

Quantum Mechanics and Machine Learning Synergies: Graph Attention Neural Networks to Predict Chemical Reactivity

Mohammadamin Tavakoli, Aaron Mood, David Van Vranken, and Pierre Baldi*



Cite This: *J. Chem. Inf. Model.* 2022, 62, 2121–2132



Read Online

ACCESS |



Metrics & More

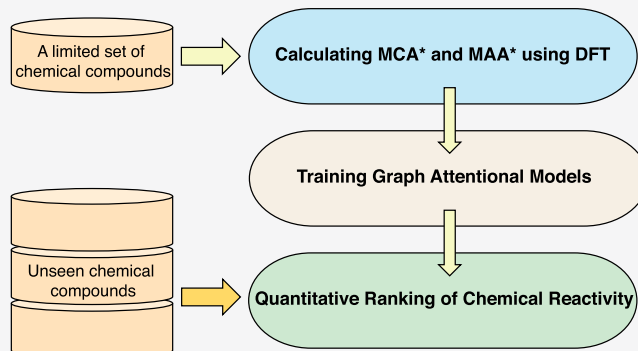


Article Recommendations



Supporting Information

ABSTRACT: There is a lack of scalable quantitative measures of reactivity that cover the full range of functional groups in organic chemistry, ranging from highly unreactive C–C bonds to highly reactive naked ions. Measuring reactivity experimentally is costly and time-consuming, and no single method has sufficient dynamic range to cover the astronomical size of chemical reactivity space. In previous quantum chemistry studies, we have introduced Methyl Cation Affinities (MCA*) and Methyl Anion Affinities (MAA*), using a solvation model, as quantitative measures of reactivity for organic functional groups over the broadest range. Although MCA* and MAA* offer good estimates of reactivity parameters, their calculation through Density Functional Theory (DFT) simulations is time-consuming. To circumvent this problem, we first use DFT to calculate MCA* and MAA* for more than 2,400 organic molecules thereby establishing a large data set of chemical reactivity scores. We then design deep learning methods to predict the reactivity of molecular structures and train them using this curated data set in combination with different representations of molecular structures. Using 10-fold cross-validation, we show that graph attention neural networks applied to a relational model of molecular structures produce the most accurate estimates of reactivity, achieving over 91% test accuracy for predicting the $\text{MCA}^* \pm 3.0$ or $\text{MAA}^* \pm 3.0$, over 50 orders of magnitude. Finally, we demonstrate the application of these reactivity scores to two tasks: (1) chemical reaction prediction and (2) combinatorial generation of reaction mechanisms. The curated data sets of MCA* and MAA* scores is available through the ChemDB chemoinformatics web portal at cdb.ics.uci.edu under Chemical Reactivities data sets.



INTRODUCTION

In general terms, the chemical reactivity of an atom in a molecule is its propensity toward being an electron donor or acceptor in a polar chemical reaction. Being able to assign reactivity scores to atoms and molecules can be useful to better understand chemical reactions and their mechanisms in different areas such as chemical synthesis, atmospheric chemistry, drug design, and materials sciences. Reaction rates have been measured experimentally for a long time, but this is typically a time-consuming process, which becomes increasingly costly as one tries to explore the most challenging reactions. Moreover, true solution-phase reaction rates are bounded by the rates of molecular diffusion which complicates quantifying the extremes of reactivity. Mayr and his colleagues have pioneered the empirical study of chemical reactivity by laboriously measuring the reactivity of the main organic functional groups and deriving corresponding scales of reactivity.^{1,2} However, due to the experimental limitations, their scale covers only a limited range of electrophiles and nucleophiles. An alternative approach to derive reactivity scores is to use quantum mechanical (QM) simulations.

Recently, we used QM with Density Functional Theory (DFT) simulations to investigate this problem for electron

donor and electron acceptor functional groups.^{3,4} Specifically, we applied DFT to over 100 diverse molecular structures and showed that, in general, methyl ion affinities, using a solvation model, are highly correlated to the Mayr reactivity scale. However, while using QM is faster than running laboratory experiments and can potentially cover a larger range of electrophiles and nucleophiles, the underlying DFT calculations still take up several hours for a molecule with only 20 atoms. Therefore, here we develop a more efficient approach leveraging the synergies between QM and machine learning,⁵ where we first use DFT to produce a substantial training set of chemical reactivity scores and then develop and train machine learning methods to predict chemical reactivity scores in real-time. These machine learning methods, in particular deep learning methods,⁶ have already been successfully applied to a variety

Special Issue: From Reaction Informatics to Chemical Space

Received: November 16, 2021

Published: January 12, 2022



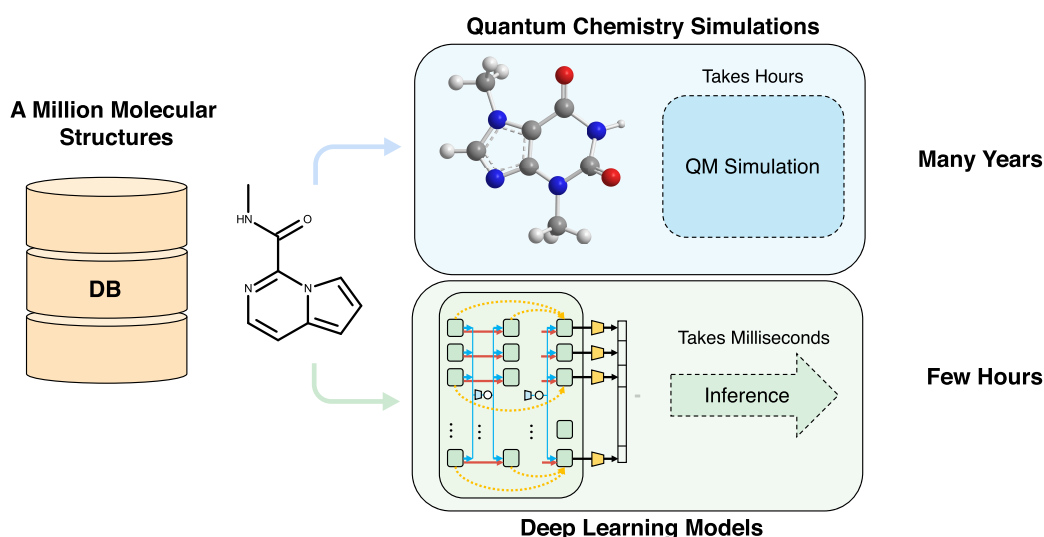


Figure 1. Typical speed differences in computing a given molecular property for a database of 1 M molecules using DFT versus machine learning. Assuming QM simulations take approximately 5 h per molecular structure, the total processing time for 1 M structures is $5 \times 3600(\text{s}) \times 10^6 \approx 500$ years. This processing time can be reduced to a few hours, after training a suitable neural architecture, assuming that inference time in a trained deep learning model is in the range of 5 ms: $5 \times 10^{-3}(\text{s}) \times 10^6 \approx \text{few hours}$.

of chemoinformatics problems, including the prediction of molecular properties^{7–9} and reactions.^{10–18} Training these models may take some time, but once trained, they are orders of magnitude faster than QM calculations and can generalize, making them a suitable complement to time-consuming DFT simulations (Figure 1).

In what follows, we first review some of the related work on chemical reactivity. Because the concept of chemical reactivity has different interpretations, we then describe the interpretation used in this work and how it correlates with methyl ion affinities. We then describe the curation of a training and testing data set obtained using DFT calculations. We develop different machine learning methods, tuned to different molecular representations, and train, test, and compare them using the curated data set. Finally, we demonstrate and discuss the potential applications of this approach for estimating the relative reactivity of atoms for the tasks of chemical reaction prediction and combinatorial generation of organic mechanisms.

RELATED WORK ON PREDICTING CHEMICAL REACTIVITY

The notion of chemical reactivity has different interpretations depending on whether one emphasizes thermodynamic or kinetic properties and corresponding properties such as activation energy, yield, rate constants, and stereoselectivity. These properties are usually measured experimentally or estimated via DFT calculations. Although DFT-based calculations can be accurate, their computational cost is not suitable for high-throughput analyses and for applications to a broad range of chemical structures.^{19,20} On the other hand, data-driven methods, such as deep learning-based predictive models, are orders of magnitude faster than DFT calculations and can be applied to a wide range of chemical structures.

To predict reaction yields and stereoselectivities, Sandfort et al.²¹ proposed a transferable machine learning method using the QM9 data set.^{22,23} In a similar direction, Schwaller et al.²⁴ proposed an attentional neural network to predict the reaction yield for Buchwald-Hartwig and Suzuki reactions with an R^2 coefficient greater than 0.6. Although the reaction yield and

selectivity, under certain conditions (e.g., pressure and temperature), are often interpreted as measures of chemical reactivity, this interpretation has several drawbacks.²⁵ For instance, a high yield can result from a low energy barrier; the converse is not necessarily true. A low energy barrier may still be associated with a low yield because of the presence of side reactions yielding other products. In addition, prediction of yield or stereoselectivity must be performed in the context of a chemical reaction, and therefore, it cannot be applied to an atom, functional group, or molecule in isolation.

To predict activation energy barriers, Mulliner et al.²⁶ used linear regression models to predict the activation energies (ΔE) of Michael acceptors using a data set derived by DFT calculations. Similarly, Palazzesi et al.²⁷ used DFT-computed (ΔE) data for the reactions between a limited set of acrylamides and cysteines to train a tree regressor model using physicochemical descriptors. The tree regressor model predicts the reactivity of covalent warheads with an R^2 regression coefficient greater than 0.5. In this line of work, DFT calculations are used to produce chemically narrow data sets that are used to train corresponding predictors of reactivity operating over a narrow range of chemical structures.

A different data-driven approach¹⁷ trains graph neural networks using the patent mining work of Lowe²⁸ to learn how to rank pairs of atoms with respect to the likelihood of their participation in a chemical reaction. This likelihood, interpreted as the reactivity of the corresponding pair of atoms, is also reaction-dependent and cannot be applied in more general settings to single atoms, functional groups, or molecules.

Here we use a new metric for measuring chemical reactivity, recently introduced in Mood et al.³ and Kadish et al.,⁴ based on methyl ion affinities. This metric has several advantages. First, it provides a reliable measure of chemical reactivity, as it is highly correlated with the Mayr reactivity scale^{2,29} which is an experimentally derived scale. Second, this metric is applicable to a broad range of chemistry, over 180 orders of magnitude of reactivity.^{3,4} Using this new metric and DFT calculations, we generate a new data set to train corresponding predictors of reactivity.

REACTIVITY METRIC

The extent to which a chemical compound can donate or accept electrons (referred to as nucleophilicity and electrophilicity, respectively) has always been an essential concept in organic chemistry. Until the pioneering work of Mayr and his team, the very idea that nucleophilicity or electrophilicity might be quantified on independent scales had eluded chemists. Through a massive experimental and theoretical undertaking, Mayr and his collaborators have made comprehensive and systematic measurements of reaction rates for reactions of various electrophiles and nucleophiles in the laboratory. Furthermore, they showed that the solution-phase electrophilicity E and solvent-dependent nucleophilicity N could be independently quantified using a logarithmic scale that correlates with the free energy of activation, allowing useful predictions of reaction rate constants, below diffusion control, using the equation $\log_{10} k_{25^\circ} = (E + N)s_N$, where s_N is a nucleophile-dependent parameter typically near unity, and k is the rate constant. The success of this equation centers around a focus on reactions that form bonds to carbon atoms; given the importance of solution-phase organic chemistry in biochemistry and medicine, this is a reasonable restriction.^{1,2}

Prior to Mayr's work, a limited range of nucleophilic parameters was available for the early Swain-Scott³⁰ and Ritchie relationships,³¹ which quantify nucleophilicity toward methyl bromide and esters, respectively. However, Mayr reactivity parameters are available for about 350 electrophiles and over 1200 nucleophile/solvent combinations. While experimental parameters seem plentiful, parameters for many of the canonical functional groups (C–C bonds, esters, amides, alkylolithiums, *tert*-butyl cation, and methyl anion) seem far out of reach. Additionally, nucleophilicity is far more difficult to quantify than electrophilicity. After over almost a century of study,³² there is still no comprehensive list of nucleophilicities of all the canonical organic functional groups, ranging from C–C bonds to naked alkyl anions, against any reference electrophile in any reference environment (gas phase or solvent system). The most common approach is to correlate equilibrium basicities with nucleophilicities. Basicity, in the form of pK_{aH} s, is readily available from titrations, but pK_{aH} s cannot be directly measured for functional groups far less acidic than the solvent or far more acidic than the solvent's conjugate acid.^{33–35} In particular, pK_{aH} s are generally not available for carbon nucleophiles with low basicity, such as alkenes and arenes. Mayr's team has already measured log-scale nucleophilicity (Ns_N) across 24 orders of magnitude versus a reference electrophile (4-MeOPh)₂CH⁺ (Mayr $E = 0$), ranging from toluene to an alpha-phenylpropionitrile carbanion. Equilibrium pK_{aH} s are available for nucleophiles across a broader range and correlate well with $\log k$ for a nucleophilic attack within functional group families, but they correlate poorly with $\log k$ between diverse functional group families.³⁶ pK_a s in DMSO can be accurately calculated,^{37,38} and the correlation between kinetic nucleophilicity and equilibrium basicity can be improved by including extra parameters, such as molar refractivity in the well-known Edwards equation;³⁹ but a better parameter than pK_{aH} is needed for training systems to predict organic reactivity.

Methyl cation affinity (MCA) is defined as the energy difference resulting from combining a methyl cation with a nucleophile; similarly, methyl anion affinity (MAA) is defined as the energy difference resulting from combining a methyl anion with an electrophile (Figure 2a). Mayr, Ofial, Zipse, and others

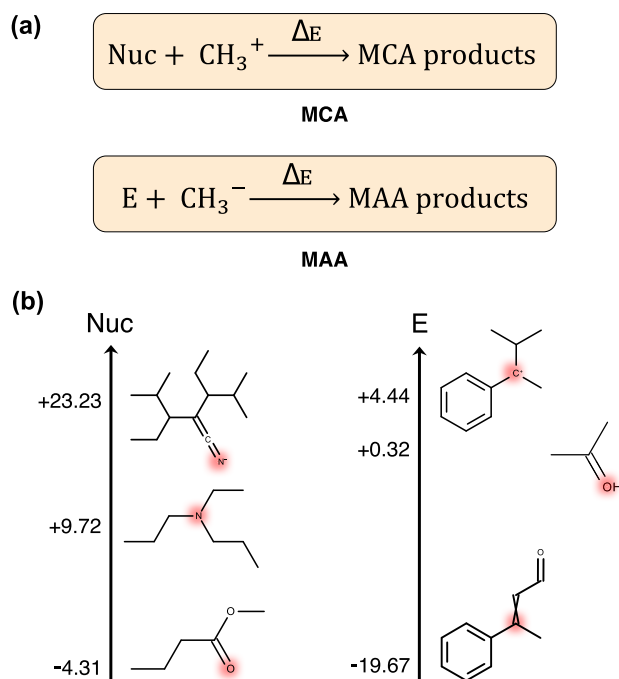


Figure 2. (a) Schematic definition of MCA and MAA: Nuc = nucleophile, E = electrophile, MCA = the negative of the energy difference resulting from combining a methyl cation with a nucleophile, and MAA = the negative of the energy difference resulting from combining a methyl anion with an electrophile. (b) A few molecules from the curated data set and the corresponding reactivity scores shown on two separate scales for nucleophilicity and electrophilicity.

have shown that MCA and MAA correlate with the experimentally measured nucleophilicities and electrophilicities on a relatively small range of reactivity^{40–44} (~30 orders of magnitude^{45,46}).

Recently, Van Vranken and his collaborators have shown that calculated methyl cation affinities and methyl anion affinities, with the inclusion of a solvation model, (MCA*) and (MAA*) where * denotes the solvation model, are highly correlated with measured nucleophilicity Ns_N and electrophilicity E over a broad range of molecules containing first- and second-row atoms.^{3,4} They used this correlation to expand the lower and upper ends of the nucleophilicity and electrophilicity scale produced by Mayr. Their work introduced MCA* and MAA* as new metrics to estimate reactivity parameters without carrying laboratory experiments. Calculated MCA* and MAA* provide useful reactivity scores across a much broader range of chemical reactivity than previous work (up to ~180 orders of magnitude). Since calculating MCA* or MAA* for only ground state electronic structure calculations, this approach is relatively fast compared to experimental approaches. However, electronic structure calculations are still resource intensive, particularly for larger molecules, hence the need to develop faster approaches.

The resulting chemical reactivity scores (MCA* and MAA*), formulated as the difference between the energy of reactants (nucleophile and methyl cation or electrophile and methyl anion, respectively) and products of the reactions shown in Figure 2a, can be interpreted in slightly different ways, depending on the entity to which the score is attributed to. Van Vranken et al. interpreted these quantities as the electrophilicity and nucleophilicity of the reacting functional groups.^{3,4} To a first order approximation, the cation affinity of a

Table 1. First Half of the Table Corresponds to the Prediction Accuracy Metric (Where the Prediction Is within ± 3.0 of the Actual Reactivity Value) and the Second Half Corresponds to the R^2 Coefficient Metric^a

	representation	atomic	group	molecular
		± 3.0 Accuracy		
electrophilicity	informative fingerprint	91.04 $\pm$ 0.01	—	—
	ECFP	—	—	80.21 $\pm$ 0.51
	SMILES	—	—	73.66 $\pm$ 0.44
	GCN	90.03 $\pm$ 0.63	86.97 $\pm$ 0.72	86.24 $\pm$ 0.66
	GAT	91.17 $\pm$ 0.67	87.66 $\pm$ 0.45	87.93 $\pm$ 0.79
nucleophilicity	informative fingerprint	92.01 $\pm$ 0.02	—	—
	ECFP	—	—	81.04 $\pm$ 0.41
	SMILES	—	—	72.59 $\pm$ 0.39
	GCN	90.91 $\pm$ 0.43	86.68 $\pm$ 0.77	87.42 $\pm$ 0.53
	GAT	92.14 $\pm$ 0.52	87.03 $\pm$ 0.51	87.24 $\pm$ 0.60
		R^2		
electrophilicity	informative fingerprint	0.90 $\pm$ 0.02	—	—
	ECFP	—	—	0.82 $\pm$ 0.02
	SMILES	—	—	0.72 $\pm$ 0.01
	GCN	0.92 $\pm$ 0.02	0.88 $\pm$ 0.01	0.88 $\pm$ 0.02
	GAT	0.94 $\pm$ 0.02	0.87 $\pm$ 0.01	0.89 $\pm$ 0.01
nucleophilicity	informative fingerprint	0.91 $\pm$ 0.01	—	—
	ECFP	—	—	0.83 $\pm$ 0.02
	SMILES	—	—	0.79 $\pm$ 0.02
	GCN	0.90 $\pm$ 0.01	0.85 $\pm$ 0.01	0.86 $\pm$ 0.01
	GAT	0.92 $\pm$ 0.02	0.89 $\pm$ 0.01	0.86 $\pm$ 0.01

^aAll the numbers are reported in the form of mean $\pm$ SD over a 10-fold cross-validation experiment. For each interpretation of the reactivity (atomic, group, molecular), the highest means are bolded.

small molecule that represents a canonical functional group is not expected to be different when the functional group is embedded in a larger molecule. In this case, the score (reactivity) can be attributed to the entire molecule. Lastly, the methyl ion affinities (MCA* and MAA*) are specific to individual atoms in a functional group, as shown in Figure 2a. Different methyl ion affinities could, in theory, be assigned to different atoms in an ambident functional group, but the overall reaction score would be dependent on the reacting partner. Since the goal of this work is to predict bond formation between atoms, we use the third attribution at the level of atoms. Throughout the rest of this paper, MCA* and MAA* are referred to as the nucleophilicity (electron donor) and electrophilicity scores (electron acceptor).

CURATED DATA SET OF CHEMICAL REACTIVITIES

Following the method for calculating chemical reactivity in Mood et al.³ and Kadish et al.,⁴ we compute the methyl cation affinities (MCA*) of 1232 nucleophiles and the methyl anion affinities (MAA*) of 1189 electrophiles. These molecules contain simple carbon skeleton structural variations to improve the generalizability of the trained models such as the ones shown in Figure 2b. The nucleophilic functional groups include the following: amines, ethers, amide anions, alkyl carbanions, aldehydes, ketones, esters, carboxylic acids, amides, enolates, nitronate anions, diazo compounds, cyanoalkyl anions, imines, nitriles, isonitriles, and bis(cyano)alkyl anions. The electrophilic functional groups include the following: iminium ions, imines, oxonium ions, aldehydes, ketones, esters, amides, benzyl cations, allyl cations, alkyl cations, carbonyl Michael acceptors, nitrile Michael acceptors, and nitro Michael acceptors. Methyl ion affinities (MCA* and MAA*) were calculated using TURBO-

MOLE V7.3^{47,48} at the PBE0(disp)(ref 49)/DEF2-TZVP(ref 50) COSMO(∞)^{51,52} level of theory.^{53,54} It must be noted that the MAA* and MCA* tend to correlate with experimentally determined Mayr E and Mayr N_s within a range of 3.0 orders of magnitude.^{3,4} For example, the DFT calculated MCA* for the compound at the top of the nucleophilicity scale in Figure 2b is +23.23, which means the actual reactivity score is in the range of +20.23 to +26.23.

DEEP LEARNING TO PREDICT CHEMICAL REACTIVITY

Deep learning⁶ methods have many applications in chemoinformatics, from molecular property prediction^{8,9,18,55} and optimization,^{14,16,56} to reaction prediction,^{10–17,57} to the acceleration of QM calculations.^{58,59} Although it takes time to train deep learning methods, at inference time they are fast and tend to generalize well. Deep learning methods in chemoinformatics must be developed in tandem with the underlying chemical representations: while a feedforward neural network can be applied to vectorial or tensorial representations of fixed size, a recursive or graph neural network must be used in the case of variable size, structured, representations.⁶⁰ Next, we describe the representations and architectures we use for the problem of predicting electrophilicity and nucleophilicity.

Representation of Molecular Structures. As previously mentioned, there are several ways of attributing the reactivity score introduced in Mood et al.³ and Kadish et al.⁴ This score is defined as the difference between the free energy of the reactants and products in the reactions shown in Figure 2a. It is also calculated for the most reactive atom of the most reactive functional group in each molecule. Therefore, one can interpret this as (1) the reactivity of the atom which is bonding to the

methyl ion during the reaction (atomic reactivity); (2) the reactivity of the functional group which contains the atom that is bonding to the methyl ion (group reactivity); or (3) the reactivity of the molecule which is reacting with the methyl ion (molecular reactivity). Using our know-how and some preliminary exploration to avoid looking at all possible combinations of attribution and representations, we converged on the following list of possible representations: (1) informative fingerprint vector representation to learn the atomic reactivity; (2) extended connectivity fingerprint (ECFP) to learn the molecular reactivity; (3) SMILES string text representation to learn the molecular reactivity; and (4) graph representations to learn atomic, functional group, and molecular reactivity. Next, we develop deep learning models congruous with each representation.

■ DEEP LEARNING MODELS

Informative Fingerprint Vector Representation and Model. To perform an atom level prediction, each atom has to be mapped to a vector. The first method to find this mapping is to use chemical features associated with each atom in a molecule. These features fall into two categories: 1) graph-topological and 2) physical-chemical. Graph-topological features reflect patterns of connectivity and the neighborhood of atoms, as in standard molecular fingerprint representations. The physical-chemical features instead capture properties of the atom itself. Examples of physical-chemical features are the presence and type of filled and unfilled orbitals, electronegativity, and the location of the atom in the periodic table. Here, we associate a feature vector of length 52 to each atom. This vector corresponds to the concatenation of 44 graph topological features and 8 physical chemical features. We refer to this vector as the informative fingerprint vector representation of the corresponding atom. Then, we train two separate neural networks, one for electrophilicity prediction and one for nucleophilicity prediction. After a hyperparameter optimization phase carried out using Sherpa⁶¹ with the random search option, both resulting networks comprise two hidden layers with 32 and 16 units and one output linear unit for the regression task using the mean squared error. We use dropout^{62,63} on the hidden layers with a rate of 0.4. We use stochastic gradient descent with a learning rate initialized at 0.001 and a momentum of 0.85. We also use SPLASH activation functions initialized to ReLU activations⁶⁴ in the hidden layers. The results of this experiment for both electrophilicity and nucleophilicity are shown in [Table 1](#).

Extended Connectivity Fingerprint Representation and Model. When the reactivity score is attributed to the molecule (molecular reactivity), we use molecular fingerprints to represent the entire molecule as a binary or integer (count) vector. In this work, we use the well-known extended-connectivity fingerprints (ECFP)⁶⁵ to train two separate networks for electrophilicity and nucleophilicity prediction. The length of the fingerprints must be adjusted as there is a basic trade-off: longer fingerprints capture more information; however, they increase the risk of overfitting. Considering the length of the fingerprints and the radius as hyperparameters, after a grid search stage for choosing the best length and radius, we converge on a size of 512 with a radius of four. The hyperparameter search carried using Sherpa⁶¹ with the random search option yields an architecture with 64 and 32 units in the hidden layers, followed by a linear output unit. This hyperparameter search also results in a stochastic-gradient-descent learning rate initialized at 0.001 with a momentum of 0.85 and

an $L2$ regularization coefficient of 0.20 which is a relatively large value for $L2$ regularization and consistent with avoiding overfitting on a relatively small training data.

SMILES String Representation and Model. We also use canonical SMILES strings⁶⁶ to represent the molecules. In order to apply deep learning methods, the SMILES strings must be converted to a numerical format. The most straightforward and widely used embedding is the character level embedding.^{57,67,68} However, it is more efficient to use atomic symbols as the embedding units. Not only does this reduce the extra computations required by atoms represented with multiple characters, but it also provides a clearer separation between pairs of atoms versus atoms represented by multiple characters (e.g., Sc can be seen as either a sulfur atom connected to an aromatic carbon or a scandium atom). In this embedding, each atom and special character in a SMILES string is mapped into a high dimensional vector. This can be done through an embedding layer whose weights are learned during the training process. However, because we are using a relatively small training data set of reactivity scores, we avoid adding extra embedding parameters to the model by using a pretrained embedding of atoms. We use the trained atom embedding vectors used in Fooshee et al.¹³ for an atom classification task within a reaction prediction pipeline. In this case, atoms are mapped onto a 10-dimensional vector space. Figure 3 shows the t-SNE 2D visualization of the embedded atoms preserving some of their physical-chemical properties and separating them from special characters.

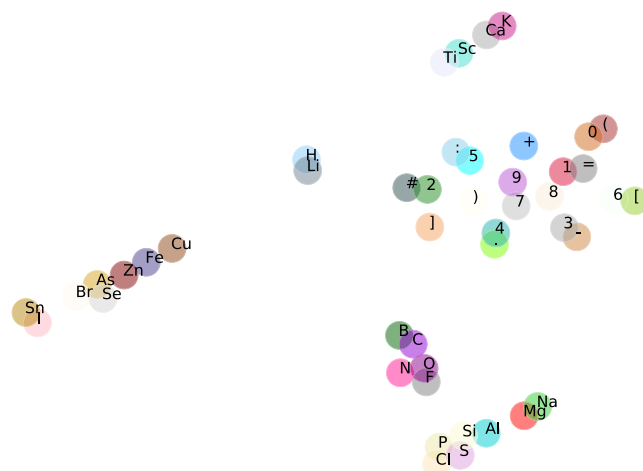


Figure 3. A two-dimensional t-SNE visualization of the numerical embedding for atom symbols and special characters found in SMILES strings. Special characters are clustered together and far from the atom symbols. The embeddings of atoms with similar properties are close to each other.

To predict the reactivity as the molecular property (i.e., molecular reactivity), the molecules are represented as SMILES strings. Two neural networks with the same architecture are employed to predict electrophilicity and nucleophilicity. The architecture of the neural networks consists of a pretrained atom embedding layer followed by a one-dimensional convolution layer with a window size of five. This is followed by a bidirectional LSTM layer with 16 hidden units. Then, the output is computed by a linear unit fully connected to the previous layer. To avoid the risk of overfitting because of the large number of parameters, an $L2$ regularization with $\lambda = 0.20$ is applied for

each weight of the convolution and LSTM layers. We use dropout with a 0.5 rate at each layer, and the learning rate was fixed at 0.001. All the hyperparameters are chosen using Sherpa⁶¹ hyperparameter optimization software with the random search option. The results of this experiment are shown in Table 1.

Graph Representation and Model. Finally, there are deep learning methods that can be applied directly to graph-structured data, such as knowledge graphs, social networks, parse trees, and molecules.^{8,60,69–72} Within these methods, the graph convolutional network (GCN), or outer recursive neural network approach, is particularly suited for processing molecular graphs through an iterative message passing mechanism that aggregates information about each atom's over increasing larger neighborhoods. Here we follow an approach similar to Schlichtkrull et al.⁷⁰ to model a molecule as a relational graph structure. Each molecule is first represented as an undirected labeled graph $G = (V, E, R, S)$, where the nodes in V correspond to the atoms and the edges in E to the bonds. The labels for the vertices in R correspond to atom types (e.g., C, O). The labels for the edges in S correspond to edge types (single, double, triple, aromatic). In the beginning, each node of the graph is associated with a vector representation (initial mapping) carrying information about the node. The edges of the graph can be treated in different ways; they can be mapped to real-valued vectors using learnable embedding weights, or they can be mapped to binary vectors using the one-hot encoding of the bond types. Since we are focusing on organic structures, there are only four types of bonds in the data set, and therefore, the one-hot encoding approach is both reasonable and economical. Thus, a molecule with n atoms and a feature vector of length d for each atom (initial mapping) can be represented by an $n \times d$ node-feature matrix H^0 together with an $n \times n \times d'$ adjacency tensor A (here $d' = 4$). Each row in the node-feature matrix is the vector representation of the corresponding atom. In the adjacency tensor, for a pair of vertices i and j , the corresponding vector of length four represents the one-hot encoding of the corresponding bond type. For atoms that are not bonded, the corresponding vector is (0, 0, 0, 0). Using this representation, we can apply a graph convolutional neural network to recursively update the representation of each atom. The recursive propagation of information in the convolutional neural network is given by

$$h_i^{l+1} = h_i^l + \sigma \left(\sum_{j \in N(i)} W^l h_j^l \right) \quad (1)$$

Here h_i^l is the vector representation associated with node i at level l . W^l denotes the shared weights of the convolution applied from level l to level $l + 1$. $N(i)$ denotes the neighborhood of vertex i , consisting of its immediate neighbors, and σ is a nonlinear transformation. In order to take different bond types into account, we use the approach in Schlichtkrull et al.⁷⁰ using a different set of weights for each bond type, resulting in the form

$$h_i^{l+1} = h_i^l + \sigma \left(\sum_{k \in S} \sum_{j \in N_k(i)} (W_k^l h_j^l) \right) \quad (2)$$

where $N_k(i)$ denotes the set of nodes connected to node i through the edge type $k \in S$. To train this graph neural network, we optimize a multiobjective loss function. In addition to minimizing the error between the predicted reactivity and the

actual value of MCA* or MAA* for the node of interest, we penalize the network whenever the predicted reactivity of other nodes is greater than the predicted reactivity of the node of interest. This can be formulated as follows

$$\mathcal{L} = \|f(h_i^L) - y\|^2 + \beta [\text{ReLU}(\max_j (f(h_j^L) - f(h_i^L)))]^2 \quad (3)$$

where i is the index of the atom which reacts with methyl ions, h^L is the atom representation produced by the last convolutional layer of the GCN (L), and $f(\cdot)$ is a fully connected linear function from $\mathbb{R}^d$ to $\mathbb{R}$ which outputs the reactivity prediction. The first loss term is the standard least-squares regression loss for predicting the target value of the MCA* or MAA* masked to the node of interest i . The second term looks at all the other nodes j and the corresponding reactivity prediction $f(h_j^L)$, computes the maximum reactivity difference with respect to node i , retains it only in the unwanted case where this difference is positive (through the ReLU function), and applies a square loss with a weighting hyperparameter β . In other words, the reactions shown in Figure 2a have the highest reaction rate among all other plausible reactions with the same set of reactants. For the initial mapping of each atom (h_i^0) we use a one-hot vector encoding of the atom type concatenated with eight physical-chemical features. As can be seen from eq 2, applying the first level of graph convolution updates the node representation using information from its immediate (one-hop) neighbors. To incorporate information from nodes up to three hops away into an atom's representation, we apply three layers of graph convolution with node representations of length 38, 24, 16, and 8 chosen after some exploratory experimentation. We also concatenate a count vector of 10 predefined molecular graph connectivity patterns to the output of the top convolutional layer. These predefined patterns have graph lengths greater than three and thus cannot be captured by the three-layer GCN. Adding these patterns to the node representations results in a slight performance improvement without requiring any additional graph attention layers. ReLU activation functions are used for each of the graph convolutional layers. An L_2 regularization with $\lambda = 0.2$ is applied to all the weights to avoid any overfitting. All the hyperparameters including the length of atom representation at each layer, L_2 regularization, and the learning rate are chosen using Sherpa⁶¹ hyperparameter optimization software with the random search option. The GCN performance is shown in Table 1.

There are a number of graph pooling mechanisms, such as those described in Lee et al.⁷³ and Murphy et al.,⁷⁴ that can be used to predict molecular reactivity and functional group reactivity after the graph convolutions. Because of our relatively small training data set, to avoid the risk of overfitting, we try simple pooling mechanisms that require minimal additional learning steps. In addition to the function f described above, we use an element-wise max-pooling, average-pooling, and the pooling mechanism introduced in Duvenaud et al.⁶⁹ which is described in eq 4, where $G(V)$ is the final representation of the entire graph, and N is the number of input nodes to the pooling operation. When predicting molecular reactivity, $N = |V|$; when predicting group reactivity, N is the number of atoms within the functional group which reacts with the methyl ions. As previously seen, h_i^L is a vector of size n_L representing the node i in the top convolutional layer L , and W_p is the shared pooling weight matrix of size (n_L, m) . Since this pooling is known to work better with relatively longer node representations,⁶⁹ we set $m =$

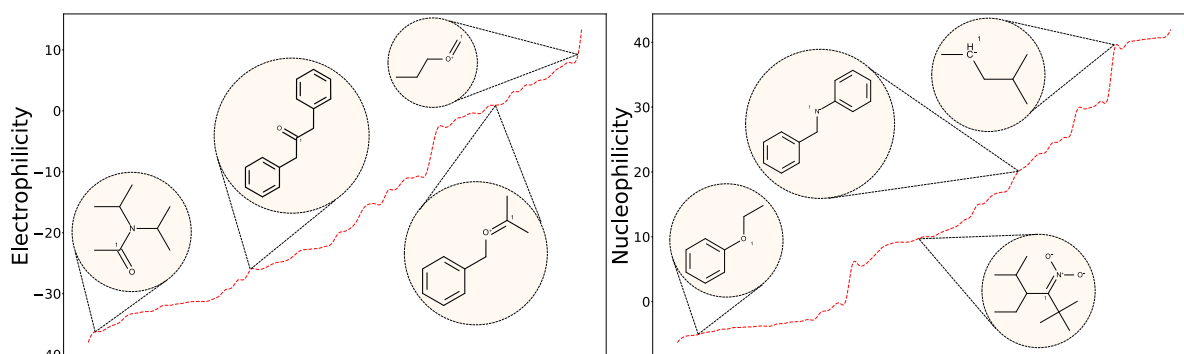


Figure 4. Range of electrophilicity and nucleophilicity in the curated data set, covering ~ 50 orders of magnitude in each case, together with five examples of electrophiles and nucleophiles. The most reactive atom in each sample is labeled as 1.

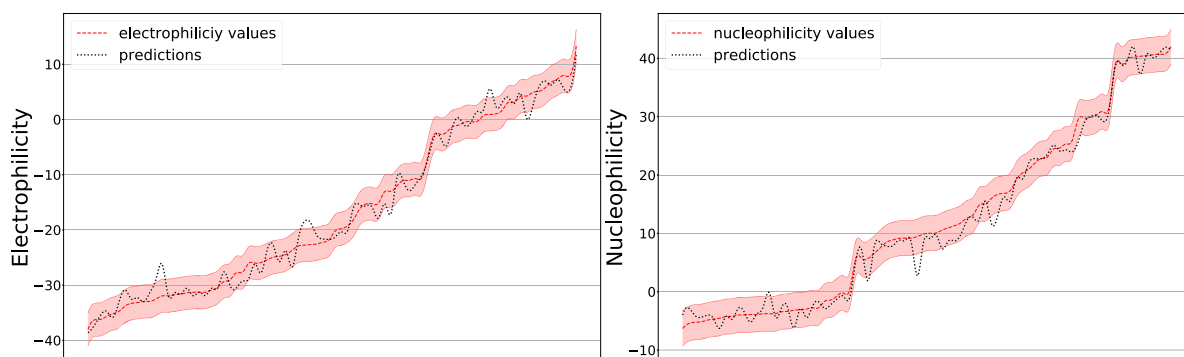


Figure 5. Predictions of the best performing GAT models for (left) electrophilicity and (right) nucleophilicity. The red strip is showing the actual electrophilicity and nucleophilicity plus or minus three orders of magnitude for the test samples. The black dotted line is the predicted reactivity.

n_i . In our experiments, however, we found that for the prediction of molecular reactivity and group reactivity max-pooling and average-pooling work slightly better than the mechanism introduced in Duvenaud et al.⁶⁹

$$G(V) = \sum_{i=1}^N \frac{\exp(o_i)}{\sum_{i=1}^N \exp(o_i)}, \quad o_i = h_i^L W_p \quad (4)$$

Graph Attention Networks. Another recently introduced type of neural network which can operate on graph data is the Graph Attention Network (GAT).⁷¹ In this approach, node representations are updated by a weighted message passing scheme between neighbors, where the weights are calculated through an attention mechanism as follows:

$$h_i^{l+1} = \text{softmax}_m \left(\sum_{j \in N_i} \alpha_{(m)}^{(ij)} h_j^l W_{(m)}^l \right) \quad \text{with} \quad \alpha_{(m)}^{(ij)} = \frac{\exp(A(W_{(m)}^l h_i \| W_{(m)}^l h_j))}{\sum_{k \in N_i} \exp(A(W_{(m)}^l h_i \| W_{(m)}^l h_k))} \quad (5)$$

Here P is the number of attention mechanisms (also known as attention heads⁷¹), m ranges from 1 to P , $W_{(m)}^l$ denotes the shared weights of layer l and attention mechanism m , and A represents a shallow neural network with a leaky-ReLU output activation mapping the input vector into a scalar. The symbol $\|$ represents the concatenation operation, and h_i^{l+1} , as the updated node representation, is the concatenation of the output of different attention mechanisms. Through our experiments, we find that using a larger node representation at the last layer of the GAT together with averaging (instead of concatenation) works

better. We use three GAT layer representations of size 38, 24, 16, and 12 and concatenate the representation of the top layer with the same count vector of predefined patterns used with the GCN. The GAT is trained using the same loss function as described in eq 3. We experimented with several variations in terms of the number of attentions heads and bond representations and report only the best results ($P = 3$), although the differences observed were minor. The GAT results for the prediction of reactivity are shown in Table 1.

RESULTS

Data. The full curated data set of 2421 molecules and their scores, covering 53 orders of magnitude of chemical reactivity, is available for download from the UCI ChemDB chemoinformatics web portal at cdb.ics.uci.edu. Figure 4 illustrates the range of electrophilicity and nucleophilicity in the curated data set.

Comparative Analysis and Predictions. Table 1 shows the results obtained with four different optimized neural networks with the corresponding molecular representations. Each number in Table 1 corresponds to the average 10-fold cross-validation accuracy, together with the corresponding standard deviation, for both electrophilicity and nucleophilicity predictions. As previously mentioned, the reactivity numbers are valid within 3 orders of magnitude. Thus, a prediction is considered to be correct if the predicted reactivity is within 3 orders of magnitude of the actual MCA* or MAA*. Although the informative vector representation of atoms shows a good performance, it has several downsides and limitations. The informative features are specifically tailored for our data set and are not generalizable to unseen molecular structures. Also, there

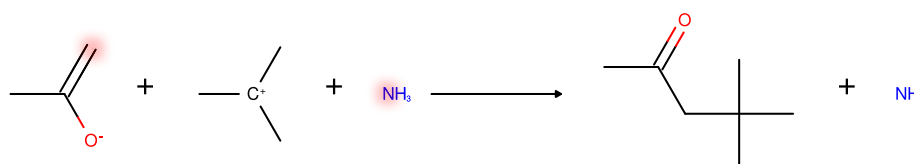


Figure 6. An example of a chemical reaction that can be problematic for an automated reaction prediction system. This reaction has two potential electron donors, marked in red. While both donors are reasonable, the nitrogen's reactivity score is considerably smaller than the carbon's reactivity score.

Table 2. Improvement in the Atom Classification Task Described in Fooshee et al.¹³ Obtained by Reranking the Atoms Based on Their Predicted MCA* and MAA* Values

	model	top-5		top-10	
		before	after	before	after
electrophilicity (electron sink)	informative fingerprint	0.43	0.51	0.35	0.42
	GAT	0.43	0.50	0.35	0.41
nucleophilicity (electron source)	informative fingerprint	0.41	0.48	0.32	0.40
	GAT	0.41	0.48	0.32	0.39

is no guarantee that the extracted features are the best representation of atoms. Since human chemists have designed these features based on their experience, another chemist might come up with a different set of features. On the other hand, for the SMILES representation and the corresponding networks, although there are no manually tailored features, these networks do not show comparable performance in comparison to other representations. The reasons for this poor performance might be the implicit rules of writing SMILES string which convert the molecular graphs into text representations. For instance, the long dependency between two parentheses that specify a branch of atoms might not be captured with recurrent neural networks such as LSTMs. Finally, the graph representation of molecular structure does not have any of the aforementioned drawbacks, and the corresponding models produce the best results. Figure 5 illustrates the performance of the best graph attention model for both electrophilicity and nucleophilicity predictions.

When comparing the performance of different models associated with different interpretations of chemical reactivity (atomic, group, molecular), one observes that the models based on atomic reactivity consistently have higher accuracy and lower standard deviation than the models associated with the group and molecular reactivity. One reason for this performance difference is the presence of nonreactive atoms (atoms with low electrophilicity or low nucleophilicity) in the reacting functional groups and molecules, which may play a distracting role. These nonreactive atoms contribute to the final (after pooling) representations in the models based on group or molecular reactivity, whereas by definition, these nonreactive atoms are not present in the final representation of the models based on the atomic reactivity.

These results provide at least a proof-of-concept that the high-throughput prediction of the MCA* and MAA* scores is feasible. Next, we demonstrate how such a system could enable other tasks, such as chemical reaction prediction and combinatorial generation of chemical reaction mechanisms.

Chemical Reaction Prediction. In recent years, several deep-learning-based methods have been introduced to predict the outcome of chemical reactions under certain conditions. These systems operate primarily either on string (SMIRKS) representations⁵⁷ or molecular graph representations.^{10–13,75} All the molecular graph-based methods seek to identify the most reactive atoms within the reactants. For instance, Coley et al.¹²

identify the most reactive atoms (i.e., reaction centers) using a binary classifier that classifies atoms as members of the reaction center or not. Fooshee et al.¹³ focus on reaction mechanisms by identifying electron sources and sinks and ranking the corresponding pairings. Although these methods yield reasonably successful reaction predictors, they are not always able to accurately identify the most reactive atoms, especially in the presence of multiple reagent molecules. The method presented here for estimating reactivity scores for electrophiles and nucleophiles could be used to address some of these problems by using the scores to rank the atoms. Such ranking may accurately identify reaction centers and electron donors and acceptors, bypassing or complementing the complex methods described, for instance, in Coley et al.¹² and Fooshee et al.¹³ An example of this approach is depicted in Figure 6. For this example, a chemical reaction predictor such as Fooshee et al.¹³ might predict both the nitrogen (amine group) and the highlighted carbon as the potential electron donors (since both atoms are labeled as electron donors in the training set of such system). However, our predicted nucleophilicity score for the carbon is higher than the reactivity score predicted for the nitrogen (+21.64 vs +9.71), leading to the correct identification of the electron donor and the corresponding reaction.

To further evaluate the viability of our method in predicting MCA* and MAA*, we collect a test set of 100 organic mechanisms where the reacting functional groups are similar to the electrophiles and nucleophiles in the training data described in the Curated Data Set of Chemical Reactivities. These 100 organic mechanisms are extracted from the literature and are given in the Supporting Information. Using these 100 mechanisms, we follow the atom classification task presented in Fooshee et al.¹³ where atoms are labeled as electron donor/acceptor. Then, for those reactions where the true electron donor and the true electron acceptor atoms are predicted within top-k predictions, we rerank the top-k predictions based on their actual predicted MCA* and MAA*. We use the Mean Reciprocal Rank (MRR)⁷⁶ score to measure the success of the reranking procedure. The MRR score is defined in eq 6, where N is the number of test reactions, and a_i is the reactive atom (either the electron donor or the electron acceptor) of the i th test reaction. The MRR score ranges from 0.0 to 1.0, where a higher MRR corresponds to a more accurate ranking.

Table 3. Two Examples of Combinatorial Reaction Mechanisms^a

templates		substituents		$N(T_1) - N(S_1)$	$E(T_2) - E(S_2)$
T_1 (donor)	T_2 (acceptor)	S_1 (donor)	S_2 (acceptor)		
N[R]	C[CH+][R]	O=C([R])NC	[R]C(OC)=O	+11.47	+28.21
C=C([R])[O-]	C/[N+](C)=C\[R]	N#C[R]	O=C([R])N(C)C	+7.6	+18.48

^a T_1 and T_2 correspond to electron donor and electron acceptor templates, respectively. Similarly, S_1 and S_2 correspond to electron donor and electron acceptor substituents, respectively. $E(.)$ and $N(.)$ denote the predicted electrophilicity and nucleophilicity values. The last two columns show the difference between the reactivity of templates and substituents for both electron donors and electron acceptors. The positive numbers indicate that the predicted reactivity of the templates is higher than the predicted reactivity of the substituents (see text).

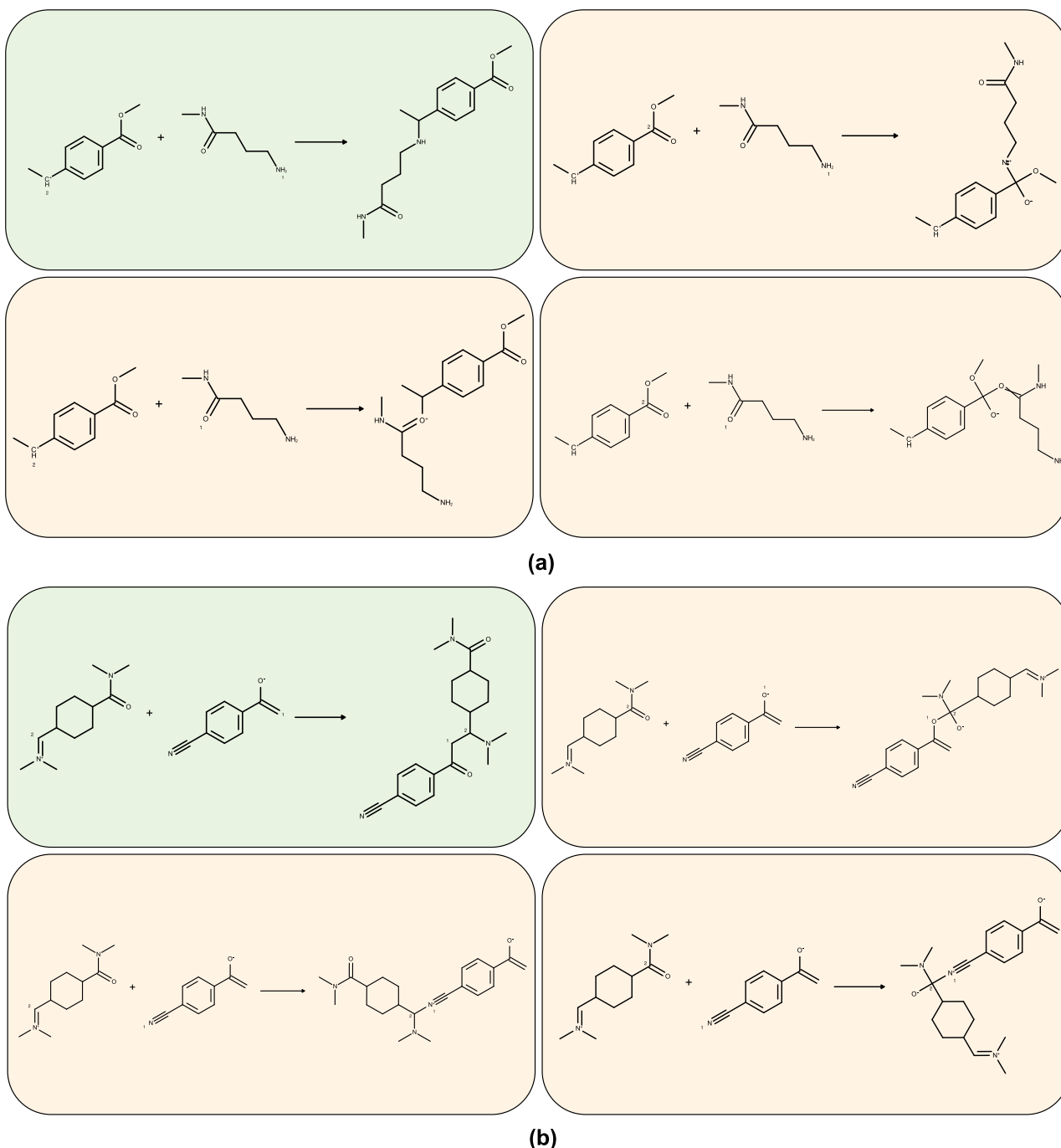


Figure 7. (a) The four possible reactions corresponding to the first row of Table 3 [$T_1 + S_2$; $T_2 + S_1$; $T_1 + T_2$; $S_1 + S_2$]. (b) Likewise, the four possible reactions corresponding to the second row of Table 3. In each figure, the reaction predicted using the reactivity scores is depicted inside a green box, whereas the remaining three less plausible reactions are depicted inside an orange box. These implausible reactions can be automatically filtered out during the process of combinatorial reaction generation. The plausibility and implausibility of these reactions are verified by expert chemists.

$$\text{MRR} = \frac{1}{N} \sum_{i=1}^N \frac{1}{\text{rank}(a_i)} \quad (6)$$

In Table 2, we show the improvement of the MRR score for $k = 5$ and 10 before and after the reranking. As it is shown, reranking the top- k predicted atoms based on their predicted MCA* and MAA* values improves the atom classification in Fooshee et al.¹³ by 18% in MRR score. This 18% improvement in MRR score corresponds to 25 reactions out of the 100 test reactions, having their reactive atoms (electron donor and electron acceptor) being ranked higher.

Combinatorial Generation of Chemical Reaction Mechanisms. Fooshee et al.¹³ introduced a method to combinatorially augment a training set of reactions. This method uses two fixed scaffolds respectively containing one electron donor group and one electron acceptor group (also called templates) and then combinatorially varies the decorative atoms (also called substituents) attached to the templates within realistic chemical constraints. Following this process, one can generate large numbers of elementary reactions covering a range of fundamental reaction classes. The most important constraint to enable this method to generate plausible reactions is that the atoms within the attaching substituent functional groups must be less reactive than the template atoms. Automatically enforcing this constraint requires a ranking of atoms based on their reactivity within different functional groups, and this can be done using the scales proposed here. A demonstration is given in Table 3 and Figure 7 showing that the proposed scales can be used to filter out the nonplausible combinatorially generated mechanisms. The first row of Table 3 and Figure 7a consider four functional groups in two reactants. The higher nucleophilicity of the amino group over the carboxamide and the higher electrophilicity of the carbocation over the ester carbonyl correctly predict the attack of the amino group on the carbocation (the green reaction in Figure 7a). Similarly, the second row of Table 3 and Figure 7b consider four other functional groups in two more reactants. The higher nucleophilicity of the enolate relative to the nitrile and the higher electrophilicity of the iminium ion over the carboxamide carbonyl group correctly predicts a Mannich reaction involving the addition of the enolate to the iminium ion (the green reaction in Figure 7b).

CONCLUSION

Methyl cation affinity and methyl anion affinity have been shown to be highly correlated to the reactivity of atoms in functional groups over a broad range of organic chemistry. Leveraging this correlation, we used DFT calculations to curate a data set of relative reactivity scores for 2,421 electrophilic and nucleophilic functional groups covering 53 orders of magnitude of chemical reactivity. This curated data set is available to the community and was used to train several deep neural networks, with different representations, to estimate reactivity. Through experiments, we have shown that graph attention neural networks outperform other methods and representations and can accurately estimate the reactivity with a 10-fold cross-validation accuracy of 92% showcasing another synergistic application of QM and Machine Learning methods.

In the future, it may be useful to incorporate the experimentally validated reactivity parameters from the Mayr database and to further expand the curated data set of DFT-derived reactivity scores (MCA* and MAA*) to a broader range

of molecular structures. All the proposed methods in this work are trained using the 2421 structures with decent coverage of basic functional groups over 53 orders of magnitude of chemical reactivity. However, the method in Mood et al.⁴ and Kadish et al.³ for calculating chemical reactivity is applicable to molecular structures with a chemical reactivity covering 180 orders of magnitude. Thus, there is significant room for curating larger data sets and training more refined machine learning systems to predict chemical reactivity, as well as using the machine learning results to guide the DFT calculations toward the most informative regions of chemical reaction space.

Data and Software Availability. The data sets of MCA* and MAA* scores along with the set of 100 test mechanisms are available through the ChemDB chemoinformatics web portal at cdb.ics.uci.edu under Chemical Reactivities Data sets. The models and training scripts with a description of the required packages are available at github.com/amintavakol/rrs_prediction.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.1c01400>.

Study on generalization capability of models and set of 100 test reactions (PDF)

AUTHOR INFORMATION

Corresponding Author

Pierre Baldi – Department of Computer Science, University of California, Irvine, Irvine, California 92697, United States;
orcid.org/0000-0001-8752-4664; Email: pfbaldi@uci.edu

Authors

Mohammadamin Tavakoli – Department of Computer Science, University of California, Irvine, Irvine, California 92697, United States

Aaron Mood – Department of Chemistry, University of California, Irvine, Irvine, California 92697, United States

David Van Vranken – Department of Chemistry, University of California, Irvine, Irvine, California 92697, United States

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jcim.1c01400>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work is partially supported by NSF grants 1633631 and 195811 to D.V.V. and P.B. We are grateful to OpenEye for their academic license.

REFERENCES

- (1) Mayr, H.; Ofial, A. R. Do general nucleophilicity scales exist? *J. Phys. Org. Chem.* **2008**, *21*, 584–595.
- (2) Mayr, H.; Ofial, A. A quantitative approach to polar organic reactivity. *SAR QSAR Env. Res.* **2015**, *26*, 619–646.
- (3) Mood, A.; Tavakoli, M.; Gutman, E.; Kadish, D.; Baldi, P.; Van Vranken, D. L. Methyl Anion Affinities of the Canonical Organic Functional Groups. *J. Org. Chem.* **2020**, *85*, 4096–4102.
- (4) Kadish, D.; Mood, A. D.; Tavakoli, M.; Gutman, E. S.; Baldi, P.; Van Vranken, D. L. Methyl Cation Affinities of Canonical Organic Functional Groups. *J. Org. Chem.* **2021**, *86*, 3721.

- (5) Sadowski, P.; Fooshee, D.; Subrahmanya, N.; Baldi, P. Synergies Between Quantum Mechanics and Machine Learning in Reaction Prediction. *J. Chem. Inf. Model.* **2016**, *56*, 2125–2128.
- (6) Baldi, P. *Deep Learning in Science: Theory, Algorithms, and Applications*; Cambridge University Press: Cambridge, UK, 2021; in press.
- (7) Swamidass, S. J.; Chen, J.; Bruand, J.; Phung, P.; Ralaivola, L.; Baldi, P. Kernels for Small Molecules and the Prediction of Mutagenicity, Toxicity, and Anti-Cancer Activity. *Bioinformatics* **2005**, *21*, i359–368. Proceedings of the 2005 ISMB Conference.
- (8) Lusci, A.; Pollastri, G.; Baldi, P. Deep architectures and deep learning in chemoinformatics: the prediction of aqueous solubility for drug-like molecules. *J. Chem. Inf. Model.* **2013**, *53*, 1563–1575.
- (9) Lusci, A.; Fooshee, D.; Browning, M.; Swamidass, J.; Baldi, P. Accurate and efficient target prediction using a potency-sensitive influence-relevance voter. *J. Cheminf.* **2015**, *7*, 63.
- (10) Kayala, M.; Azencott, C.; Chen, J.; Baldi, P. Learning to predict chemical reactions. *J. Chem. Inf. Model.* **2011**, *51*, 2209–2222.
- (11) Kayala, M.; Baldi, P. ReactionPredictor: Prediction of Complex Chemical Reactions at the Mechanistic Level Using Machine Learning. *J. Chem. Inf. Model.* **2012**, *52*, 2526–2540.
- (12) 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.
- (13) Fooshee, D.; Mood, A.; Gutman, E.; Tavakoli, M.; Urban, G.; Liu, F.; Huynh, N.; Van Vranken, D.; Baldi, P. Deep learning for chemical reaction prediction. *Mol. Syst. Des. Eng.* **2018**, *3*, 442.
- (14) De Cao, N.; Kipf, T. MolGAN: An implicit generative model for small molecular graphs. *arXiv preprint arXiv:1805.11973*, 2018. <https://arxiv.org/pdf/1805.11973.pdf> (accessed 2021-12-29).
- (15) You, J.; Liu, B.; Ying, Z.; Pande, V.; Leskovec, J. Graph convolutional policy network for goal-directed molecular graph generation. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 6410–6421.
- (16) Zhou, Z.; Kearnes, S.; Li, L.; Zare, R. N.; Riley, P. Optimization of molecules via deep reinforcement learning. *Sci. Rep.* **2019**, *9*, 10752.
- (17) 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.
- (18) Tavakoli, M.; Baldi, P. Continuous Representation of Molecules Using Graph Variational Autoencoder. *arXiv preprint arXiv:2004.08152*, 2020. <https://arxiv.org/abs/2004.08152> (accessed 2021-12-29).
- (19) Mendoza-Wilson, A. M.; Glossman-Mitnik, D. CHIH-DFT study of the electronic properties and chemical reactivity of quercetin. *J. Theor. Chem.* **2005**, *716*, 67–72.
- (20) Vargas-Sánchez, R.; Mendoza-Wilson, A.; Balandrán-Quintana, R.; Torrescano-Urrutia, G.; Sánchez-Escalante, A. Study of the molecular structure and chemical reactivity of pinocembrin by DFT calculations. *Comput. Theor. Chem.* **2015**, *1058*, 21–27.
- (21) Sandfort, F.; Strieth-Kalthoff, F.; Kühnemund, M.; Beecks, C.; Glorius, F. A structure-based platform for predicting chemical reactivity. *Chem.* **2020**, *6*, 1379–1390.
- (22) Ruddigkeit, L.; Van Deursen, R.; Blum, L. C.; Reymond, J.-L. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *J. Chem. Inf. Model.* **2012**, *52*, 2864–2875.
- (23) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **2014**, *1*, 140022.
- (24) Schwaller, P.; Vaucher, A. C.; Laino, T.; Reymond, J.-L. Prediction of chemical reaction yields using deep learning. *Mach. Learn.: Sci. Technol.* **2021**, *2*, 015016.
- (25) Jorner, K.; Tomberg, A.; Bauer, C.; Sköld, C.; Norrby, P.-O. Organic reactivity from mechanism to machine learning. *Nat. Rev. Chem.* **2021**, *5*, 240–255.
- (26) Mulliner, D.; Wondrousch, D.; Schüürmann, G. Predicting Michael-acceptor reactivity and toxicity through quantum chemical transition-state calculations. *Org. Biomol. Chem.* **2011**, *9*, 8400–8412.
- (27) Palazzesi, F.; Hermann, M. R.; Grundl, M. A.; Pautsch, A.; Seeliger, D.; Tautermann, C. S.; Weber, A. Blreactive: a machine-learning model to estimate covalent warhead reactivity. *J. Chem. Inf. Model.* **2020**, *60*, 2915–2923.
- (28) Lowe, D. M. *Extraction of chemical structures and reactions from the literature*. Ph.D. thesis, University of Cambridge, 2012.
- (29) Mayer, H.; Gomez, F.; Wierstra, D.; Nagy, I.; Knoll, A.; Schmidhuber, J. A system for robotic heart surgery that learns to tie knots using recurrent neural networks. *Adv. Robot.* **2008**, *22*, 1521–1537.
- (30) Swain, C. G.; Scott, C. B. Quantitative correlation of relative rates. Comparison of hydroxide ion with other nucleophilic reagents toward alkyl halides, esters, epoxides and acyl halides. *J. Am. Chem. Soc.* **1953**, *75*, 141–147.
- (31) Ritchie, C. D. Cation-anion combination reactions. XIII. Correlation of the reactions of nucleophiles with esters. *J. Am. Chem. Soc.* **1975**, *97*, 1170–1179.
- (32) Harris, J. M.; McManus, S. P. *Nucleophilicity*; American Chemical Society: 1987; DOI: 10.1021/ba-1987-0215.
- (33) Bordwell, F. G. Equilibrium acidities in dimethyl sulfoxide solution. *Acc. Chem. Res.* **1988**, *21*, 456–463.
- (34) Bordwell, F. G.; Algrim, D. J. Acidities of anilines in dimethyl sulfoxide solution. *J. Am. Chem. Soc.* **1988**, *110*, 2964–2968.
- (35) Breslow, R.; Grant, J. L. Electrochemical determination of the basicities of benzyl, allyl, and propargyl anions, and a study of solvent and electrolyte effects. *J. Am. Chem. Soc.* **1977**, *99*, 7745–7746.
- (36) Mayr, A. R. P. F.; Ofial, H. Equilibrium Constants. *Acc. Chem. Res.* **2016**, *49*, 952–965.
- (37) Klamt, A.; Eckert, F.; Diedenhofen, M.; Beck, M. E. First principles calculations of aqueous p K a values for organic and inorganic acids using COSMO-RS reveal an inconsistency in the slope of the p K a scale. *J. Phys. Chem. A* **2003**, *107*, 9380–9386.
- (38) Ding, F.; Smith, J. M.; Wang, H. First-principles calculation of p K a values for organic acids in nonaqueous solution. *J. Org. Chem.* **2009**, *74*, 2679–2691.
- (39) Edwards, J. O. Correlation of relative rates and equilibria with a double basicity scale. *J. Am. Chem. Soc.* **1954**, *76*, 1540–1547.
- (40) Schindele, C.; Houk, K.; Mayr, H. Relationships between Carbocation stabilities and Electrophilic reactivity parameters, E: quantum mechanical studies of benzhydryl cation structures and stabilities. *J. Am. Chem. Soc.* **2002**, *124*, 11208–11214.
- (41) Seeliger, F.; Błażej, S.; Bernhardt, S.; Mąkosza, M.; Mayr, H. Reactions of nitroheteroarenes with carbanions: bridging aromatic, Heteroaromatic, and vinylic electrophilicity. *Chem.—Eur. J.* **2008**, *14*, 6108–6118.
- (42) Troshin, K.; Schindele, C.; Mayr, H. Electrophilicities of Symmetrically Substituted 1, 3-Diaryllallyl Cations. *J. Org. Chem.* **2011**, *76*, 9391–9408.
- (43) Allgäuer, D. S.; Jangra, H.; Asahara, H.; Li, Z.; Chen, Q.; Zipse, H.; Ofial, A. R.; Mayr, H. Quantification and Theoretical Analysis of the Electrophilicities of Michael Acceptors. *J. Am. Chem. Soc.* **2017**, *139*, 13318–13329.
- (44) Böttger, G. M.; Fröhlich, R.; Würthwein, E.-U. Electrophilic Reactivity of a 2-Azaallenium and of a 2-Azaallylium Ion. *Eur. J. Org. Chem.* **2000**, *2000*, 1589–1593.
- (45) Pellerite, M. J.; Brauman, J. I. Intrinsic barriers in nucleophilic displacements. A general model for intrinsic nucleophilicity toward methyl centers. *J. Am. Chem. Soc.* **1983**, *105*, 2672–2680.
- (46) Wei, Y.; Sastry, G. N.; Zipse, H. Methyl cation affinities of commonly used organocatalysts. *J. Am. Chem. Soc.* **2008**, *130*, 3473–3477.
- (47) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. Electronic structure calculations on workstation computers: The program system turbomole. *Chem. Phys. Lett.* **1989**, *162*, 165–169.
- (48) Furche, F.; Ahlrichs, R.; Hättig, C.; Klopper, W.; Sierka, M.; Weigend, F. Turbomole. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 91–100.
- (49) Perdew, J. P.; Ernzerhof, M.; Burke, K. Rationale for mixing exact exchange with density functional approximations. *J. Chem. Phys.* **1996**, *105*, 9982–9985.

- (50) Jensen, F. Atomic orbital basis sets. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, 3, 273–295.
- (51) Klamt, A.; Schüürmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin 2* **1993**, 799–805.
- (52) Klamt, A. *COSMO-RS: from quantum chemistry to fluid phase thermodynamics and drug design*; Elsevier: 2005.
- (53) Li, Z.; Jangra, H.; Chen, Q.; Mayer, P.; Ofial, A. R.; Zipse, H.; Mayr, H. Kinetics and Mechanism of Oxirane Formation by Darzens Condensation of Ketones: Quantification of the Electrophilicities of Ketones. *J. Am. Chem. Soc.* **2018**, 140, 5500–5515.
- (54) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, 32, 1456–1465.
- (55) Feinberg, E. N.; Sur, D.; Wu, Z.; Husic, B. E.; Mai, H.; Li, Y.; Sun, S.; Yang, J.; Ramsundar, B.; Pande, V. S. PotentialNet for molecular property prediction. *ACS Cent. Sci.* **2018**, 4, 1520–1530.
- (56) Olivecrona, M.; Blaschke, T.; Engkvist, O.; Chen, H. Molecular de-novo design through deep reinforcement learning. *J. Cheminf.* **2017**, 9, 48.
- (57) 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.
- (58) Snyder, J. C.; Rupp, M.; Hansen, K.; Müller, K.-R.; Burke, K. Finding Density Functionals with Machine Learning. *Phys. Rev. Lett.* **2012**, 108, 253002.
- (59) Hermann, J.; Schätzle, Z.; Noé, F. Deep-neural-network solution of the electronic Schrödinger equation. *Nat. Chem.* **2020**, 12, 891–897.
- (60) Urban, G.; Subrahmanya, N.; Baldi, P. Inner and Outer Recursive Neural Networks for Chemoinformatics Applications. *J. Chem. Inf. Model.* **2018**, 58, 207–211.
- (61) Hertel, L.; Collado, J.; Sadowski, P.; Baldi, P. *Sherpa: hyperparameter optimization for machine learning models*. 2018.
- (62) Srivastava, N.; Hinton, G. E.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, 15, 1929–1958.
- (63) Baldi, P.; Sadowski, P. The Dropout Learning Algorithm. *Artificial Intelligence* **2014**, 210, 78–122.
- (64) Tavakoli, M.; Agostinelli, F.; Baldi, P. Splash: Learnable activation functions for improving accuracy and adversarial robustness. *Neural Networks* **2021**, 140, 1.
- (65) Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **2010**, 50, 742–754.
- (66) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, 28, 31–36.
- (67) Liu, B.; Ramsundar, B.; Kawthekar, P.; Shi, J.; Gomes, J.; Luu Nguyen, Q.; Ho, S.; Sloane, J.; Wender, P.; Pande, V. Retrosynthetic reaction prediction using neural sequence-to-sequence models. *ACS Cent. Sci.* **2017**, 3, 1103–1113.
- (68) Ikebata, H.; Hongo, K.; Isomura, T.; Maezono, R.; Yoshida, R. Bayesian molecular design with a chemical language model. *J. Comp. Aid. Mol. Des.* **2017**, 31, 379–391.
- (69) Duvenaud, D. K.; Maclaurin, D.; Iparraguirre, J.; Bombarell, R.; Hirzel, T.; Aspuru-Guzik, A.; Adams, R. P. Convolutional networks on graphs for learning molecular fingerprints. *Adv. Neural Inf. Process. Syst.* **2015**, 2224–2232.
- (70) Schlichtkrull, M.; Kipf, T. N.; Bloem, P.; Van Den Berg, R.; Titov, I.; Welling, M. Modeling relational data with graph convolutional networks. *Euro. Sem. Web. Conf.* **2018**, 10843, 593–607.
- (71) Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. *Graph attention networks*. *arXiv preprint arXiv:1710.10903*, 2017. <https://arxiv.org/abs/1710.10903> (accessed 2021-12-29).
- (72) Kearnes, S.; McCloskey, K.; Berndl, M.; Pande, V.; Riley, P. Molecular graph convolutions: moving beyond fingerprints. *J. Comp. Aid. Mol. Des.* **2016**, 30, 595–608.
- (73) Lee, J.; Lee, I.; Kang, J. Self-attention graph pooling. *Int. Conf. Mach. Learn.* **2019**, 3734–3743.
- (74) Murphy, R.; Srinivasan, B.; Rao, V.; Ribeiro, B. Relational pooling for graph representations. *Int. Conf. Mach. Learn.* **2019**, 4663–4673.
- (75) Do, K.; Tran, T.; Venkatesh, S. Graph transformation policy network for chemical reaction prediction. *International Conference on Knowledge Discovery & Data Mining* **2019**, 750–760.
- (76) Radev, D. R.; Qi, H.; Wu, H.; Fan, W. Evaluating Web-based Question Answering Systems. *LREC*; 2002.

Recommended by ACS

What Does the Machine Learn? Knowledge Representations of Chemical Reactivity

Joshua A. Kammeraad, Paul M. Zimmerman, *et al.*

FEBRUARY 24, 2020
JOURNAL OF CHEMICAL INFORMATION AND MODELING

READ 

Deep Learning of Activation Energies

Colin A. Grambow, William H. Green, *et al.*

MARCH 27, 2020
THE JOURNAL OF PHYSICAL CHEMISTRY LETTERS

READ 

Predicting Reaction Products and Automating Reactive Trajectory Characterization in Molecular Simulations with Support Vector Machines

Gianmarco Grazioli, Carter T. Butts, *et al.*

MAY 07, 2019
JOURNAL OF CHEMICAL INFORMATION AND MODELING

READ 

Development and Application of a Data-Driven Reaction Classification Model: Comparison of an Electronic Lab Notebook and Medicinal Chemistry L...

Gian Marco Ghiandoni, Valerie J. Gillet, *et al.*

SEPTEMBER 17, 2019
JOURNAL OF CHEMICAL INFORMATION AND MODELING

READ 

Get More Suggestions >