

Data-Efficient, Chemistry-Aware Machine Learning Predictions of Diels–Alder Reaction Outcomes

Angus Keto, Taicheng Guo, Morgan Underdue, Thijs Stuyver, Connor W. Coley, Xiangliang Zhang, Elizabeth H. Krenske, and Olaf Wiest*



Cite This: *J. Am. Chem. Soc.* 2024, 146, 16052–16061



Read Online

ACCESS |



Metrics & More

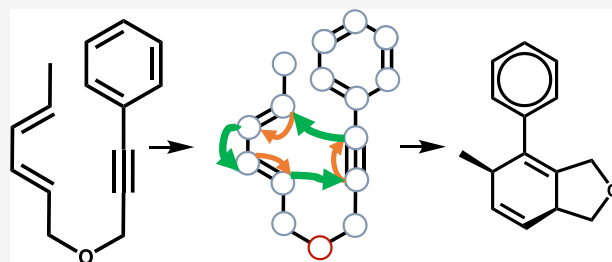


Article Recommendations



Supporting Information

ABSTRACT: The application of machine learning models to the prediction of reaction outcomes currently needs large and/or highly featurized data sets. We show that a chemistry-aware model, NERF, which mimics the bonding changes that occur during reactions, allows for highly accurate predictions of the outcomes of Diels–Alder reactions using a relatively small training set, with no pretraining and no additional features. We establish a diverse data set of 9537 intramolecular, hetero-, aromatic, and inverse electron demand Diels–Alder reactions. This data set is used to train a NERF model, and the performance is compared against state-of-the-art classification and generative machine learning models across low- and high-data regimes, with and without pretraining. The predictive accuracy (regio- and site selectivity in the major product) achieved by NERF exceeds 90% when as little as 40% of the data set is used for training. Another high-performing model, Chemformer, requires a larger training data set (>45%) and pretraining to reach 90% Top-1 accuracy. Accurate predictions of less-represented reaction subclasses, such as those involving heteroatomic or aromatic substrates, require higher percentages of training data. We also show how NERF can use small amounts of additional training data to quickly learn new systems and improve its overall understanding of reactivity. Synthetic chemists stand to benefit as this model can be rapidly expanded and tailored to areas of chemistry corresponding to the low-data regime.



INTRODUCTION

The prediction of reaction outcomes using machine learning (ML) has attracted significant interest due to its prospects for expediting synthetic route design and execution, e.g., for new pharmaceuticals.^{1,2} Although reaction prediction tools have existed since the 1970s in the form of expert-guided rule-based programs,^{3,4} a strong focus on data-driven approaches led to a proliferation of computational and big-data technologies.^{5–12} Reaction prediction by ML is, however, a nontrivial task. It requires careful attention to both model architecture and feature selection. Graph-based approaches^{5,6,8,9,12,13} (which encode chemically interpretable connectivity information) and transformers^{7,10,11} (which apply language translation methods) have both shown promise when developed with large and diverse data sets. However, a current limitation is encountered when making predictions for certain reactions that are underrepresented in the benchmark data sets. The main ML models used for reaction prediction are notoriously data-hungry, requiring large numbers of data points across the chemical or reaction space of interest and/or chemically meaningful features for training and usage. A need for dense data sets that incorporate many reaction types suggests that a model is not highly robust or capable of effective interpolation. This in turn limits the application of ML-assisted synthesis or other techniques to common low-data regimes¹⁴ in chemistry

such as underrepresented reactions and/or substrate classes. When applied to burgeoning areas of chemistry, data-efficient models need fewer experiments to be carried out to develop a sufficient model.

Here, we investigate the question of data efficiency of ML models for the prediction of Diels–Alder reaction products, a task entailing questions of site-, regio-, and diastereoselectivity (Figure 1). First reported in 1928,¹⁵ the Diels–Alder reaction is today a mainstay of the synthetic chemist's toolkit, albeit more so in academia than industry.^{16–18} It enables the simultaneous formation of two new bonds with perfect atom economy and typically high selectivity. One noteworthy application is the construction of polycyclic structures via transannular Diels–Alder reactions of macrocycles, an application of particular significance to the fragrance and flavor industries.¹⁶ From a technical viewpoint, the Diels–Alder reaction embodies a challenging reaction class for machine learning prediction due to the diversity of distinct

Received: March 3, 2024

Revised: May 21, 2024

Accepted: May 22, 2024

Published: June 1, 2024



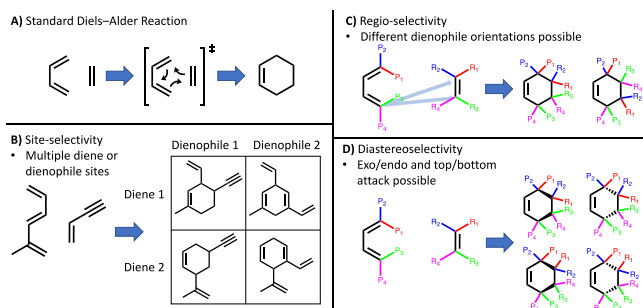


Figure 1. (A) Standard Diels–Alder reaction between butadiene and ethene and the corresponding bond change. (B) Site-selectivity illustrated by reactions between different possible dienes and dienophiles. (C) Regioselectivity shown through different diene and dienophile atom pairings. (D) Diastereoselectivity arising through different combinations of exo/endo and top/bottom attack.

subclasses, e.g., normal electron demand (NED) vs inverse electron demand (IED), aliphatic vs aromatic vs heteroatom-containing substrates, and intermolecular vs intramolecular (including transannular). These technical challenges, coupled with the reaction's widespread value to synthetic chemistry and the associated diversity of substrates, motivated our interest in developing predictive models for the Diels–Alder reaction.

In 2019, Grzybowski and co-workers¹⁹ reported a random forest model for predicting the outcomes of Diels–Alder reactions. Trained on subsets of a data set of 6355 reactions—all of which were intermolecular reactions—the model predicted the regiochemistry, site-selectivity, and diastereoselectivity (our definition of diastereoselectivity in Figure 1 may differ) of intermolecular Diels–Alder reactions (major products) with accuracies of 93.6, 91.3, and 89.2%, respectively. The best-performing feature set incorporated molecule fingerprints and substituent constants (Hammett constants and topological steric indices²⁰). They showed that these features provided superior interpolation accuracy compared to quantum mechanical (QM)-derived features (e.g., Parr functions). Other models including a neural network (NN) classifier²¹ and a graph-based NN⁶ gave lower accuracies. It was suggested that the use of chemically relevant descriptors representing reaction center and substituent effects was important for accurately predicting reaction outcomes. More recently, Alexandrova and co-workers²² developed a predictive model for Diels–Alder reactions that used topological descriptors of charge density to predict barrier heights in solution, while Grayson and co-workers²³ applied ML and transfer learning techniques to improve semiempirical QM-derived Diels–Alder barrier heights.

We report the first examples of ML models that display useful levels of predictive accuracy across all of the major subclasses of Diels–Alder reactions. We demonstrate that generative ML architectures display superior performance for predicting reaction outcomes, such as regio- and site-selectivities with accuracies of >95%. Notably, the best-performing model architecture learns the chemical principles of the reaction, namely the cyclic movement of electrons, as opposed to template models using curated features in analogy to the physics-aware models²⁴ encoding the underlying physics of the reactions described. The best performance is achieved solely based on connectivity information rather than on specific chemical descriptors. High predictive accuracy is achieved by two different models: a generative graph-based

model, NERF, that makes predictions based on chemically interpretable edge/bond changes without pretraining, and a pretrained transformer model, Chemformer,¹¹ based on natural language processing of SMILES strings. Both models give high levels of predictive accuracy across diverse reaction classes, viz. normal/inverse-electron demand, aliphatic/aromatic/hetero, and intermolecular/intramolecular/transannular. The latter case is particularly noteworthy: we achieve the first successful predictions (accuracies >90%) of intramolecular reactions, indicating the development of models that can correctly describe these geometrically more challenging reactions. In contrast to Chemformer, new chemistry can be included in NERF with minimal training; therefore representing an ideal platform for future incorporation into retrosynthesis programs.

METHODS

Data Set Curation. Experimentally reported Diels–Alder reactions were extracted from Reaxys,²⁵ which provided substance, reaction, and property data sourced from patents and literature articles spanning the full-time frame since the discovery of the Diels–Alder reaction. Through the use of Diels–Alder-related search terms, we extracted an initial 37,891-entry data set (April 6, 2021) containing 27,349 unique reactions. A filtering process shown in Figure 2a was then implemented to produce a working data set of

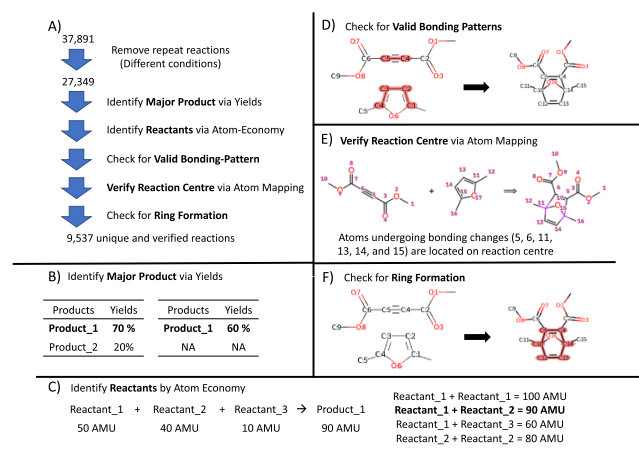


Figure 2. (A) Workflow used for generating the working data set of 9537 Diels–Alder reactions. (B) Identification of major product via yield. (C) Identification of reactants accounting for atom economy (sum of reactant masses must equal the product mass). (D) Detection of Diels–Alder bonding patterns. (E) Verification that atoms undergoing bond changes are located on the reaction center. The reaction center is defined here as the reacting diene and dienophile atoms. (F) Detection of the formation of a new six-membered ring.

9537 reactions. The sequence of filtration steps involved (i) determining reaction SMILES by identifying the major products and taking into account the atom economy of the Diels–Alder reaction (Figure 2b,c), (ii) checking for the presence of correct cycloadduct bonding patterns through substructure searches (Figure 2d), (iii) performing reaction mapping with RXNMapper²⁶ to atom-map the reaction SMILES and check that the atoms undergoing bonding changes were located on a Diels–Alder reaction center (Figure 2e), and (iv) eliminating tautomerizations and other non-Diels–Alder reactions by confirming the formation of a six-membered ring (Figure 2f).

Density Functional Theory (DFT) Calculations. A range of molecular features were generated for the reactants in our data set using RDKit,²⁷ XTB,^{28,29} and Gaussian 16.³⁰ To create starting geometries for DFT optimizations, 3D coordinates were generated in RDKit using the MMFF94³¹[8] force field and the ETKDGV2^{32,33}

conformational search algorithm. Geometry optimizations and single-point calculations were then performed in Gaussian 16 using M06-2X/6-311+G(d,p)//B3LYP-D3(BJ)/6-31G(d).^{34–38} This DFT method was chosen to provide a cost-effective treatment of dispersion effects across chemically diverse systems. Because the experimental data from which the data set was derived were obtained in a wide range of solvents and the Diels–Alder reaction generally shows a low dependence on solvents, all DFT calculations were performed in the gas phase for internal consistency. This automated approach provided DFT-optimized structures for 7171 of the 9537 reactions, herein referred to as the DFT subset. Its purpose was to probe the effects of DFT-calculated features, such as charges and atomic contributions to molecular orbitals, on model performance.

Fingerprinting Methodology. To support an analysis of the diversity of the data set, an adaptation of a fingerprinting method originally reported by Hu and co-workers³⁹ was applied. Molecule-wide descriptors were removed and the atom and bond matrices were restricted to the reaction center and neighboring atoms only (Figure 3). Atoms and bonds in the reaction center region were characterized

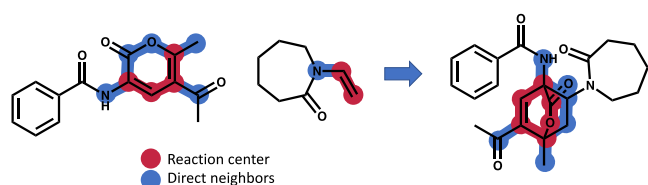


Figure 3. Visualization of reaction center and direct neighbors upon which atom and bond features were extracted and used for clustering.

by various factors such as formal charge, hybridization, and bond type using one-hot encoding. Ring membership and size were added as additional atomic features. A sequential combination of principal component analysis (PCA) reductions and variational auto encoding (VAE) was then applied to detect patterns in the data.

Machine Learning Architectures. We examined the performance of a range of machine learning model architectures:

- (1) A template-based graph neural network (GNN) as used by Stuyver et al.^{40,41} which had shown strong predictive performance for substitution reactions.
- (2) Other template-based baseline models⁴² including support vector machines, multilayer perceptron, linear regression, Bayesian ridge, kernel ridge, and random forest.
- (3) The generative natural language processing (NLP) models, Chemformer¹¹ and TSCHEM,¹⁰ are examples of chemical applications of NLP⁴³ and are both capable of SMILES2SMILES translation and product prediction. In both cases, a pretrained model trained on subsets of the USPTO⁴⁴ data set was fine-tuned on our data set.
- (4) The generative graph-based model NERF¹² (Nonautoregressive Electron Redistribution Framework). Molecular graphs, which are reminiscent of 2D representations of molecules, represent atoms by nodes and bonds by edges. The network allows for message passing between nodes, leading to atoms learning about their surrounding environment. See Friederich and co-workers⁴⁵ for a detailed review on GNNs, in which molecular graphs can be applied, in chemistry. To the best of our knowledge, NERF has not previously been applied to predicting products for specific reaction classes; it has only been applied to benchmark predictions using USPTO data. In NERF, nodes are featurized by atom type, charge, aromaticity, segment embedding, and positional embedding. Multiple-order bonds are designated by multiple edge values. Using this graph, NERF assesses the likelihood of interaction between each node and every other node for edge formation and edge breaking before summing up and rounding these values to determine the change in edge value (Figure 4). Changes in edges thus resemble changes in bonds and the associated electron flow, even though it is important to note that this is a result of the

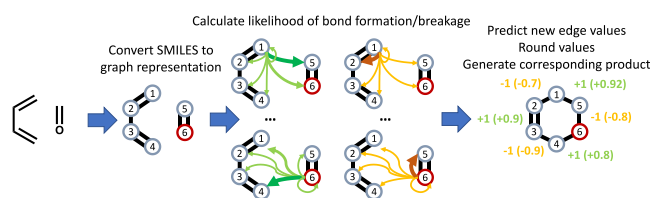


Figure 4. Simplified visualization of product prediction from SMILES by the NERF model.

equivalency of edges and covalent bonds used in molecular graphs. We anticipated that this property would be appropriate for Diels–Alder reactions, which involve a shuffling of bonds around a cyclic array.

Data Set Splits. To account for the effects of randomness in the data set splits, 10 random splits of the data set were used to train 10 models for each experimental run of each model architecture. Performances are reported as the average $\pm$ standard deviation (SD) of the 10 models. This approach was used for all model types except for the GNN from Stuyver and Coley,⁴⁰ where each fold from the 10-fold cross-validation was used to calculate standard deviations.

RESULTS AND DISCUSSION

Data Set Diversity. As the basis for our predictive model building, we constructed a data set of 9537 Diels–Alder reactions listed by their Reaxys IDs in the [Supporting Information](#). The diversity of the data set—and the applicability of any models trained on it—is determined by various metrics including selectivity and substrate type shown in Figure 5. Illustrating its diversity, our data set includes 5226

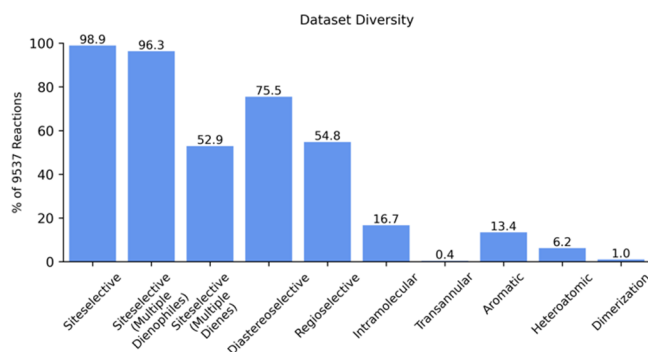


Figure 5. Percentage distribution of reaction subclasses contained in the data set.

regioselective reactions, 9436 site-selective reactions, and 7200 diastereoselective reactions (see Figure 1 for definitions). With regards to site selectivity, 5044 reactions had multiple possible dienes while 9187 had multiple possible dienophiles (multiple double, aromatic, and triple bonds found on the dienophile). Our detection of site selectivity does not entail any judgment on chemical reasonableness but rather on whether another outcome is theoretically possible. The ratio of intermolecular to intramolecular reactions in the full data set was 5:1. Among the intramolecular reactions, 2.4% (38) were transannular. Aromatic and hetero-Diels–Alder reactions comprised 13.4 and 6.2%, respectively, of the data set. The normal electron demand (NED) to inverse electron demand (IED) ratio was 8.2:1, as calculated by comparing HOMO/LUMO energies for 5576 intermolecular reactions from the DFT subset. Details about substructures are available in the [Supporting Information](#).

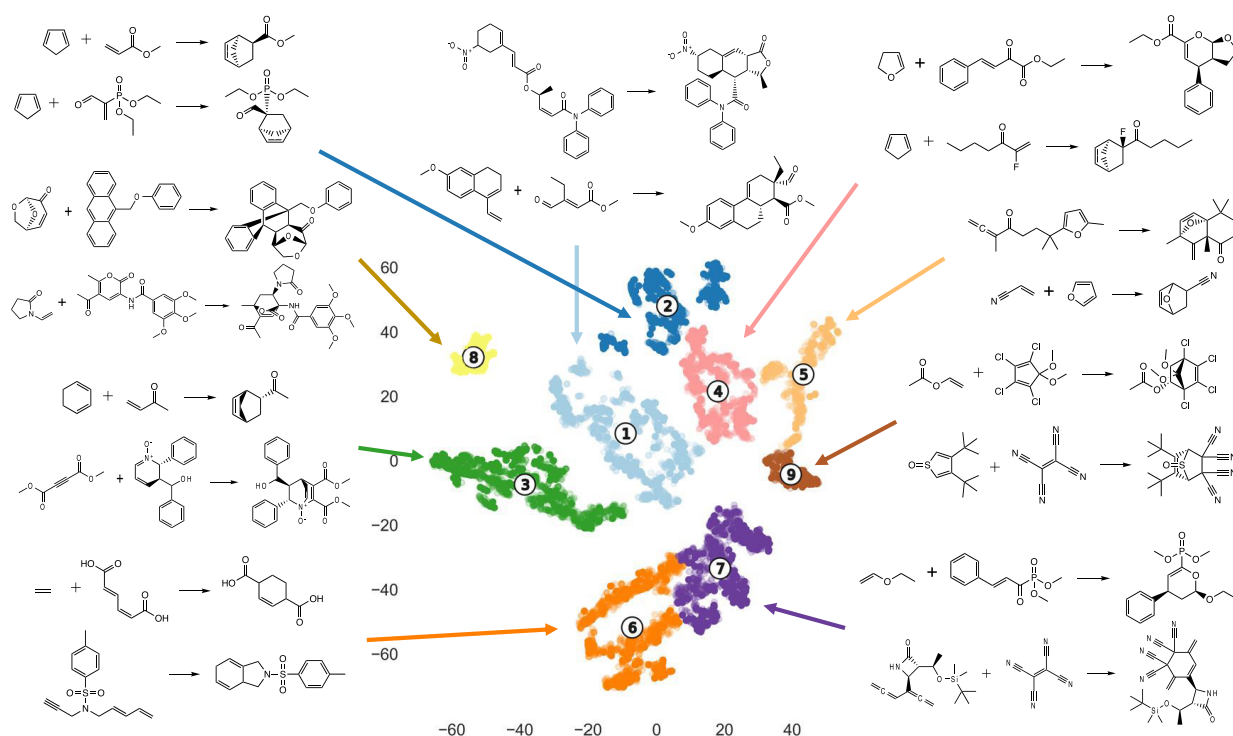


Figure 6. Diversity of the data set, as visualized using T-SNE dimensionality reduction and agglomerative clustering. Features were derived from the reaction center and immediate neighboring atoms. Different colors demarcate different clusters. Example reactions from each cluster (cluster numbering is shown in the central panel) are provided.

To further dissect reaction subclasses, we applied a clustering method adapted from Hu and co-workers³⁹ which analyzed the atom and bond information around the reaction center. These clusters, visualized using T-SNE dimensionality reduction, are shown in Figure 6 along with selected representative reactions. The clustering grouped reactions together into distinct, chemically meaningful sets. For example, two clusters (5 and 8) contained predominantly aromatic reaction centers (93.7 and 100%, respectively) but differed with respect to whether the center was contained in a 5- or 6-membered ring. Another cluster (7) contained twice the proportion of hetero-Diels–Alder reactions (13.6%) compared to the full data set. Further evidence of the chemically distinct character of the clusters was provided by the challenge that each one posed when it was used as the testing set for our ML models (see SI for more details).

Product Prediction. A requirement for any model is that it must have high accuracy of predicting the experimentally observed product as the Top-1 ranked choice. We used a target of >90% Top-1 accuracy across the wide diversity of Diels–Alder chemistry in our data set. We investigated a variety of model architectures with different molecule representations and capabilities. Template-based approaches were explored as they are conceptually simple while generative models were used on account of their straightforward input and versatility. Herein we explore how NERF on account of its chemistry-aware design, is able to learn the Diels–Alder reaction, as exemplified by its performance across different data set splits, substrate classes, and extrapolation tasks.

Template-Based Models. Template-based models were used as a starting point because they are simply a classification task between products, in our case generated using RDChiral⁴⁶ and templates allowing for any combination of carbon, nitrogen, oxygen, and sulfur atoms, even if located in aromatic

rings. These template-based models can handle site- and regioselectivity but not diastereoselectivity. These selectivity variables, along with symmetry, and the need to maintain consistent data set sizes with models using reactant DFT features (7171 reactions) reduced the number of distinct templates. As a result, our primary data set of 6198 reactions was a subset of the full 9537-reaction data set.

The two best models were a random forest (RF) model with Gasteiger charges^{27,47} features for the six reaction center atoms and the template-based GNN. For the former, Gasteiger charges proved more effective descriptors compared to the HOMO/LUMO contributions with respective accuracies of 85.3 and 39.5%. In comparison, the Diels–Alder trained GNN model achieved an accuracy of 81.1% for the same set of reactions and templates. Reducing the number of products to choose from by restricting atom type and aromaticity in the templates reduced the amount of training data and subsequent performance; a 4621-point carbon-only template data set had an accuracy of just 77.8% (see SI). In contrast, for the 6198 reaction data set, the generative models Chemformer and NERF had respective Top-1 accuracies of 92.1% and 96.0%. As these template models could not exceed 90% accuracy, they were therefore not considered further in this study.

Generative Models. An advantage of generative models over template-based methods is that in cases where there are multiple possibilities of dienes and dienophiles (e.g., in aromatic systems), the different possibilities do not have to be enumerated in multiple templates.^{48,49} Template-based methods are constrained by the specific reactions and mechanisms they are designed to identify, making it impractical to comprehensively outline every possible reaction. In contrast, generative models are capable of discerning intricate patterns directly from data, without the need for predefined templates. This capability enables them to general-

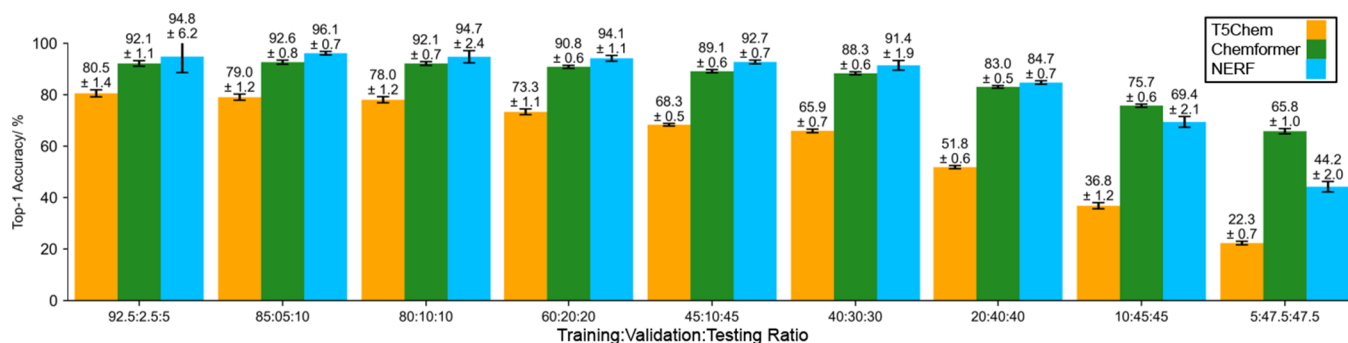


Figure 7. Top-1 accuracies of T5Chem, Chemformer, and NERF for predicting the regio- and site-selectivity of Diels–Alder reactions across data set splits of decreasing training set size. Error bars represent the standard deviation over 10 replicates.

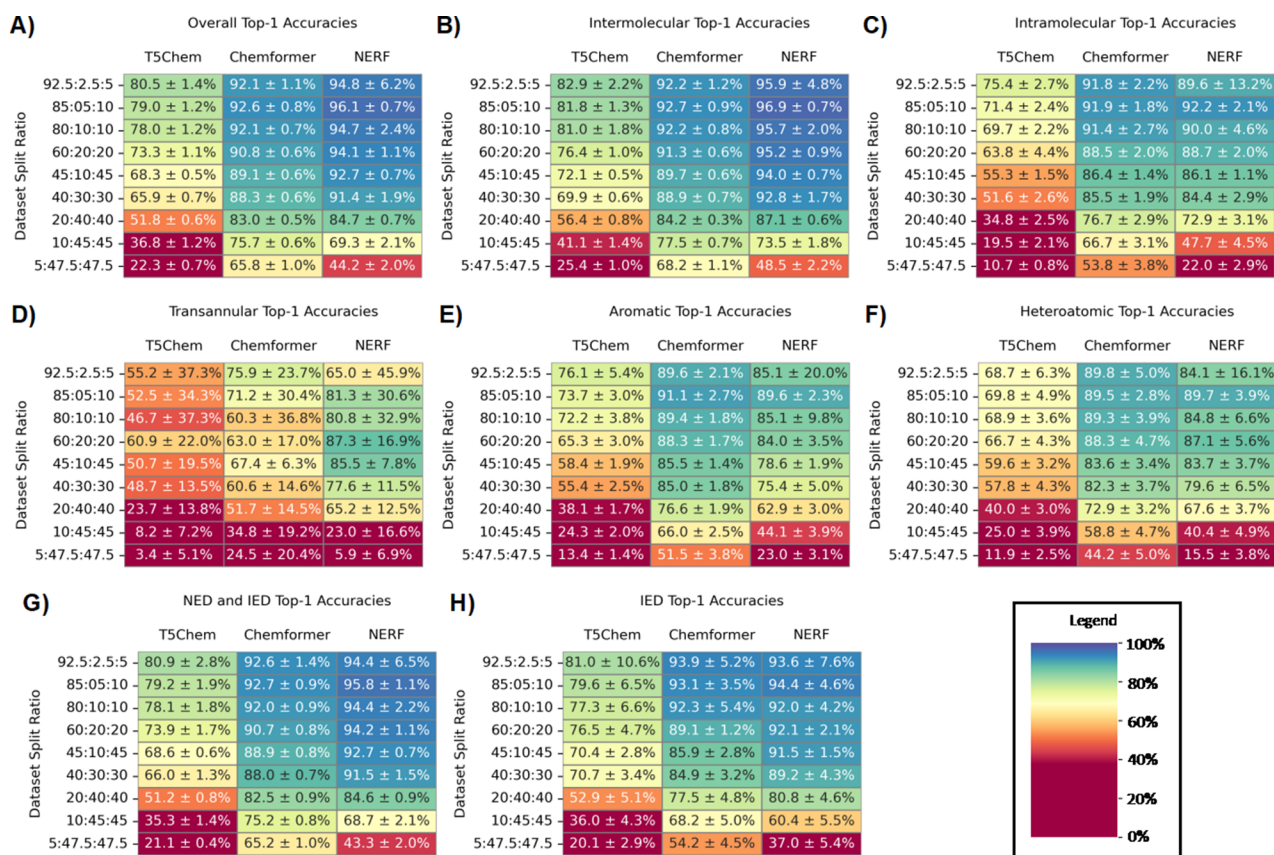


Figure 8. Accuracies of T5Chem, Chemformer, and NERF for predicting the regio- and site-selectivity of Diels–Alder reactions across data set splits of decreasing training set size. The accuracies for the overall (A), intermolecular (B), intramolecular (C), transannular (D), aromatic (E), heteroatomic (F), NED and IED (G), and IED (H) substrate subclasses are shown. Values are shown as average performance across 10 replicates ± the standard deviation.

ize to new, unseen data and to accurately predict reactions across a broader range of conditions.

We examined two types of generative models: transformers and graph-based models. One potential advantage of transformers is their ease of application due to the minimal input data needed. Here, we investigated two popular transformers, Chemformer¹¹ and T5Chem.¹⁰ In contrast, graph-based models potentially offer advantages associated with their ability to encode connectivity and atomic features. Here, we examined the graph-based model NERF which predicts simultaneous edge changes that occur on going from reactants to products.

We investigated how the predictive accuracies of the generative models are affected by the size of the training data set. Models were built using splits of 5:47.5:47.5 to 92.5:2.5:5 (training:validation:testing) from the full data set. We first considered the accuracy of predictions of regio- and site-selectivity, ignoring any stereochemical features of the reaction. The performance of the generative models across the range of data regimes is shown in Figure 7, with Top-1 accuracy referring to when the most confident prediction is correct and Top-5 accuracy (Table S2) referring to when one of the top 5 most confident predictions is correct.

With a typical split of 80% training data, T5Chem does not exceed 90% accuracy while Chemformer and NERF do with

respective Top-1 accuracies of 78.0, 92.1, and 94.7%, respectively. Using a larger split of 85% training data, the Top-1 accuracies increase to 79.0, 92.6, and 96.1%. NERF can even achieve >95% accuracy while T5Chem was unable to achieve >90% accuracy using any of our splits. At the highest split (92.5% training data), the performance of Chemformer and NERF decreases slightly but not significantly, a consequence of the random composition of the testing set in each experiment.

We then stepwise reduced the amount of training data to 40% of the data set to test the robustness of the models as shown in Figure 7. The respective accuracies for T5Chem, Chemformer, and NERF of 65.9, 88.3, and 91.4%. NERF exceeded the 90% threshold when as little as 40% of the data set was used for training. Similarly, Chemformer required >45% to achieve comparable accuracy but T5Chem never reaches it. While NERF performs slightly better than Chemformer for predicting Top-1 accuracy at the 40% split, the two methods have comparable performance when judged according to Top-5 accuracy (92.3 vs 92.6% for Chemformer and NERF respectively). In the lowest training data regime, where 5% of the data set was used for training, the Top-1 accuracies provided by T5Chem, Chemformer, and NERF were 22.3, 65.8, and 44.2% (Figure 7). Chemformer has a clear advantage here and with the other low-data set split of 10% training data.

We sought to understand the impact of pretraining for the best-performing model in this low-data regime, Chemformer. The starting checkpoint for Chemformer as used in Figure 7 was pretrained¹¹ on the task of reaction product prediction using the USPTO-MIT data set with roughly 400,000 training reactions that do not contain stereochemical information and according to previous examination,⁵⁰ seven Diels–Alder reactions. If no fine-tuning is performed, the Top-1 prediction accuracy (on the testing set of the 5:47.5:47.5 split) decreases to 3.3%. We then built models without a checkpoint (i.e., not pretrained) and trained them on Diels–Alder reactions. The data set splits tested ranged from 40% - 92.5% of the data set as training and the corresponding accuracies ranged from 11.2% to 65.1%. We conclude that while the pretrained checkpoint offers no direct benefit to predicting Diels–Alder chemistry, it does provide an understanding that is necessary to achieve a Top-1 accuracy of >90%. Understanding the syntax of SMILES and the preservation of structure through a reaction are key aspects that could be learned from pretraining.

We next explored the ability of the models to predict diastereoselectivity (where present) in addition to regio- and site-selectivity. These kinds of predictions are possible with the transformer models but not with the current version of NERF, which encodes connectivity but not stereochemistry. Even in the highest training regime of 92.5% training data, the Top-1 accuracies of T5Chem and Chemformer were 47.0 and 43.2%, respectively, far from the target threshold of 90% (see SI). Prediction of stereoselectivity is nontrivial and would likely require knowledge of steric interactions that are not directly encoded in SMILES strings.

To probe the general applicability of the three generative models across different Diels–Alder reaction subclasses, we examined their respective performance in predictions of intramolecular, hetero-, aromatic, and inverse electron demand (IED) Diels–Alder reactions. This was done by filtering the predictions obtained on the complete data set and shown in Figure 7 to these respective categories. The results shown in

Figure 8 indicate that intramolecular reactions (16.7% of the data set) present an especially challenging test for the models as they entail steric and geometric constraints associated with forming viable cyclic structures and were consequently excluded in the study of Beker et al.¹⁹ When an 80:10:10 split is used, the Top-1 accuracies are 81.0%, 92.2%, and 95.7% for intermolecular reactions while those for intramolecular reactions are 0.8–11.3% lower at 69.7%, 91.4%, and 90.0%. This is the first time that machine learning models achieved >90% Top-1 accuracies for intramolecular Diels–Alder reactions. For transannular reactions (0.4% of the data set), T5Chem and Chemformer perform best with a 92.5:2.5:5 split, resulting in corresponding accuracies of 55.2 and 75.9%, while NERF's best result of 87.3% is with a more balanced 60:20:20 split; none however exceed the 90% threshold. The decrease in performance for NERF at higher data set splits is likely statistical errors, resulting from limited testing examples (a 5% testing set has 2 testing examples on average). The token positioning in SMILES strings or connectivity of the reactant graphs may suffice for understanding the steric environment of simple intramolecular reactions but presumably not for complex transannular cases.

Aromatic reaction centers also present a challenge as the model must differentiate between multiple reactive sites that have different impacts on conjugation and molecular stability. Within the data set, 63.4% (6047) of the reactions involve substrates that contain aromatic rings, but only 13.4% of the data set involve direct participation of the aromatic ring as the diene or dienophile. When compared against the overall accuracies for the three methods (78.0, 92.1, and 94.7% using an 80:10:10 split), the corresponding accuracies for reactions involving aromatic reaction centers were 2.7–9.6% lower, at 72.2, 89.4, and 85.1%. If the slightly higher split of 85:5:10 is used, Chemformer can achieve >90% accuracy (91.1%) on these same reaction centers while NERF comes close at 89.6%. Analysis of the failed NERF predictions showed that reactions involving aromatic reaction centers had significantly higher rates of invalid double bond placement, ring formation at the wrong sites, partial reaction of the diene, or no reaction at all compared to nonaromatic substrates. However, the failed predictions, aromatic or not, were unlikely (1–5.5% probability) to produce a valid Diels–Alder product. Nonetheless, the failed predictions mirror the problems many human chemists would encounter in predicting the correct regioselectivity of this type of reaction.

Hetero-Diels–Alder reactions (e.g., those involving oxygen, nitrogen, phosphorus, or sulfur atoms as part of the reaction center) proved difficult for all models. Using an 80:10:10 split, the T5Chem, Chemformer, and NERF models had respective overall Top-1 accuracies of 78.0, 92.1, and 94.7%, while the values for hetero-Diels–Alder reactions were 2.8–9.9% lower, at 68.9, 89.3, and 84.8%, respectively. This drop in performance is comparable to that observed for aromatic reaction centers but on account of their lower representation in the data set (6.2% hetero- as compared to 13.4% aromatic), hetero-Diels–Alder reactions perhaps do not represent as fundamental a shift in reactivity. This reemphasizes again the need for a sufficient number of training examples for accurate predictions.

Finally, we examined the performance of the models for predicting the products of normal versus inverse electron demand Diels–Alder (IED) reactions using the subset of 5576 intermolecular reactions (vide supra). For NED and IED reactions considered together, the Top-1 accuracies for the

80:10:10 split were 78.1, 92.0, and 94.4% for T5Chem, Chemformer, and NERF respectively. In comparison for IED reactions, the Top-1 accuracies were only 0.3–2.4% lower at 77.3, 92.3, and 92.0%. Although NED and IED reactions represent distinctly different reaction subclasses from a chemical perspective in that the molecular orbitals involved are reversed, most models are capable of accurate predictions of both since they do not make explicit use of HOMO and LUMO. We also considered whether reactions with major and minor products with similar reactivity, as determined by yield differences, were more difficult. However, the composition of these reactions and their small data subset size made a conclusive determination difficult (see [Supporting Information](#)).

To further probe the accuracy of NERF for predictions of the most challenging reaction motifs, we examined the accuracies achieved when isolating all intramolecular, aromatic, and hetero-Diels–Alder reactions to the test set (see [SI](#)). The Top-1 performances for testing on intramolecular, aromatic, and hetero-Diels–Alder reactions were 9.6, 0, and 29.8% respectively. The significant drop in performance seen in this challenge highlights the difficulties associated with generalizing from one subclass of Diels–Alder reaction to another and the importance of data set diversity and breadth. The 0% accuracy for aromatic reactions is particularly noteworthy: if never exposed to aromatic reaction centers during training, the model does not learn the fundamental electronic reasons why an aromatic ring would undergo a Diels–Alder reaction.

It is noteworthy that high accuracy can be achieved by NERF without pretraining whereas Chemformer had to be extensively pretrained. Furthermore, NERF can achieve this accuracy for Diels–Alder reactions at lower training splits. Although Chemformer provides predictions of diastereoselectivity, which NERF does not consider, the Top-1 accuracy is below 50% and thus not practically useful. When examining the regio- and site-selective performance across substrate classes, Chemformer was the most consistent (overall vs substrate class accuracy) but NERF was able to achieve higher accuracies in most cases. We hypothesize that the chemistry-aware aspects of the NERF model, namely the cyclic change in edges calculated that is common to all Diels–Alder reactions, allow the model to transfer learning between different reaction subclasses and correctly predict the products using small training samples. Future work will have to elucidate if this also holds true for other, similar pericyclic reactions.

Features. We investigated the hypothesis of chemistry-aware learning of NERF further by testing if the predictive accuracy could be improved by incorporating features related to the Woodward–Hoffmann rules and orbital aspects of the Diels–Alder reaction.^{51–53} The NERF model is based on the likelihood of interactions between atom pairs. We introduced electronic features, such as charge, nucleophilicity/electrophilicity, and atomic contributions to molecular orbitals describing the node properties of HOMO and LUMO, to mimic a chemist's understanding of atoms and possible reaction centers. However, we found for the DFT-featurized subset of 7171 reactions, with a 60:20:20 and 10:45:45 split, that all features led to accuracies that were lower and/or within one standard deviation of the default NERF model performance (see [Supporting Information](#)). Simplifications, either one-hot encoding (1/0 values) or “phase” encoding (−1/0/1 values) of MO contributions and electronic charges slightly improve performance over the original continuous value. This

suggests that the NERF model does not prioritize curated features over learned features in its architecture, perhaps because the curated features may be too noisy, inferior, or already present in the model's latent understanding.

Extrapolation. Finally, we examined the ability of the models to predict reactions from outside the data set. As an initial test case, we investigated the Diels–Alder reactions of 1,2,4-triazenes.⁵⁴ Reactions of triazenes were not present in our original data set because their cycloadditions are followed by spontaneous extrusion of N₂ and are thus filtered out during the atom-economy filtering step. Chemformer and NERF models were tested on a set of 17 Diels–Alder reactions involving 1,2,4-triazenes using models trained on 80:10:10 splits ([Figure 9a](#), see [SI](#) for other splits).

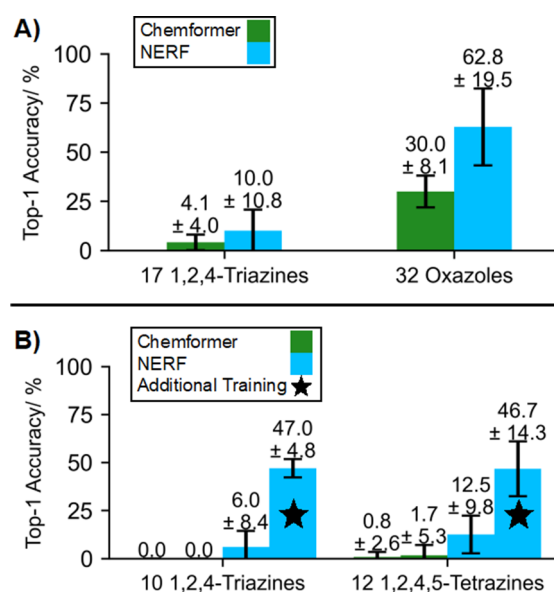


Figure 9. (A) Top-1 accuracies for Chemformer and NERF models trained on an 80:10:10 split and tested on 17 1,2,4-triazines and 32 oxazoles. (B) Top-1 accuracies for Chemformer and NERF models trained on an 80:10:10 split with and without 17 1,2,4-triazines and 32 oxazoles as additional training data. Testing is performed on 10 1,2,4-triazines and 12 1,2,4,5 tetrazines.

Overall, the predictive accuracies for the triazene Diels–Alder reactions were low. NERF often failed to identify the correct reaction center, while Chemformer led to incorrect bonding, involved another reaction, or simply had no Top-1 prediction (see [Supporting Information](#)). Given that these reactions involve simultaneously aromatic, IED, and hetero-Diels–Alder reaction centers, systems that are present in the training data set, the low predictive accuracy is perhaps not surprising.

Another set of 32 reactions was used as a challenge for the models, this time involving intermolecular reactions of oxazoles and the elimination of the bridging oxygen as water proceeding the Diels–Alder reaction.⁵⁴ Oxazole motifs appear in 22 reactions within our data set and are directly involved in the reaction center in 15 of these, but in each case reacting in an intramolecular manner. The 32 test reactions therefore provide a molecularly-based extrapolation challenge. While the Top-1 accuracies for Chemformer and NERF of 30.0 and 62.8% respectively ([Figure 9a](#)) are far from the threshold of 90%, NERF appears to better utilize these limited training examples and extrapolate.

Having demonstrated the limitations associated with extrapolation to unseen reactions, we retrained NERF with these 49 triazene and oxazole reactions included in the training set. The revised model's performance was then evaluated on a set of 22 additional reactions^{54,55} between alkenes or alkynes and 1,2,4-triazines (10 examples) or 1,2,4,5-tetrazines (12 examples). As opposed to the previous 49 examples, the triazine reactions differ by having alkyne dienophiles. The Top-1 accuracy of NERF improved from 6.0 to 47.0% for triazines and from 12.5 to 46.7% for tetrazines, corresponding to respective improvements of 41.0 and 34.2% (Figure 9b). In contrast, Chemformer shows no improvement for 1,2,4-triazines (0–0%) and negligible improvement for 1,2,4,5-tetrazines (0.8–1.7%). This shows that in agreement with the hypothesis above, NERF can effectively leverage small amounts of training data to get significant improvements in accuracy on related systems (triazines with alkynes) and even unseen reaction centers (tetrazines). The NERF model's performance also increased on the original test set from 94.7 to 95.9%, because of improved recognition of aromatic (85.1–89.6%), heteroatomic (84.8–88.5%), IED (92.0–93.5%), and intramolecular (89.7–92.8%) reaction centers. The improvement in the latter is likely a byproduct of the improvement in the other areas. Strategies including more data mining, data augmentation, and/or transfer learning, will be needed to reach the performance target of 90% for triazines and tetrazines but the data efficiency of NERF suggests this is achievable in the small-data regime commonly encountered in chemistry.

CONCLUSIONS

By using a chemistry-aware model, NERF, it is possible to accurately predict the site- and regio-outcomes of Diels–Alder reactions across a variety of subclasses, including hetero-, aromatic, and intramolecular reactions, in a data-efficient manner and with greater performance over state-of-the-art pretrained models. The changes in edges/bonds predicted by NERF are analogous to the cyclic reorganization of electrons within a Diels–Alder transition state and can be learned simply from the connectivity implied in SMILES strings. Neither simple features from RDKit nor complex DFT features representing key chemical aspects of the reaction are needed.

Using only 40% of the data set, NERF achieved a Top-1 accuracy of 91.4% while the next best model, the pretrained Chemformer, scored 88.3%. It is also possible to further increase the accuracy of NERF to 96.1% accuracy by using a larger training:validation:testing split of 85:5:10. NERF is data-efficient by needing smaller training sets, no pretraining, and no additional features, aspects that make it exceptionally adaptable and deployable to new reactions.

This work also represents the first successful development of highly accurate models for intramolecular Diels–Alder reactions, an inherently more difficult task because of the geometric and distortion effects of intramolecular reactions. Chemformer and NERF both achieved $\geq 90\%$ Top-1 accuracy for intramolecular reactions when allocating at least 80% of the data to training. Similarly, hetero- and aromatic Diels–Alder reactions were also predicted with accuracies close to or exceeding 90%.

NERF was also able to effectively learn from small additions of data (49 1,2,4-triazines and oxazoles) to increase performance on the original testing set as well as extrapolate to related reactions. In comparison, Chemformer was not able to effectively learn from this additional data using our fine-tuning

strategy. The effectiveness and data efficiency of NERF make it highly suited for both high-data high-throughput screening and low-data synthetic lab environments, whereby new experimental results can be rapidly incorporated. Chemical principles should be considered in the design of future machine-learning model architectures.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/jacs.4c03131>.

Common structures, diastereoselectivity, inverse electron demand (IED) reactions, template models, generative models: overall accuracies, generative models: accuracies for diastereoselectivity predictions, Chemformer pretraining effects, reactant subclass testing sets, major minor product reactivity difficulty, performance of models in different reaction clusters, features, aromatic vs nonaromatic systems, extrapolation, link to github, Reaxys reaction IDs, and link to Jupyter notebook to regenerate data sets (PDF)

AUTHOR INFORMATION

Corresponding Author

Olaf Wiest – Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556, United States; orcid.org/0000-0001-9316-7720; Email: owiest@nd.edu

Authors

Angus Keto – School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia

Taicheng Guo – Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, Indiana 46556, United States

Morgan Underdue – Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556, United States

Thijs Stuyver – Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States; Present Address: Ecole Nationale Supérieure de Chimie de Paris, Université PSL, CNRS, Institute of Chemistry for Life and Health Sciences, 75005 Paris, France; orcid.org/0000-0002-8322-0572

Connor W. Coley – Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States; orcid.org/0000-0002-8271-8723

Xiangliang Zhang – Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, Indiana 46556, United States

Elizabeth H. Krenke – School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/jacs.4c03131>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the United States National Science Foundation under the NSF Center for Computer-Assisted Synthesis (C-CAS), grant number CHE-2202693, and by the Australian Research Council (DP180103047). A.K. acknowledges the support of an Australian Government Research Training Program Scholarship and an American Australian Association Graduate Education Fund Scholarship. Computer resources were provided by the Notre Dame Center for Research Computing, the Australian National Computational Infrastructure, and the University of Queensland Research Computing Centre. Additional supplemental reaction IDs used by Grzybowski et al.¹⁹ were kindly provided by the Grzybowski group (April 23, 2021).

■ REFERENCES

- (1) DiMasi, J. A.; Grabowski, H. G.; Hansen, R. W. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J. Health Econ.* **2016**, *47*, 20–33.
- (2) Wouters, O. J.; McKee, M.; Luyten, J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009–2018. *J. Am. Med. Assoc.* **2020**, *323*, 844–853.
- (3) Pensak, D. A.; Corey, E. J. LHASA-Logic and Heuristics Applied to Synthetic Analysis. In *Computer-Assisted Organic Synthesis*, ACS Symposium Series; ACS: 1977; Vol. 61, pp 1–32.
- (4) Wipke, W. T.; Braun, H.; Smith, G.; Choplin, F.; Sieber, W., SECS-Simulation and Evaluation of Chemical Synthesis: Strategy and Planning. In *Computer-Assisted Organic Synthesis*, ACS Symposium Series; 1977; Vol. 61, pp 97–127.
- (5) Kayala, M. A.; Azencott, C. A.; Chen, J. H.; Baldi, P. Learning to predict chemical reactions. *J. Chem. Inf. Mod.* **2011**, *51*, 2209–22.
- (6) Jin, W.; Coley, C. W.; Barzilay, R.; Jaakkola, T. In *Predicting Organic Reaction Outcomes with Weisfeiler-Lehman Network*, NeurIPS 2017; 2017.
- (7) Schwaller, P.; Laino, T.; Gaudin, T.; Bolgar, P.; Hunter, C. A.; Bekas, C.; Lee, A. A. Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction. *ACS Cent. Sci.* **2019**, *5*, 1572–1583.
- (8) Coley, C. W.; Jin, W.; Rogers, L.; Jamison, T. F.; Jaakkola, T. S.; Green, W. H.; Barzilay, R.; Jensen, K. F. A graph-convolutional neural network model for the prediction of chemical reactivity. *Chem. Sci.* **2019**, *10*, 370–377.
- (9) 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.
- (10) Lu, J.; Zhang, Y. Unified Deep Learning Model for Multitask Reaction Predictions with Explanation. *J. Chem. Inf. Mod.* **2022**, *62*, 1376–1387.
- (11) Irwin, R.; Dimitriadis, S.; He, J.; Bjerrum, E. J. Chemformer: a pre-trained transformer for computational chemistry. *Mach. Learn.: Sci. Technol.* **2022**, *3*, No. 015022.
- (12) Bi, H.; Wang, H.; Shi, C.; Coley, C.; Tang, J.; Guo, H., Non-Autoregressive Electron Redistribution Modeling for Reaction Prediction. In *Proc. 38th Int. Conf. Mach. Learn.*, Meila, M.; Zhang, T., Eds.; 2021; Vol. 139, pp 904–913.
- (13) Thakkar, A.; Kogej, T.; Raymond, J. L.; Engkvist, O.; Bjerrum, E. J. Datasets and their influence on the development of computer assisted synthesis planning tools in the pharmaceutical domain. *Chem. Sci.* **2020**, *11*, 154–168.
- (14) Shim, E.; Tewari, A.; Cernak, T.; Zimmerman, P. M. Machine Learning Strategies for Reaction Development: Toward the Low-Data Limit. *J. Chem. Inf. Mod.* **2023**, *63*, 3659–3668.
- (15) Diels, O.; Alder, K. Synthesen in der hydroaromatischen Reihe. *Liebigs Ann. Chem.* **1928**, *460*, 98–122.
- (16) Funel, J. A.; Abele, S. Industrial applications of the Diels-Alder reaction. *Angew. Chem. Int. Ed.* **2013**, *52*, 3822–3863.
- (17) Gregoritz, M.; Brandl, F. P. The Diels-Alder reaction: A powerful tool for the design of drug delivery systems and biomaterials. *Eur. J. Pharm. Biopharm.* **2015**, *97*, 438–53.
- (18) Briou, B.; Ameduri, B.; Boutevin, B. Trends in the Diels-Alder reaction in polymer chemistry. *Chem. Soc. Rev.* **2021**, *50*, 11055–11097.
- (19) Beker, W.; Gajewska, E. P.; Badowski, T.; Grzybowski, B. A. Prediction of Major Regio-, Site-, and Diastereoisomers in Diels-Alder Reactions by Using Machine-Learning: The Importance of Physically Meaningful Descriptors. *Angew. Chem., Int. Ed.* **2019**, *58*, 4515–4519.
- (20) Cao, C.; Liu, L. Topological Steric Effect Index and Its Application. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 678–687.
- (21) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754.
- (22) Vargas, S.; Hennefarth, M. R.; Liu, Z.; Alexandrova, A. N. Machine Learning to Predict Diels-Alder Reaction Barriers from the Reactant State Electron Density. *J. Chem. Theor. Comp.* **2021**, *17*, 6203–6213.
- (23) Espley, S. G.; Farrar, E. H. E.; Buttar, D.; Tomasi, S.; Grayson, M. N. Machine learning reaction barriers in low data regimes: a horizontal and diagonal transfer learning approach. *Digital Disc.* **2023**, *2*, 941–951.
- (24) Zubatiuk, T.; Isayev, O. Development of multimodal machine learning potentials: Toward a physics-aware artificial intelligence. *Acc. Chem. Res.* **2021**, *54*, 1575–1585.
- (25) Reaxys; Elsevier Information Systems GmbH: 2022.
- (26) Schwaller, P.; Hoover, B.; Raymond, J.-L.; Strobelt, H.; Laino, T. Extraction of organic chemistry grammar from unsupervised learning of chemical reactions. *Sci. Adv.* **2021**, *7*, No. eabe4166.
- (27) Landrum, G. *RDKit: Open-source cheminformatics*; 2006.
- (28) Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. Extended tight-binding quantum chemistry methods. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2021**, *11*, No. e1493.
- (29) 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. Theor. Comp.* **2019**, *15*, 1652–1671.
- (30) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams, Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. C.01*; Wallingford, CT, 2016.
- (31) Halgren, T. A. Merck Molecular Force Field III Molecular Geometries and Vibrational Frequencies for MMFF94. *J. Comput. Chem.* **1996**, *17*, 553–586.
- (32) Wang, S.; Witek, J.; Landrum, G. A.; Riniker, S. Improving Conformer Generation for Small Rings and Macrocycles Based on Distance Geometry and Experimental Torsional-Angle Preferences. *J. Chem. Inf. Mod.* **2020**, *60*, 2044–2058.
- (33) Guba, W.; Meyder, A.; Rarey, M.; Hert, J. Torsion Library Reloaded: A New Version of Expert-Derived SMARTS Rules for Assessing Conformations of Small Molecules. *J. Chem. Inf. Mod.* **2016**, *56*, 1–5.
- (34) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, non-covalent interactions, excited states, and transition elements: two new

functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

(35) Frisch, M. J.; Pople, J. A.; Binkley, J. S. Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. *J. Chem. Phys.* **1984**, *80*, 3265–3269.

(36) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(37) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–65.

(38) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.

(39) Hadipour, H.; Liu, C.; Davis, R.; Cardona, S. T.; Hu, P. Deep clustering of small molecules at large-scale via variational autoencoder embedding and K-means. *BMC Bioinf.* **2022**, *23*, 132.

(40) Stuyver, T.; Coley, C. W. Quantum chemistry-augmented neural networks for reactivity prediction: Performance, generalizability, and explainability. *J. Chem. Phys.* **2022**, *156* (8), No. 084104.

(41) Guan, Y.; Coley, C. W.; Wu, H.; Ranasinghe, D.; Heid, E.; Struble, T. J.; Pattanaik, L.; Green, W. H.; Jensen, K. F. Regioselectivity prediction with a machine-learned reaction representation and on-the-fly quantum mechanical descriptors. *Chem. Sci.* **2020**, *12*, 2198–2208.

(42) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

(43) Ozturk, H.; Ozgur, A.; Schwaller, P.; Laino, T.; Ozkirimli, E. Exploring chemical space using natural language processing methodologies for drug discovery. *Drug Disc. Today* **2020**, *25*, 689–705.

(44) Lowe, D. M. Chemical reactions from US patents (1976-Sep2016). https://figshare.com/articles/dataset/Chemical_reactions_from_US_patents_1976-Sep2016_/5104873.

(45) Reiser, P.; Neubert, M.; Eberhard, A.; Torresi, L.; Zhou, C.; Shao, C.; Metni, H.; van Hoesel, C.; Schopmans, H.; Sommer, T.; Friederich, P. Graph neural networks for materials science and chemistry. *Commun. Mater.* **2022**, *3*, 93.

(46) Coley, C. W.; Green, W. H.; Jensen, K. F. RDChiral: An RDKit Wrapper for Handling Stereochemistry in Retrosynthetic Template Extraction and Application. *J. Chem. Inf. Mod.* **2019**, *59*, 2529–2537.

(47) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity A Rapid Access To Atomic Charges. *Tetrahedron* **1980**, *36*, 3219–3288.

(48) 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* (5), 434–443.

(49) Jin, W.; Coley, C. W.; Barzilay, R.; Jaakkola, T. Predicting Organic Reaction Outcomes with Weisfeiler-Lehman Network. In *Adv. Neur. Inf. Proc. Sys. (NIPS)*; 2017; pp 2607–2616.

(50) Beker, W.; Gajewska, E. P.; Badowski, T.; Grzybowski, B. A. Prediction of Major Regio-, Site-, and Diastereoisomers in Diels-Alder Reactions by Using Machine-Learning: The Importance of Physically Meaningful Descriptors. *Angew. Chem., Int. Ed. Engl.* **2019**, *58* (14), 4515–4519.

(51) Brown, F. K.; Houk, K. N. The STO-3G Transition Structure of the Diels-Alder Reaction. *Tetrahedron* **1984**, *25*, 4609–4612.

(52) Houk, K. N.; Li, Y.; Evanseck, J. D. Transition Structures of Hydrocarbon Pericyclic Reactions. *Angew. Chem., Int. Ed.* **1992**, *31*, 682–708.

(53) Houk, K. N.; Liu, F.; Yang, Z.; Seeman, J. I. Evolution of the Diels-Alder Reaction Mechanism since the 1930s: Woodward, Houk with Woodward, and the Influence of Computational Chemistry on Understanding Cycloadditions. *Angew. Chem., Int. Ed.* **2021**, *60*, 12660–12681.

(54) Boger, D. L. Diels-Alder Reactions of Heterocyclic Azadienes: Scope and Applications. *Chem. Rev.* **1986**, *86*, 781–793.

(55) Liu, F.; Liang, Y.; Houk, K. N. Theoretical elucidation of the origins of substituent and strain effects on the rates of Diels-Alder reactions of 1,2,4,5-tetrazines. *J. Am. Chem. Soc.* **2014**, *136*, 11483–11493.