green, W. H. Reaction Rate Prediction via Group Additivity Part 1: H Abstraction from Alkanes by H and CH_3. *J. Phys. Chem. A* **2001**, *105*, 6910–6925.

(53) Sumathi, R.; Green Jr, W. H. A Priori Rate Constants for Kinetic Modeling. *Theor. Chem. Acc.* **2002**, *108*, 187–213.

(54) Green, W. H. Predictive Kinetics: A New Approach for the 21st Century. *Adv. Chem. Eng.* **2007**, *32*, 1–50.

(55) Liu, M.; Grinberg Dana, A.; Johnson, M. S.; Goldman, M. J.; Jocher, A.; Payne, A. M.; Grambow, C. A.; Han, K.; Yee, N. W.; Mazeau, E. J. et al. Reaction Mechanism Generator v3.0: Advances in Automatic Mechanism Generation. *J. Chem. Inf. Model.* **2021**, *61*, 2686–2696.

(56) Wang, K.; Dean, A. M. Rate Rules and Reaction Classes. *Comput. Aided Chem. Eng.* **2019**, *45*, 203–257.

(57) Johnson, M. S.; Green, W. H. A Machine Learning Based Algorithm for Rate Estimation. 2019 AIChE Annual Meeting. 2019.

(58) Heid, E.; Goldman, S.; Sankaranarayanan, K.; Coley, C. W.; Flamm, C.; Green, W. H. EHreact: Extended Hasse Diagrams for the Extraction and Scoring of Enzymatic Reaction Templates. *J. Chem. Inf. Model.* **2021**, *61*, 4949–4961.

(59) Hawkins, P. C. Conformation Generation: The State of the Art. *J. Chem. Inf. Model.* **2017**, *57*, 1747–1756.

(60) Mansimov, E.; Mahmood, O.; Kang, S.; Cho, K. Molecular Geometry Prediction Using a Deep Generative Graph Neural Network. *Sci. Rep.* **2019**, *9*, 1–13.

(61) Simm, G. N.; Hernández-Lobato, J. M. A Generative Model for Molecular Distance Geometry. *arXiv* **2019**, *1909.11459*, Accessed 2022-03-08.

(62) Grambow, C. A.; Pattanaik, L.; Green, W. H. Deep Learning of Activation Energies. *J. Phys. Chem. Lett.* **2020**, *11*, 2992–2997.

(63) Stuyver, T.; Coley, C. W. Quantum Chemistry-Augmented Neural Networks for Reactivity Prediction: Performance, Generalizability, and Explainability. *J. Chem. Phys.* **2022**, *156*, 084104.

(64) 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*, 1163–1175.

(65) Heinen, S.; von Rudorff, G. F.; von Lilienfeld, O. A. Toward the Design of Chemical Reactions: Machine Learning Barriers of Competing Mechanisms in Reactant Space. *J. Chem. Phys.* **2021**, *155*, 064105.

(66) Zhong, S.; Hu, J.; Fan, X.; Yu, X.; Zhang, H. A Deep Neural Network Combined with Molecular Fingerprints (DNN-MF) to Develop Predictive Models for Hydroxyl Radical Rate Constants of Water Contaminants. *J. Hazard. Mater.* **2020**, *383*, 121141.

(67) Lu, J.; Zhang, H.; Yu, J.; Shan, D.; Qi, J.; Chen, J.; Song, H.; Yang, M. Predicting Rate Constants of Hydroxyl Radical Reactions with Alkanes Using Machine Learning. *J. Chem. Inf. Model.* **2021**, *61*, 4259–4265.

(68) Greaves, T. L.; McHale, K. S. S.; Burkart-Radke, R. F.; Harper, J. B.; Le, T. C. Machine Learning Approaches to Understand and Predict Rate Constants for Organic Processes in Mixtures Containing Ionic Liquids. *Phys. Chem. Chem. Phys.* **2021**, *23*, 2742–2752.

(69) Stewart, J. J. Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. *J. Mol. Model.* **2007**, *13*, 1173–1213.

(70) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. A Climbing Image Nudged Elastic Band Method for Finding Saddle Points and Minimum Energy Paths. *J. Chem. Phys.* **2000**, *113*, 9901–9904.

(71) Peters, B.; Heyden, A.; Bell, A. T.; Chakraborty, A. A Growing String Method for Determining Transition States: Comparison to the Nudged Elastic Band and String Methods. *J. Chem. Phys.* **2004**, *120*, 7877–7886.

(72) Behn, A.; Zimmerman, P. M.; Bell, A. T.; Head-Gordon, M. Efficient Exploration of Reaction Paths via a Freezing String Method. *J. Chem. Phys.* **2011**, *135*, 224108.

(73) Dewyer, A. L.; Argüelles, A. J.; Zimmerman, P. M. Methods for Exploring Reaction Space in Molecular Systems. *WIREs Comput. Mol. Sci.* **2018**, *8*, e1354.

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

(75) Takahashi, K.; Miyazato, I. Rapid Estimation of Activation Energy in Heterogeneous Catalytic Reactions via Machine Learning. *J. Comput. Chem.* **2018**, *39*, 2405–2408.

(76) Houston, P. L.; Nandi, A.; Bowman, J. M. A Machine Learning Approach for Prediction of Rate Constants. *J. Phys. Chem. Lett.* **2019**, *10*, 5250–5258.

(77) Nandi, A.; Bowman, J. M.; Houston, P. A Machine Learning Approach for Rate Constants. II. Clustering, Training, and Predictions for the $O(^3P) + HCl \rightarrow OH + Cl$ Reaction. *J. Phys. Chem. A* **2020**, *124*, 5746–5755.

(78) Peterson, A. A. Acceleration of Saddle-Point Searches with Machine Learning. *J. Chem. Phys.* **2016**, *145*, 074106.

(79) Pozun, Z. D.; Hansen, K.; Sheppard, D.; Rupp, M.; Müller, K.-R.; Henkelman, G. Optimizing Transition States via Kernel-Based Machine Learning. *J. Chem. Phys.* **2012**, *136*, 174101.

(80) Hermes, E. D.; Sargsyan, K.; Najm, H. N.; Zádor, J. Accelerated Saddle Point Refinement Through Full Exploitation of Partial Hessian Diagonalization. *J. Chem. Theory Comput.* **2019**, *15*, 6536–6549.

(81) Komp, E.; Valleau, S. Machine Learning Quantum Reaction Rate Constants. *J. Phys. Chem. A* **2020**, *124*, 8607–8613.

(82) von Rudorff, G. F.; Heinen, S. N.; Bragato, M.; von Lilienfeld, O. A. Thousands of Reactants and Transition States for Competing E2 and S2 Reactions. *Machine Learning: Sci. Tech.* **2020**, *1*, 045026.

(83) Bragato, M.; von Rudorff, G. F.; von Lilienfeld, O. A. Data Enhanced Hammett-Equation: Reaction Barriers in Chemical Space. *Chem. Sci.* **2020**, *11*, 11859–11868.

(84) Komp, E.; Valleau, S. Low-Cost Prediction of Molecular and Transition State Partition Functions via Machine Learning. *arXiv* **2022**, *2203.02621*, Accessed 2022-03-08.

(85) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: the Δ-Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087–2096.

(86) Meuwly, M. Machine Learning for Chemical Reactions. *Chem. Rev.* **2021**, *121*, 10218–10239.

(87) Lewis-Atwell, T.; Townsend, P. A.; Grayson, M. N. Machine Learning Activation Energies of Chemical Reactions. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2021**, e1593, `https://doi.org/10.1002/wcms.1593`.

(88) Komp, E.; Janulaitis, N.; Valleau, S. Progress Towards Machine Learning Reaction Rate Constants. *Phys. Chem. Chem. Phys.* **2022**, *24*, 2692–2705.

(89) Suleimanov, Y. V.; Green, W. H. Automated Discovery of Elementary Chemical Reaction Steps Using Freezing String and Berny Optimization Methods. *J. Chem. Theory Comput.* **2015**, *11*, 4248–4259.

(90) Koerstz, M.; Rasmussen, M. H.; Jensen, J. H. Fast and Automated Identification of Reactions with Low Barriers: the Decomposition of 3-Hydroperoxypropanal. *SciPost Chemistry* **2021**, *1*, 003.

(91) Grambow, C. A.; Pattanaik, L.; Green, W. H. Reactants, Products, and Transition States of Elementary Chemical Reactions Based on Quantum Chemistry. *Sci. Data* **2020**, *7*, 1–8.

(92) Heid, E.; Green, W. H. Machine Learning of Reaction Properties via Learned Representations of the Condensed Graph of Reaction. *J. Chem. Inf. Model.* **2021**, *62*, 2101–2110.

(93) Spiekermann, K. A.; Pattanaik, L.; Green, W. H. High Accuracy Barrier Heights, Enthalpies, and Rate Coefficients for Chemical Reactions. 2022; `https://doi.org/10.5281/zenodo.6618262`.

(94) Grambow, C. A.; Li, Y.-P.; Green, W. H. Accurate Thermochemistry with Small Data Sets: A Bond Additivity Correction and Transfer Learning Approach. *J. Phys. Chem. A* **2019**, *123*, 5826–5835.

(95) Scott, A. P.; Radom, L. Harmonic Vibrational Frequencies: An Evaluation of Hartree-Fock, Møller-Plesset, Quadratic Configuration Interaction, Density Functional Theory, and Semiempirical Scale Factors. *J. Phys. Chem.* **1996**, *100*, 16502–16513.

(96) Alecu, I.; Zheng, J.; Zhao, Y.; Truhlar, D. G. Computational Thermochemistry: Scale Factor Databases and Scale Factors for Vibrational Frequencies Obtained from Electronic Model Chemistries. *J. Chem. Theory Comput.* **2010**, *6*, 2872–2887.

(97) Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39*, 2887–2893.

(98) Landrum, G., et al. RDKit: Open-Source Cheminformatics. **2006**; `https://www.rdkit.org`.

(99) Yang, K.; Swanson, K.; Jin, W.; Coley, C.; Eiden, P.; Gao, H.; Guzman-Perez, A.;

Hopper, T.; Kelley, B.; Mathea, M. et al. Analyzing Learned Molecular Representations for Property Prediction. *J. Chem. Inf. Model.* **2019**, *59*, 3370–3388.

(100) Wang, A. Y.-T.; Murdock, R. J.; Kauwe, S. K.; Oliynyk, A. O.; Gurlo, A.; Brgoch, J.; Persson, K. A.; Sparks, T. D. Machine Learning for Materials Scientists: An Introductory Guide Toward Best Practices. *Chem. Mater.* **2020**, *32*, 4954–4965.

(101) Guan, Y.; Coley, C. W.; Wu, H.; Ranasinghe, D.; Heid, E.; Struble, T. J.; Pattanaik, L.; Green, W. H.; Jensen, K. F. Regio-Selectivity Prediction with a Machine-Learned Reaction Representation and On-The-Fly Quantum Mechanical Descriptors. *Chem. Sci.* **2021**, *12*, 2198–2208.

(102) Artrith, N.; Butler, K. T.; Coudert, F.-X.; Han, S.; Isayev, O.; Jain, A.; Walsh, A. Best Practices in Machine Learning for Chemistry. *Nat. Chem.* **2021**, *13*, 505–508.

(103) Greenman, K. P.; Green, W. H.; Gomez-Bombarelli, R. Multi-Fidelity Prediction of Molecular Optical Peaks with Deep Learning. *Chem. Sci.* **2022**, *13*, 1152–1162.

(104) Bacciu, D.; Errica, F.; Micheli, A.; Podda, M. A Gentle Introduction to Deep Learning for Graphs. *Neural Networks* **2020**, *129*, 203–221.

(105) Varnek, A.; Fourches, D.; Hoonakker, F.; Solov'ev, V. P. Substructural Fragments: An Universal Language to Encode Reactions, Molecular and Supramolecular Structures. *J. Comput.-Aided Mol. Des.* **2005**, *19*, 693–703.

(106) Spiekermann, K. A.; Pattanaik, L.; Green, W. H.; Yang, K.; Swanson, K.; Jin, W.; Coley, C.; Eiden, P.; Gao, H.; Guzman-Perez, A. et al. `https://github.com/kspieks/chemprop/tree/barrier_prediction`, Accessed 2022-05-16.

(107) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. Schnet–A Deep Learning Architecture for Molecules and Materials. *J. Chem. Phys.* **2018**, *148*, 241722.

(108) Devereux, C.; Smith, J. S.; Davis, K. K.; Barros, K.; Zubatyuk, R.; Isayev, O.; Roitberg, A. E. Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens. *J. Chem. Theory Comput.* **2020**, *16*, 4192–4202.

(109) Westermayr, J.; Gastegger, M.; Marquetand, P. Combining SchNet and SHARC: The SchNarc Machine Learning Approach for Excited-State Dynamics. *J. Phys. Chem. Lett.* **2020**, *11*, 3828–3834.

(110) Klicpera, J.; Giri, S.; Margraf, J. T.; Günnemann, S. Fast and Uncertainty-Aware Directional Message Passing for Non-Equilibrium Molecules. *arXiv* **2020**, *2011.14115*, Accessed 2022-03-08.

(111) Klicpera, J.; Groß, J.; Günnemann, S. Directional Message Passing for Molecular Graphs. *arXiv* **2020**, *2003.03123*, Accessed 2022-03-08.

(112) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; Von Lilienfeld, O. A. Quantum Chemistry Structures and Properties of 134 Kilo Molecules. *Sci. Data* **2014**, *1*, 1–7.

(113) Spiekermann, K. A.; Pattanaik, L.; Green, W. H. `https://github.com/kspieks/DimeReaction`, Accessed 2022-04-11.

(114) Dietterich, T. G. Ensemble Methods in Machine Learning. Multiple Classifier Systems. MCS 2000. Lecture Notes in Computer Science, vol 1857. Springer, Berlin, Heidelberg. 2000; pp 1–15, https://link.springer.com/chapter/10.1007/3-540-45014-9_1.pdf.

(115) Vermeire, F. H.; Green, W. H. Transfer Learning for Solvation Free Energies: From Quantum Chemistry to Experiments. *Chem. Eng. J.* **2021**, *418*, 129307.

(116) Kelley, B. Descriptor Computation (Chemistry) and (Optional) Storage for Machine Learning. DescriptaStorus, version 2.2.0. `https://github.com/bp-kelley/descriptastorus`, Accessed 2021-09-03.

Supporting Information:

Fast Predictions of Reaction Barrier Heights:

Towards Coupled-Cluster Accuracy

Kevin A. Spiekermann, Lagnajit Pattanaik, and William H. Green*

Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, United States

E-mail: whgreen@mit.edu

Contents

S1: Graph Network Description **S3**

S1.1: Architecture and Featurization S3

S1.2: Training and Hyperparameter Optimization S6

S1.3: Fitting Error vs. Level Of Theory S8

S2: DimeReaction **S9**

References **S12**

S1: Graph Network Description

S1.1: Architecture and Featurization

For our experiments, we rely on Chemprop, a deep learning property prediction framework built by Yang et al.[S1] Chemprop takes SMILES[S2] strings as input and outputs the single or multiple properties of interest. While Chemprop is really a general-use framework that can train a range of machine learning models, the heart of Chemprop's innovation is their directed message passing neural network (D-MPNN) architecture, which we adapt here. Crucially, we include the additional architectural choices made by the original authors (i.e. input and output neural layers), which makes the D-MPNN architecture and the resultant Chemprop property prediction framework so successful.

The D-MPNN method builds on the formalization of message passing neural networks (MPNNs) developed by Gilmer et al.[S3] Generally, MPNNs operate on molecules by representing them as attributed graphs with nodes (atoms) and edges (covalent bonds) with associated feature vectors. During the message passing phase, hidden representations of nodes are updated with information from neighboring nodes and edges; this builds updated feature vectors for all nodes in the graph. During the readout phase, atom representations are aggregated and fed through a dense layer for property prediction. Yang et al.[S1] introduce messages over directed edges rather than over nodes, which makes the resulting D-MPNN architecture more expressive than traditional MPNNs. Gasteiger et al. recently formalized the D-MPNN concept as message passing over the *line graph* of the original molecular graph, which we see as a useful way of thinking about the D-MPNN architecture.

Since we are learning properties of reactions and not individual molecules, we must featurize multiple molecular graphs (i.e. reactants and products). Grambow et al.[S5] accomplished this by individually featurizing the reactant and product graphs with D-MPNNs and subtracting the learned product atomic representations from those of the reactant. We do not use this strategy in our work. Here, we use the established condensed graph of re-

action (CGR) representation,[S6] which was recently incorporated in Chemprop.[S7] CGR is a superposition of the reactant and product graphs; thus it requires atom-mapped reactions to compare the atoms and bonds between the reactant and product (i.e. we must know which reactant atoms led to which product atoms, usually defined by an annotation within the reaction SMILES string). A key aspect of the CGR representation is that it removes disjoint graphs present in bimolecular reactions and allows message passing between all atoms, a limitation of the Grambow et al. method. Using the CGR representation as input allows the rest of the Chemprop architecture to remain unchanged.

Upon creating the new condensed graph, initial feature vectors are created with RDKit[S8] for each atom and bond for both the reactant and product. This creates three possible combinations of features for the single CGR. First, the feature vectors from the reactant and product can be concatenated together (`reac_prod`). Second, the reactant feature vector can be concatenated with the difference of the product and reactant feature vectors (`reac_diff`). Third, the product feature vector can be concatenated with the difference of the reactant and product feature vectors (`prod_diff`). More detail about CGR in Chemprop, along with performance on benchmark datasets, can be found in the original publication.[S7]

Heid and Green[S7] reported that the `reac_diff` CGR representation usually performed best. However, the `reac_diff` representation of forward and reverse reactions is identical to the `prod_diff` representation of reverse and forward reactions. Since we had augmented the datasets at each level of theory with the reverse reactions, we arbitrarily chose the `reac_diff` CGR representation. Empirically, Table S1 shows that all three CGR representations give very similar performance; all models use the optimal hyperparameters described in Table S4 and Table S5.

Table S1: Comparison of CGR Representations. Errors are in kcal mol^{-1}, and results are from the first fold of cross-validation.

	B97-D3		ωB97X-D3		CCSD(T)-F12a	
Representation	MAE	RMSE	MAE	RMSE	MAE	RMSE
`reac_diff`	4.81	7.88	3.06	5.15	2.98	4.96
`prod_diff`	4.75	7.78	3.06	5.20	2.84	5.00
`reac_prod`	4.89	7.80	3.17	5.33	2.96	5.05

We tried several modifications to the model architecture. By default, the initial features in Chemprop simply assign whether the bond is in a ring of any size. We modify this by specifying the actual ring size (mapped to a one-hot vector) and incorporating this for both the atom and bond feature vectors. As shown in Table S2, this addition improves the testing RMSE by about 0.4 kcal mol^{-1}. We also allow additional features to be concatenated to the learned molecular representation that is passed to the feed forward network during readout. We find that using RDKit molecular features[1] and quantum atomic descriptors from Guan et al.[S9] have no effect.

Table S2: Sequential optimization study showing the improvement in testing error (kcal mol^{-1}). Each row includes all changes from the previous row. Results are from the first fold of cross-validation.

	B97-D3		ωB97X-D3		CCSD(T)-F12a	
Description	MAE	RMSE	MAE	RMSE	MAE	RMSE
Default	6.74	10.22	5.30	8.17	5.07	8.18
+Optimize hyperparameters	5.70	9.29	3.61	6.08	3.44	5.86
+Ring features	5.28	8.79	3.33	5.77	3.24	5.66
+Molecular RDKit features	5.41	8.90	3.42	5.90	3.22	5.66
+Atom QM Descriptors	5.25	8.72	3.39	5.83	3.13	5.60

In contrast, using reaction enthalpy as an input to the dense layer improves model performance. Thus, our model enables a workflow for predicting high-quality barrier heights without the need for a transition state (TS). Researchers only need to obtain the reaction enthalpy by running quantum calculations for the reactant(s) and product(s). One potential

[1] https://github.com/bp-kelley/descriptastorus

drawback of this approach is that quantum calculations can be expensive depending on the method and basis. An alternative workflow could instead quickly obtain a predicted value as input to the final model, an approach commonly done by other published works for property prediction.[S10,S11] For example, we could instead use the model from Grambow et al.,[S5] which was co-trained on E_a and dH (without atom or bond corrections) at the ωB97X-D3/def2-TZVP level of theory, to quickly predict dH. Similarly, we also co-trained a model on E_a and dH (with atom and bond corrections) at the CCSD(T)-F12a/cc-pVDZ-F12//ωB97X-D3/def2-TZVP level of theory in this work; testing errors are shown in the third row of Table S2. Using the predicted enthalpy from either of these models as input to our final model would be more efficient than using quantum calculations to determine the reaction enthalpy. As seen in Table S3, using either of these predicted enthalpy values yields very comparable performance, with just a small 0.17 kcal mol^{-1} decline in testing MAE as compared to using the calculated enthalpy. However, these results may be biased since both the model from Grambow et al.[S5] and from this work have already seen nearly all of these reactions during training; the accuracy when extrapolating to new reactions may be lower.

Table S3: Testing error (kcal mol^{-1}) when using reaction enthalpy calculated at the respective level of theory or predicted using a pretrained model from Grambow et al.[S5] or this work. Results are from the first fold of cross-validation.

	B97-D3		**ωB97X-D3**		**CCSD(T)-F12a**	
Description	MAE	RMSE	MAE	RMSE	MAE	RMSE
Input dH (predicted[S5])	5.18	8.26	3.40	5.59	3.18	5.36
Input dH (our predictions)	4.94	8.08	3.26	5.36	3.15	5.21
Input dH (QM calculation)	**4.81**	**7.88**	**3.06**	**5.15**	**2.98**	**4.96**

S1.2: Training and Hyperparameter Optimization

Training, validation, and testing sets are created using a scaffold split on the reactant SMILES, which partitions the data based on the Bemis-Murcko scaffold[S12] as calculated by RDKit. The exact procedure is described in ref S1. Intuitively, a scaffold split is more chal-

lenging than a random split since validation and testing molecules are deliberately chosen to have some class imbalance relative to the training set. This causes the testing performance to be a better measure of extrapolation to new molecules rather than interpolation with molecules very similar to those in the training set. As described in the main text, scaffold splitting was done based on the reactants from the forward reactions. Each pair of forward and reverse reactions is placed in the same set, which ensures that each set is independent. The data is split into 85% training, 5% validation, and 10% testing data. To avoid bias that may be introduced from using just one split, we use 5-fold cross-validation. Performance on the validation set is used to determine the best model weights as well as to choose the optimal hyperparameters. We use the Noam learning rate scheduler from Vaswani et al.[S13] during training, which starts by linearly increasing the learning rate from the initial to the maximum value over a specified number of warm-up epochs. Then the learning rate is exponentially decreased to the final value over the remaining epochs.

We utilize data at all levels of theory during training. First, the model is pretrained with the lowest level data from B97-D3/def2-mSVP. The model is then fine-tuned with the ωB97X-D3/def2-TZVP data and subsequently with the high-quality CCSD(T)-F12a/cc-pVDZ-F12//ωB97X-D3/def2-TZVP data. We find that model performance is slightly better when the message passing weights are not frozen after the first round of pretraining. Instead, the model weights are initialized using the best weights from the previous training run, and all weights are fine-tuned. We use the provided hyperparameter search code from Chemprop[S1] to determine the hyperparameters controlling the model architecture and fitting procedure, which are summarized in Table S4 and Table S5 respectively. Hyperparameters not shown in the table used the default values from Chemprop.

Table S4: Optimized hyperparameters for model architecture.

Hyperparameter	Value
Hidden size	900
Hidden size	4
Activation function	Leaky ReLU
Aggregation	sum
Additional FFN Inputs	dH

Table S5: Optimized hyperparameters for training.

Hyperparameter	Pretraining	Fine-Tuning
Epochs	65	55
Initial learning rate	10^{-4}	10^{-5}
Maximum learning rate	10^{-3}	10^{-4}
Final learning rate	10^{-5}	10^{-6}
Warm-up epochs	5	5
Gradient clip	10	10

S1.3: Fitting Error vs. Level Of Theory

Table S6 shows the testing error from training a model on each level of theory (no transfer learning). Only reactions present in all levels are used i.e. each model had the same data splits for training, validation, and testing so the only variable changing was the level of theory for the regression target. We use the optimal hyperparameters from Table S4 and Table S5. Our results show that is about equally difficult to fit a model to each level of theory; of course, the predictions will be better from a model trained on the higher level dataset.

Table S6: Barrier height testing error (kcal mol^{-1}) from training a model on each level of theory.

Level of Theory	MAE	RMSE
B97D3	3.95	6.68
ωB97X-D3	3.94	7.04
CCSD(T)-F12a	4.07	7.01

S2: DimeReaction

Our DimeReaction model is an extension of the recently published DimeNet++,[S14] which is directional message passing network that operates on 3D coordinates. DimeNet++ was designed to predict energies and forces of molecular structures with the intent of speeding up molecular dynamics simulations, but the architecture has also shown promise in predicting molecular targets other than energies. More detail about the architecture as well as performance on benchmark datasets can be found in the original publication. In our DimeReaction architecture, the QM-optimized reactant and product coordinates are each passed through the same DimeNet++ model to create a learned representation of each molecule. The learned representation of the product is then subtracted from that of the product before passing through a dense layer to predict the barrier height. Although it may be interesting to explore other architectures, such as subtracting the learned atom representations before aggregating into a molecular representation, our results from the main text show that model performance is extremely sensitive to the TS geometry, which precluded efforts to further modify the architecture.

A scaffold split on the reactant SMILES is again used to create independent sets. As before, the data is split into 85% training, 5% validation, and and 10% testing, and each pair of forward and reverse reactions is placed in the same set. Unlike with our modified Chemprop model, training on only the coupled cluster data yields lower testing errors than using all three levels in a transfer learning approach. We use the Noam learning rate scheduler[S13] during training and use Optuna for hyperparameter search;[S15] the final hyperparameter values are shown in Table S7.

Table S7: Optimized hyperparameters for the DimeReaction model.

Hyperparameter	Value
Epochs	100
Warm-up epochs	4
Batch size	32
Learning rate	10^{-3}
Learning rate scheduler	Noam
Hidden channels	100
Output embedding channels	100
Output channels	100
Interaction embedding size	64
Basis embedding size	8
Blocks	6
Spherical harmonics	6
Radial basis functions	6
Output layers	2
Layers (FFN)	3
Activation (FFN)	SiLU

Using the optimal hyperparameters for network architecture and training procedure, a few combinations of inputs and training targets are tested. Two types of inputs are explored, either the optimized reactant and product geometries or the optimized reactant and TS geometries. The first strategy represents a realistic prediction strategy since reactants and products are much easier to identify and optimize than a TS. The second strategy better aligns with chemical intuition, since reaction barriers correspond to the difference in energies of the TS and the reactant. The dense layer could optionally receive the enthalpy as an additional input that is concatenated to the difference of the learned molecular representations. The regression target(s) are either just the barrier height or co-training the model on both the barrier height and enthalpy.

The results from training on the coupled cluster data for these combinations are summarized in Table S8. For the case when we use reactant and product geometries, using reaction enthalpy marginally improves the testing errors. However, using a 2D network, such as our modified Chemprop, yields far better performance, which is quite surprising. For the case

when we use reactant and TS geometries, testing errors improve over the 2D Chemprop formulation that had used the CGR of the reactant and product graphs. Although it is satisfying to see the TS encode more information about the reaction's barrier height than the product does, the optimized TS is better used with canonical transition state theory to calculate the rate constant.

An interesting conclusion comes from row 2 and row 4 in Table S8, both of which train a multi-task model to predict E_a and dH. When using the optimized reactant and product geometries as input (row 2), the model gives very good predictions for reaction enthalpy with a test MAE of 1.78 kcal mol^{-1}. On the contrary, the model trained with optimized reactants and TSs (row 4) results in a MAE of 5.95 kcal mol^{-1} for dH, which is quite poor. These results indicate that the DimeNet++ model functions as intended; it was designed as a neural network potential. When the output property is a function of the input geometries (i.e. reactants and products for reaction enthalpy or reactants and TSs for barrier height), DimeNet++ works well. However, when the output property is not a function of the inputs, the model performs poorly. This is unfortunate since the more convenient use case would be to predict barrier height from reactant and product structures.

Table S8: DimeReaction testing errors (kcal mol^{-1}) for combinations of model inputs and regression targets.

Optimized Input	dH Input to FFN	Target(s)	E_a MAE	E_a RMSE	dH MAE	dH RMSE
Reactant & Product	No	E_a	6.20	10.10	-	-
Reactant & Product	No	E_a & dH	6.03	9.84	1.78	2.78
Reactant & Product	Yes	E_a	6.12	9.67	-	-
Reactant & TS	No	E_a	2.25	3.91	-	-
Reactant & TS	No	E_a & dH	2.67	4.60	5.95	11.93
Reactant & TS	Yes	E_a	2.27	3.95	-	-

References

(S1) Yang, K.; Swanson, K.; Jin, W.; Coley, C.; Eiden, P.; Gao, H.; Guzman-Perez, A.; Hopper, T.; Kelley, B.; Mathea, M., et al. Analyzing Learned Molecular Representations for Property Prediction. *J. Chem. Inf. Model.* **2019**, *59*, 3370–3388.

(S2) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Model.* **1988**, *28*, 31–36.

(S3) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural Message Passing for Quantum Chemistry. International Conference on Machine Learning. 2017; pp 1263–1272.

(S4) Gasteiger, J.; Yeshwanth, C.; Günnemann, S. Directional Message Passing on Molecular Graphs via Synthetic Coordinates. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 15421–15433.

(S5) Grambow, C. A.; Pattanaik, L.; Green, W. H. Deep Learning of Activation Energies. *J. Phys. Chem. Lett.* **2020**, *11*, 2992–2997.

(S6) Varnek, A.; Fourches, D.; Hoonakker, F.; Solov'ev, V. P. Substructural Fragments: An Universal Language to Encode Reactions, Molecular and Supramolecular Structures. *J. Comput.-Aided Mol. Des.* **2005**, *19*, 693–703.

(S7) Heid, E.; Green, W. H. Machine Learning of Reaction Properties via Learned Representations of the Condensed Graph of Reaction. *J. Chem. Inf. Model.* **2021**, *62*, 2101–2110.

(S8) Landrum, G., et al. RDKit: Open-Source Cheminformatics. **2006**.

(S9) Guan, Y.; Coley, C. W.; Wu, H.; Ranasinghe, D.; Heid, E.; Struble, T. J.; Pattanaik, L.; Green, W. H.; Jensen, K. F. Regio-Selectivity Prediction with a Machine-

Learned Reaction Representation and On-The-Fly Quantum Mechanical Descriptors. *Chem. Sci.* **2021**, *12*, 2198–2208.

(S10) Greenman, K. P.; Green, W. H.; Gomez-Bombarelli, R. Multi-Fidelity Prediction of Molecular Optical Peaks with Deep Learning. *Chem. Sci.* **2022**, *13*, 1152–1162.

(S11) Chung, Y.; Vermeire, F. H.; Wu, H.; Walker, P. J.; Abraham, M. H.; Green, W. H. Group Contribution and Machine Learning Approaches to Predict Abraham Solute Parameters, Solvation Free Energy, and Solvation Enthalpy. *J. Chem. Inf. Model.* **2022**, *62*, 433–446.

(S12) Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39*, 2887–2893.

(S13) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.

(S14) Klicpera, J.; Giri, S.; Margraf, J. T.; Günnemann, S. Fast and Uncertainty-Aware Directional Message Passing for Non-Equilibrium Molecules. *arXiv* **2020**, *2011.14115*, Accessed 2022-03-08.

(S15) Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-Generation Hyperparameter Optimization Framework. Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2019.