ECAI 2024
U. Endriss et al. (Eds.)

doi:10.3233/FAIA240796

ReactAIvate: A Deep Learning Approach to Predicting Reaction Mechanisms and Unmasking Reactivity Hotspots

Ajnabiul Hoque[a,1], **Manajit Das**[a,2], **Mayank Baranwal**[b,c,*] **and Raghavan B. Sunoj**[a,d,**]

[a]Department of Chemistry, Indian Institute of Technology Bombay, India
[b]Department of Systems & Control Engineering, Indian Institute of Technology, India
[c]Tata Consultancy Services Research, Mumbai, India
[d]Centre for Machine Intelligence and Data Science, Indian Institute of Technology Bombay, India
ORCID (Ajnabiul Hoque): https://orcid.org/0000-0001-9807-3061, ORCID (Manajit Das): https://orcid.org/0000-0001-7709-8809, ORCID (Mayank Baranwal): https://orcid.org/0000-0001-9354-2826, ORCID (Raghavan B. Sunoj): https://orcid.org/0000-0002-6484-2878

Abstract. A chemical reaction mechanism (CRM) is a sequence of molecular-level events involving bond-breaking/forming processes, generating transient intermediates along the reaction pathway as reactants transform into products. Understanding such mechanisms is crucial for designing and discovering new reactions. One of the currently available methods to probe CRMs is quantum mechanical (QM) computations. The resource-intensive nature of QM methods and the scarcity of mechanism-based datasets motivated us to develop reliable ML models for predicting mechanisms. In this study, we created a comprehensive dataset with seven distinct classes, each representing uniquely characterized elementary steps. Subsequently, we developed an interpretable attention-based GNN that achieved near-unity and 96% accuracy, respectively for reaction step classification and the prediction of reactive atoms in each such step, capturing interactions between the broader reaction context and local active regions. The near-perfect classification enables accurate prediction of both individual events and the entire CRM, mitigating potential drawbacks of Seq2Seq approaches, where a wrongly predicted character leads to incoherent CRM identification. In addition to interpretability, our model adeptly identifies key atom(s) even from out-of-distribution classes. This generalizabilty allows for the inclusion of new reaction types in a modular fashion, thus will be of value to experts for understanding the reactivity of new molecules.

1 Introduction

The reliable prediction of chemical reactions holds paramount significance in pharmaceutical and materials manufacturing, and in understanding many processes in molecular biology [9, 5, 40]. A chemical reaction entails the reorganization of atoms and bonds within the initial reactants, resulting in the creation of novel molecules or compounds as the final products. To comprehend a chemical reaction, it is essential to know the underlying chemical transformations. A sequence of these transformation steps, also called elementary mechanistic steps, is generally expressed in the form of a chemical reaction mechanism (CRM). Many of these steps may involve a bond-forming/bond-breaking process characterized by the corresponding transition state. These elementary steps serve as building blocks for developing novel reactions and discerning side products. Knowledge of CRM can provide atomic-level insights into why the products are formed [19, 12].

One of the ways to identify a CRM is to perform quantum mechanical (QM) calculations [6, 3]. Such methods are often computationally demanding and often require substantial human attention rendering it a time-consuming task. In recent years, numerous ML studies have focused on retrosynthesis and forward synthesis prediction, with models mostly trained on the USPTO-50k dataset [39]. While the dataset proves valuable for direct product prediction, it lacks information about elementary reaction steps, thus lacking opportunities to understand CRM [45, 10]. In the realm of CRM prediction, navigating the complexities of multi-step reactions and ensuring atom and charge balance presents formidable challenges. The conventional transformer-based sequence-to-sequence (Seq2Seq) models, commonly employed for sequence generation, prove inadequate in handling the intricate long-term dependencies inherent in CRM [15]. Their limitations become glaring when even a single incorrect character introduced during inference can render the entire CRM meaningless. Beyond the conventional focus on atom and charge balance, ensuring both semantic and syntactic validity of product SMILES strings – representing molecules in a sequence-based format – highlights the necessity for a comprehensive reevaluation at the modeling level in CRM prediction. Recognizing these challenges, we have endeavored to craft an interpretable, swift, and dependable alternative for CRM prediction.

Herein, we propose an interpretable attention-based graph neural network (GNN) model for the elementary reaction step predictions, which are then used to generate the full CRM (Figure 1). We introduce a dataset containing 3 different families of catalytic reactions comprising 7 distinct elementary steps. In addition to the key task of

* Corresponding Author. Email: baranwal.mayank@tcs.com
** Corresponding Author. Email: sunoj@chem.iitb.ac.in
[1] Equal contribution.
[2] Equal contribution.

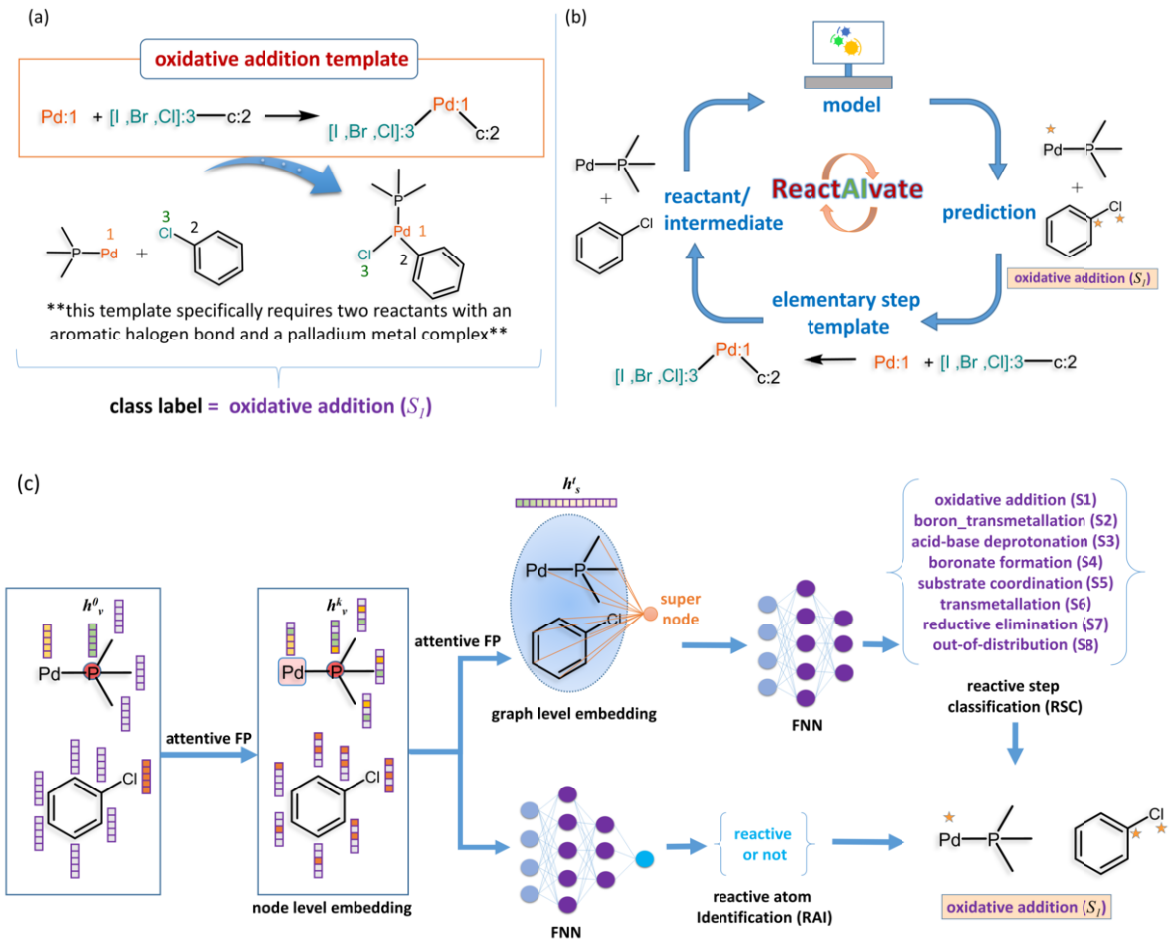

Figure 1. (a) A representative example of oxidative addition template, (b) complete workflow of our proposed ReactAIvate method, (c) process of reaction step classification and reactive atom identification using ReactAIvate

identifying the elementary steps, our model is simultaneously trained to detect the reactive atoms in such steps. In the case of out-of-distribution (OOD) samples, our model points to the reactive atoms, besides classifying them into an unseen group. This can act as a guide to experts for understanding the reactivity. Below we summarize our key contributions:

1. CRM dataset: While there are several benchmark datasets, such as USPTO-50k, comprising single-step chemical reactions, there is currently no existing dataset on CRMs within the available literature as per our knowledge. We curate a first-of-its-kind CRM dataset containing elementary mechanistic steps for transition metal-catalyzed reactions.

2. CRM identification via reaction step classification: We introduce ReactAIvate, a graph-attention-based classification model that precisely identifies the necessary elementary steps for a given combination of reactants, reagents, and catalysts. Our model's accurate intermediate-step classification is pivotal for CRM identification. Notably, our approach focuses on identifying underlying reaction rules rather than generating exact product SMILES, simplifying the problem, as our experiments reveal that SMILES generation can lead to vacuous CRM predictions due to even a single character mismatch.

3. Identification of reactive atoms/groups: Our framework is distinctly trained to minimize a composite of two distinct loss types: (a) Graph-level loss for predicting reaction classes, and (b) Node-level loss for distinguishing reactive and non-reactive atoms. As a result, our curated database includes information on reactive atoms for each specific reaction step within a CRM. This feature is primarily introduced to offer valuable insights to domain experts.

4. Visualizing reactive centers via attention Mechanism: Expanding on the previous point, we demonstrate that the inclusion of node-level loss inherently compels the attention mechanism to align with reactive centers within the molecules involved. This alignment significantly enhances the visualization of reactive centers.

5. Generalization to OOD samples: ReactAIvate avoids overconfidence by introducing an OOD eighth class, enabling accurate identification of scenarios not covered in training data. This enhances trust in predictions and allows for the incorporation of new reaction classes when relevant data becomes available. Attention visualization on OOD samples also reveals potential reaction centers, showcasing the model's adaptability.

2 Related Work

Recent years have witnessed several interesting applications of machine learning in predicting molecular properties as well as their reactions [50, 2, 31]. Intriguing ML algorithms have been developed to make complex chemical problems, such as organic synthesis, increasingly more amenable. In tasks such as forward or retrosynthesis predictions, predefined reaction templates are employed to make an intuitive connection from a set of reactants to product(s). The templates can be obtained by using data-driven approaches [43] or

through encoding by domain experts. Authors in [13] combined these templates and ML for product prediction and subsequent ranking of the products. Transformation rules in retrosynthesis tasks were leveraged in [44]. Current trends suggest the use of templates for both forward and retrosynthesis analysis owing to their efficiency and interpretability. [51, 42, 17, 20].

Interesting alternatives relying on template-free methods have emerged in very recent times. Transformer-based Seq2Seq generative model has been adopted for forward and retrosynthesis predictions using SMILES as the molecular representation [23, 41, 28]. Various models have been proposed that differ in molecular representation and/or model architecture [46, 54, 7, 24, 49]. For instance, authors in [49] included both molecular graph and SMILES representations for retrosynthesis. Although the template-free methods have gained recent attention, they are known to suffer from the generation of invalid SMILES. During translation, a single character addition or a missing one can render the generated reaction invalid.

We note that the current literature although focuses on direct prediction of products and/or retrosynthesis, there are very limited efforts for CRM predictions [37]. In [16], authors utilized deep learning on a rather limited, private dataset of elementary reactions to identify the probable electron sources and sinks and subsequent ranking of these combinations. More recently, Authors in [8] proposed ELECTRO, a graph-based generative model for electron paths movement mimicking 'arrow-pushing' diagrams, but did not consider elementary reaction steps due to the absence of such details in the USPTO-50k dataset. Consequently, ELECTRO is not suitable to multi-step reactions.

3 Preliminaries and Proposed Method

3.1 Background

Chemical reaction. A chemical reaction is a process in which one set of substances (reactants) transforms into another set of substances (products).

Elementary step. The elementary steps in a chemical reaction typically consists of bond breaking or bond forming involving the reactants/intermediates, where an electron rich source gets attached to an electron poor sink. The movement of electrons is usually denoted using 'arrow-pushing diagrams' where the arrow direction is from the electron source to the sink.

Chemical reaction mechanism (CRM). A sequence of elementary steps that describe the transformation of reactant $\mathbf{R}$ to product $\mathbf{P}$, through several transient intermediates $\mathbf{I}$, constitutes a full CRM as shown in eqn(1).

$$\mathbf{R} \xrightarrow{stepS_1} I_1 \xrightarrow{stepS_2} I_2, \cdots, I_n \xrightarrow{stepS_n} \mathbf{P} \tag{1}$$

where $S_{1:n}$ represents individual elementary steps.

Reactive atom. Reactive atoms are the active atoms involved in an elementary step that undergoes a large change in their immediate bonding/valency environment as a result of the reaction.

Reaction templates. Reaction templates are defined as predetermined sets of chemical transformation rules with specific constraints, such as the presence of a particular substructure. [27] In Figure 1a, an illustration of a template is provided using oxidative addition as a representative elementary step in a reaction. The specified constraints for this template are that any eligible reactant should possess a substructure featuring an aryl C-X bond (where X = Cl, Br, I) and that a catalytically active palladium (Pd) metal center be present.

3.2 CRM prediction via ReactAIvate

We develop an interpretable GNN model, ReactAIvate, to predict the elementary steps, which is further used to devise CRM. An extensive dataset comprising of seven such mechanistic steps for transition metal-catalyzed reactions is curated (see Section 4.1 for further details). ReactAIvate is build upon two elementary tasks:

Reaction step classification (RSC). The primary task of ReactAIvate is to predict the correct elementary step among the seven identified elementary steps. These mechanistic steps include 'oxidative addition', 'boron transmetallation', 'acid-base deprotonation', 'boronate formation', 'substrate coordination', 'transmetallation', and 'reductive elimination'.[26] Once the correct mechanistic step is classified, an off-the-shelf template based reaction rules are used to predict the product information accurately. The predicted products form the reactants for the next step, and the subsequent mechanistic step is identified again. The process is repeated until the catalyst is regenerated. Our advantage in CRM prediction stems from our focus on mechanistic step classification, providing a distinct edge over traditional sequence-based modeling approaches. More specifically, given the set of reactants $\mathbf{R}$, or intermediates I_1, I_2, ..,I_n, ReactAIvate predicts labels denoted as $\{S_i\}_{i=1}^{7}$, where each S_i corresponds to one of the seven mechanistic steps (Figure 1c).

However, while the dataset consists of the seven elementary steps, it doesn't encompass the entire range of mechanistic rules that a set of reactants may undergo in a chemical reaction. To ensure ReactAIvate doesn't erroneously predict reactions following different chemical transformation rules or force chemically non-reactive combinations into the seven elementary steps, we introduce an eighth class, denoted as S_8. Our framework is trained to classify any out-of-distribution (OOD) samples into this eighth class, bolstering confidence in ReactAIvate's predictions.

Reactive atom identification (RAI). Reactive step classification alone falls short in revealing the fundamental mechanisms and rationales underlying a CRM. Accurate identification of reaction centers within the reactant molecules is important for forward synthesis. This approach is particularly crucial when dealing with OOD samples, where the identification of reactive atoms within a given reaction class holds significant value. These insights serve as a guide for experts in recognizing feasible elementary reactions. In this work, the labels for reactive atom classification are derived from reaction templates, with '1' indicating a reactive atom and '0' indicating a non-reactive atom.

3.2.1 ReactAIvate Workflow

We now proceed to outline the operational methodology of our framework, ReactAIvate.

1. Molecular representaion via graphs A molecule can be portrayed as a graph, where atoms and bonds constitute nodes and edges, respectively [32]. Each unique atom is represented by the following set of features, encompassing nine types such as atom symbol, formal charge, hybridization, aromaticity, etc. These collectively result in a total of 39 atom features (see supporting information for further details [21]). Similarly, edge representation includes feature vectors corresponding to different bond types (single, double, triple, etc.).

In our graph representation, $G = (V, E)$, where V is the set of atoms and E is the set of edges, each atom ν is associated with a feature vector X_ν in $\mathbb{R}^D$, with $D = 39$ representing the number of features for each atom. The tasks at hand involve: (1) classifying elementary steps into predefined classes. Given an input graph

G consisting of reactant molecules, the goal is to learn the representation vector h_G^* and a linear function g_1 such that the predicted step $\hat{y}_G := g_1(h_G^*)$ aligns with the true step $y_G \in \{S_i\} i = 1^8$; (2) identifying reactive atoms, where each atom ν in V has a label y_ν ($y_\nu \in [0, 1]$). The objective is to learn the representation vector h_ν^* and a linear function g_2 for all ν such that the predicted reactivity $\hat{y}_\nu := g_2(h_\nu^*)$ aligns with the true binary label y_ν.

2. Graph attention network for RSC & RAI After encoding reactant molecules as graphs, we introduce an attention mechanism to these graphs by generating a context vector for a target atom (ν) through attention on its neighboring atoms u, where $u \in \mathcal{N}(\nu)$. The initial step in computing the context vector involves determining attention weights $\alpha_{u\nu}$ between the state vectors h_ν and h_u of the two atoms [48, 53]. Subsequently, a context operation follows, where a linear transformation is applied to h_u, the state vectors of neighboring atoms. This is succeeded by a weighted sum and a non-linear activation function, resulting in C_ν, the context vector for the target atom ν. The calculation of normalized attention coefficients can be formulated as follows:

$$\alpha_{u\nu} = \frac{\exp\left(\texttt{LeakyReLU}\left(\mathbf{W}[h_\nu, h_u]\right)\right)}{\sum_{u\in\mathcal{N}(\nu)} \exp\left(\texttt{LeakyReLU}\left(\mathbf{W}[h_\nu, h_u]\right)\right)},$$

where $\alpha_{u\nu}$ signifies the importance (weight) of neighbor atom u to target atom ν, and $\mathbf{W}$ is a trainable weight matrix, with

$$C_\nu = \texttt{ELU}\left(\sum_{u\in\mathcal{N}(\nu)} \alpha_{u\nu}.\mathbf{W}.h_u\right).$$

Modern GNNs adopt neighborhood aggregation strategies, updating the features of the target atom iteratively by incorporating the features of its neighboring atoms. The atom's representation encapsulates the structural information within the k-hop network around it after k iterations of this aggregation process. This strategy can be formulated as:

Aggregation phase

$$C_\nu^{k-1} = \sum_{u\in\mathcal{N}(\nu)} A^{k-1}(h_u^{k-1}, h_\nu^{k-1}),$$

Update phase

$$h_\nu^k = \texttt{GRU}^{k-1}(C_\nu^{k-1}, h_\nu^{k-1}),$$

where, h_ν^k represents the feature vector of atom ν after the kth layer, with $h_\nu^0 = X_\nu$ (see Figure 1c). In the aggregation phase, the graph attention mechanism, A^{k-1}, provides the most relevant information to the target atom from its neighborhoods in the form of the context vector C_ν^{k-1}. In the subsequent step, the update function $\texttt{GRU}^{k-1}$ (gated recurrent unit) takes the attention context and the previous state vector of the target atom as input to update the feature vector from the previous state h_ν^{k-1} to the current state h_ν^k.

To generate a graph-level embedding, a virtual *supernode* is introduced, connecting with all the atoms in the molecular graph. The feature vector for this supernode is obtained using sum pooling, expressed as $h_s^0 = \sum_\nu h_\nu^k$. Through a similar neighborhood aggregation method, a high-level graph embedding h_s^t is learned iteratively over t iterations. At this point, the optimally learned vector h_s^t can be considered equivalent to h_G^* (see Figure 1c), which is subsequently utilized for the elementary step prediction task. By feeding the graph-level embedding h_G^* into a feed-forward neural network (FNN), predicted values are obtained for the RSC task as:

$$\hat{y}_G = \texttt{FNN}(h_G^*).$$

ReactAIvate employs cross-entropy loss between the predicted and true labels for the RSC task:

$$\mathcal{L}_{class}^G := \texttt{CrossEntropyLoss}(\hat{y}_G, y_G).$$

In our RAI task, the objective is to classify each atom in a molecule as reactive or non-reactive, enabling the use of standard classification loss functions. ReactAIvate employs binary cross-entropy (BCE) loss for two-way classification, with a slight modification. Given that the majority of atoms in a sample are non-reactive, we modify the BCE loss by introducing a weight that strongly penalizes incorrect predictions for atoms that are genuinely reactive. This adjustment helps prevent the model from becoming biased towards predicting all atoms as non-reactive (see supporting information [21]).

Recall that for an atom ν, the associated embedding vector before the sum pooling is denoted as h_ν^k. This embedding vector is considered the optimal atom representation h_ν^*. In the node-level classification task, the updated node feature vector is passed through the FNN to predict whether the atom is reactive or non-reactive, expressed as:

$$\hat{y}_\nu = \texttt{FNN}(h_\nu^*).$$

For RAI, we consider the weighted BCE loss as discussed above:

$$\mathcal{L}_{RC}^G := \sum_\nu \texttt{WeightedBCELoss}(\hat{y}_\nu, y_\nu)$$

The overall loss for ReactAIvate is made up of the individual losses for RSC and RAI:

$$\mathcal{L} = \mathcal{L}_{class}^G + \mathcal{L}_{RC}^G$$

4 Experiments

4.1 Dataset details

In this study, we consider a diverse and representative set of transition metal-catalyzed reactions. These include Suzuki-Miyaura coupling (SMC) [30, 4], Buchwald-Hartwig amination (BHA) [35, 14], and Kumada coupling (KC) [1]. The inclusion of these reactions is motivated by their significance in drugs, agrochemicals, and pharmaceutical synthesis. To the best of our knowledge, there is currently no existing database that includes credible mechanisms for these reactions, compelling the creation of CRM datasets. For this, seven distinct elementary mechanistic steps are recognized that can account for all three reaction mechanisms: 'oxidative addition', 'boron transmetallation', 'acid-base deprotonation', 'boronate formation', 'substrate coordination', 'transmetallation', and 'reductive elimination' (see supporting information [21]).

We have created essential reaction templates for each elementary mechanistic step within the three considered reaction datasets in this study. The template is structured to yield the product based on the given reactants (refer to Figure 1a). A reaction comprises a set of substrates, catalyst and reagents. For example, in the case of SMC, key substrates include an aryl halide and a boronic acid, with a metal-phosphine complex serving as the catalyst and a base facilitating the reaction. These reaction partners are curated from primary literature [52, 22]and the PubChem database [24]. By inputting this set of molecules into the reaction template aligned with the specific reaction mechanism type, all elementary steps and the complete CRM can be derived. Our dataset encompasses a total of 100,000 elementary mechanistic steps (see supporting information for more details [21]).

4.2 Training details

The dataset is partitioned into training, validation, and test samples with a distribution ratio of 70:10:20. ReactAIvate is constructed using PyTorch [33] with the Adam optimizer [25] and a batch size of 256. Throughout this study, we maintain consistent hyperparameter values: k (number of attentive message passing layers for atom embedding) is set to 2, t (number of attentive message passing layers for molecule embedding) is set to 1, with an $L2$ weight decay of 0.000001, a learning rate of 0.001, and a dropout rate of 0.1. The number of atom features and graph feature size are specified as 39 and 200, respectively.

4.3 Baseline

First, we employed Seq2Seq approaches, such as T5Chem [29] and Transformer [36] as proxies for baselines, which are considered state-of-the-art (SOTA) for individual single step reaction prediction. T5Chem is a multi-tasking model designed for various reaction prediction tasks, leveraging the Text-To-Text Transfer Transformer (T5), an encoder-decoder model from the transformer family. The Transformer baseline is built on the original encoder-decoder framework [47].

In these frameworks, the problem is formulated as a generation task, where the model is trained to predict the full CRM given the reactants and other entities. To ensure a fair comparison, we introduce two test datasets, each comprising 1000 samples. The first is an in-distribution (ID) test dataset, where the individual reacting partners of each sample belong to the same set as the training datasets, although the samples themselves are not part of the training set. The second is an out-of-distribution (OOD) dataset, where the reaction components of each sample differ structurally from those used in training the model.

Model	Accuracy (%) ID test set	Accuracy (%) OOD test set	Train time (m)
Seq2Seq			
T5Chem	95.60 ± 0.01	0.07 ± 0.01	180
T5Chem(FI)	98.40 ± 0.01	0.11 ± 0.02	45
Transformer	11.10 ± 0.03	0.01 ± 0.00	150
Transformer(FI)	91.80 ± 0.05	0.07 ± 0.00	60
Featurization			
Morgan(r=2)	100.00 ± 0.00	31.50 ± 3.25	20
MFF	100.00 ± 0.00	48.76 ± 3.60	30
PCD	100.00 ± 0.00	48.57 ± 2.67	10
Graph			
MPNN	100.00 ± 0.00	52.80 ± 9.50	10
ReactAIvate	$\mathbf{100.00 \pm 0.00}$	$\mathbf{95.70 \pm 0.34}$	**15**

Table 1. Performance of different models for CRM prediction on ID and OOD test molecules. Training time is reported in minutes (m). **Bold** indicates best performance.

Next, we developed several DNN models as baselines that are not based on sequence data such as in seq2seq models. These models use different feature representations, including fingerprint (Morgan and multiple fingerprint feature (MFF)) [38], and rdkit-based physicochemical descriptors (PCD)) [34]. Their design targets the same objective as ReactAIvate. Additionally, we considered a Message-Passing Neural Network (MPNN) model as another benchmark [18]. It operates by passing messages between nodes (atoms) in a graph (molecule) to learn the features and interactions. More details about baseline models are provided in the supporting information [21].

4.4 Results

Table 1 illustrates the performance comparison, specifically focusing on accuracy, i.e., the percentage of samples where the generated CRM precisely matches the true CRM. For the ID dataset, ReactAIvate exhibits better accuracy compared to T5Chem and Transformer. Intriguingly, in the OOD dataset, ReactAIvate offers superior accuracy in CRM generation, reaching upto 96%. In contrast, both the baselines fail to predict the correct CRM, potentially due to incorrect predictions of one of the many tokens or atom imbalance, rendering the entire sequence invalid (see supporting information for more details [21]). The observation is not specific to CRM prediction, but has also been reported in recent findings, where several SOTA template-free models unexpectedly falter when faced with OOD in retrosynthesis prediction [11]. In addition, unlike the above-mentioned baselines, where T5Chem and Transformer-based architectures are trained to predict the entire CRM, we have conducted additional experiments where the baselines are trained to predict just the forward intermediates (FIs) at each step (exactly what these models were originally designed for), and then the predicted intermediates serve as the reactants for the next step. The process is repeated until the catalyst is regenerated (i.e., end of CRM). The results are reported above in Table 1. Even for the FI variants of the baselines, their performances remain highly inadequate for OOD samples. Although T5Chem benefits from pre-training on the Pubchem database containing **97M molecules**, its sequence-centric framework limits its effectiveness in CRM prediction, especially with OOD instances.

The poor OOD performance of seq2seq/transformer models is not due to poor hyperparameter tuning but rather the nature of character generation-based product prediction. For example, if a model is trained on molecules with Iodine and Chlorine but tested on a molecule where Iodine is replaced with Fluorine, it struggles because it was never trained to generate sequences involving Fluorine. Additionally, any incorrect symbol prediction in an intermediate step can corrupt the entire downstream task for seq2seq models. On the contrary, ReactAIvate predicts the most appropriate reaction rule (template) at each step, using these rules to accurately generate the right products at each step.

The fingerprints and PCD based DNN models exhibited enhanced performance as compared to sequence-based models on both ID and OOD samples. However, it is worth noting that these models yield lower accuracy for the OOD set as compared to ReactAIvate, this discrepancy may stem from the fact that fingerprints rely on predefined sets of molecular substructures, potentially limiting their ability to capture nuanced structural details. Whereas MPNN performed better than fingerprints and PCD featurization, they still fell short of ReactAIvate. This could be attributed to the advantages conferred by attention mechanisms and the incorporation of RAI task within ReactAIvate. This, in turn, implicitly captures the relevance between CRM and RAI, and also helps with the alignment of highly attentive atoms with likely reaction centers. Nonetheless, these models provide robust and adequate validation. These results strongly imply the necessity for a meticulous reconsideration of the modeling aspect for CRM generation.

Task	**Accuracy** (%)	**F1-Score**
ID-RSC	100.00 ± 0.00	1.00 ± 0.00
ID-RAI	96.40 ± 0.30	0.87 ± 0.01
OOD-RSC	98.60 ± 0.23	0.98 ± 0.01
OOD-RAI	94.86 ± 0.40	0.85 ± 0.02

Table 2. Performance of ReactAIvate for RSC and RAI tasks on ID and OOD test molecules.

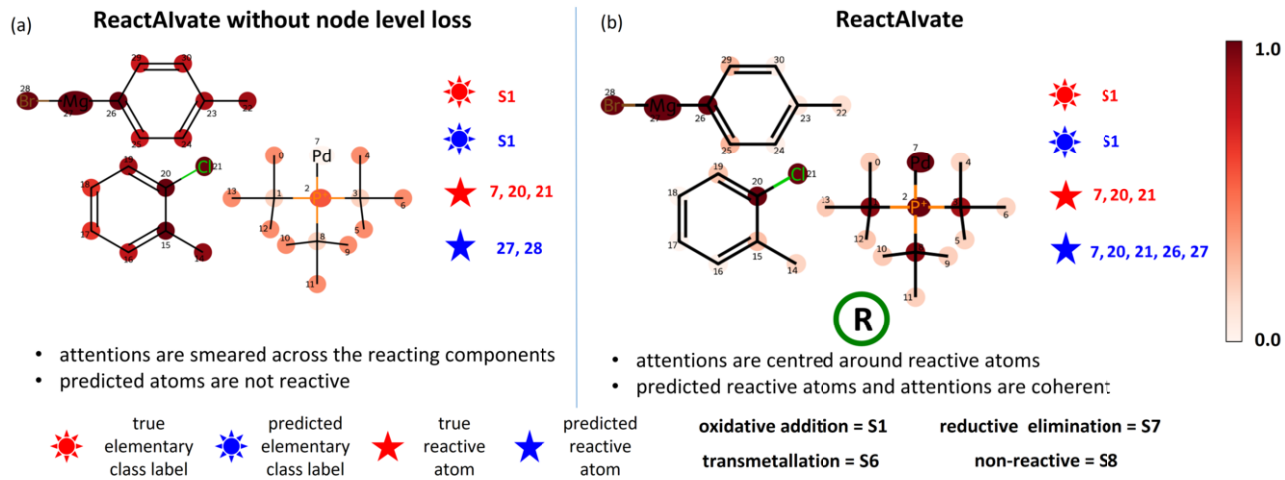

Figure 2. Effect of the inclusion of node-level loss in ReactAIvate demonstrated through attention visualization. The rightmost bar represents min-max rescaled attention values.

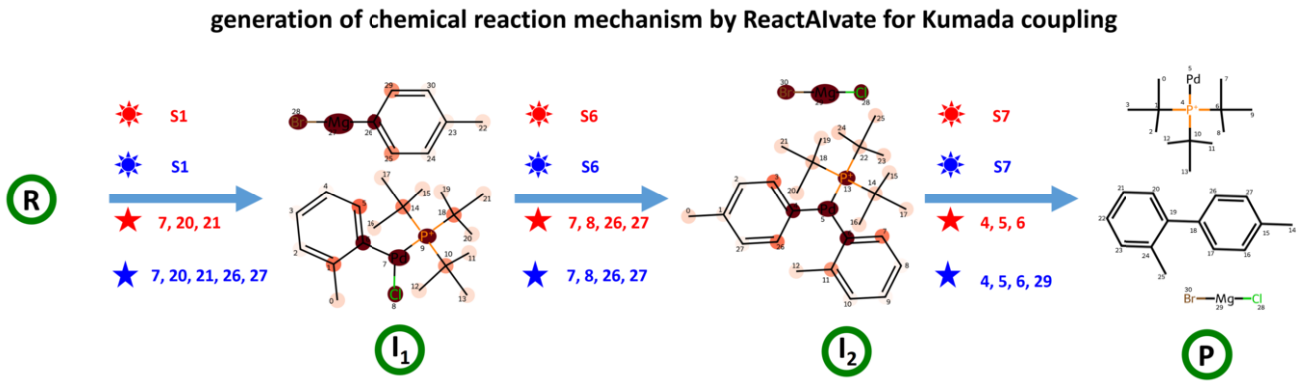

Figure 3. An illustration of the sequential generation of the full CRM for the Kumada coupling reaction

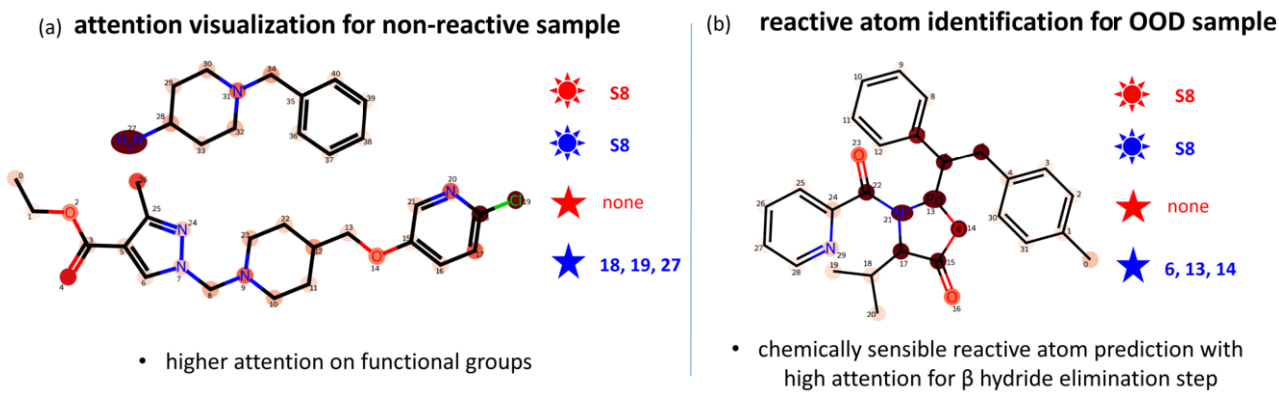

Figure 4. Attention visualization for a sample in (a) non-reactive, (b) reactive out-of-distribution (OOD) set.

We delve deeper into ReactAIvate's capability for accurately predicting RSC and RAI concurrently. The classification accuracies and F1-scores for ID and OOD reactants are presented in Table 2. Beyond accurately classifying the correct elementary step, ReactAIvate demonstrates proficiency in distinguishing between reactive and non-reactive atoms. Notably, even in the OOD dataset, ReactAIvate maintains high efficacy for both RSC and RAI tasks. This indicates that the model has a broad applicability domain and holds potential utility for domain experts.

5 Discussion and Analysis

Having demonstrated the exceptional performance of our framework ReactAIvate for both the RSC and RAI tasks, we proceed to delve into the underlying reasons for its effectiveness through a comprehensive ablation study. Additionally, we aim to elucidate domain-specific advantages that contribute to its superior performance.

Significance of incorporating node-level RAI: We begin by emphasizing the implication of integrating the node-level loss in the

reactive atom prediction task along with the graph-level loss for elementary step prediction. In Figure 2, we compare the attention visualization between ReactAIvate without the node-level loss and ReactAIvate with both losses, using an oxidative addition step from the Kumada coupling dataset. In the former model, attention weights are dispersed across the reactants and catalyst, providing limited utility. Notably, the crucial reactive metal atom Pd:7, responsible for catalyzing this step, fails to draw any attention. Thus, attention visualization without the node-level loss contributes little to understanding chemical reactivity. In the latter model with the two-level loss, attention weights are more concentrated around the reactive region of the molecules. For instance, the Pd metal center and the aryl chlorine bond garner higher attention, (Pd:7, C:20, Cl:21), representing the true reactive atoms in this elementary step. This underscores the importance of including the node-level loss, aligning the model's attention mechanism more closely with how a chemist would assign attention to reactive atoms.

Illustration of a full CRM: Moving forward, we assess ReactAIvate's capability in predicting a complete CRM for Kumada coupling, serving as a representative example (Figure 3). In the initial step, the model accurately predicts oxidative addition (S_1) as the elementary step, along with the correct identification of reactive atoms. Notably, the model also predicts two additional atoms as active, which intriguingly turn out to be reactive in the subsequent step. In the second step, ReactAIvate once again correctly predicts the elementary class, transmetallation (S_6), and accurately identifies all true reactive atoms. Finally, the model predicts the reductive elimination (S_7) step as the concluding phase. Both the predicted active atoms and attention distributions align consistently with the expected mechanism. In summary, ReactAIvate demonstrates the ability to generate a complete CRM starting from only reactants and catalysts.

Attention visualization for OOD-class: To gain deeper insights into the model, we visualize attentions in a sample belonging to the OOD class, as shown in Figure 4a. In this instance, the pair of molecules is predicted to fall into the eighth (OOD) class, accurately reflecting that the amine and the aryl halide (N:27, C:18, Cl:19) would not react in the absence of a catalyst. Intriguingly, the model predicts chlorine in one molecule and the adjacent carbon in the other as reactive atoms. Additionally, attentions are primarily distributed around functional groups. The broad dispersion of attention underscores ReactAIvate's challenge in pinpointing the exact reaction mechanism. This outcome aligns with expectations, as the combination is chemically non-reactive, and it would be counterintuitive for the model to highlight specific reaction centers leading to its prediction in the OOD class.

In our final evaluation, we test ReactAIvate for the identification of potential reactive atoms in an OOD sample involving an entirely different reaction mechanism, as depicted in Figure 4b. Please note that the molecule is not expected to undergo any transformation (both predicted and true class are S_8, i.e., no reaction). However, if it is presented with a suitable reagent, this molecule is anticipated to undergo a β hydride elimination step and the reactive centers get activated. Among the predicted reactive atoms, two (Pd:13 and α-C: 6) correspond to possible true reactive atoms. Since the molecule in this example doesn't undergo reaction, the true reactive centers are presented as an empty list. Moreover, attentions are dispersed around the reaction center and the reactive atoms. These highlights the model's versatility in predicting OOD samples and serves as a guide for comprehending the reactivity of metal-catalyzed reactions. Consequently, one can easily incorporate new elementary classes of interest to further broaden the model's applicability.

6 Conclusion and Future Work

In conclusion, we introduce ReactAIvate, a graph-attention-based Graph Neural Network (GNN) model designed for interpretable Chemical Reaction Mechanism (CRM) generation. Our model is trained on a novel dataset comprising seven distinct elementary mechanistic steps, covering the complete CRM for three different transition-metal-catalyzed processes. ReactAIvate excels in accurately classifying elementary steps and recognizing reactive atoms, demonstrating its capability to construct full CRMs. Notably, the model exhibits a prudent handling of non-reactive cases, showcasing its reliability in predictions. ReactAIvate outperforms Seq2Seq baseline models, emphasizing the limitations of the latter in CRM identification due to minor errors. The robust OOD classification performance underscores ReactAIvate's potential for exploring additional mechanisms with the availability of more data. As part of future work, we plan create a user-friendly interface for predicting entire CRMs based on user-inputted SMILES of reactants.

7 Data Availability

Data and codes related to this work are publicly available through our Github repository at https://github.com/alhqlearn/ReactAIvate.

Acknowledgements

We gratefully acknowledge the generous computing time provided by the SpaceTime supercomputing facility at IIT Bombay. M.D. expresses gratitude for the Prime Minister's Research Fellowship.

References

[1] L. Ackermann and A. Althammer. Air-stable pinp (o) h as preligand for palladium-catalyzed kumada couplings of unactivated tosylates. *Organic Letters*, 8(16):3457–3460, 2006.

[2] D. T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher, and A. G. Doyle. Predicting reaction performance in c–n cross-coupling using machine learning. *Science*, 360(6385):186–190, 2018.

[3] S. Bahmanyar, K. Houk, H. J. Martin, and B. List. Quantum mechanical predictions of the stereoselectivities of proline-catalyzed asymmetric intermolecular aldol reactions. *Journal of the American Chemical Society*, 125(9):2475–2479, 2003.

[4] I. P. Beletskaya, F. Alonso, and V. Tyurin. The suzuki-miyaura reaction after the nobel prize. *Coordination Chemistry Reviews*, 385:137–173, 2019.

[5] T. Benzinger. Thermodynamics, chemical reactions and molecular biology. *Nature*, 229(5280):100–102, 1971.

[6] T. Bhattacharya, S. Ghosh, S. Dutta, S. Guin, A. Ghosh, H. Ge, R. B. Sunoj, and D. Maiti. Combinatorial ligand assisted simultaneous control of axial and central chirality in highly stereoselective c- h allylation. *Angewandte Chemie*, page e202310112, 2023.

[7] J. Born and M. Manica. Trends in deep learning for property-driven drug design. *Current medicinal chemistry*, 28(38):7862–7886, 2021.

[8] J. Bradshaw, M. J. Kusner, B. Paige, M. H. Segler, and J. M. Hernández-Lobato. A generative model for electron paths. *arXiv preprint arXiv:1805.10970*, 2018.

[9] D. G. Brown and J. Bostrom. Analysis of past and present synthetic methodologies on medicinal chemistry: where have all the new reactions gone? miniperspective. *Journal of medicinal chemistry*, 59(10): 4443–4458, 2016.

[10] S. Chen and Y. Jung. A generalized-template-based graph neural network for accurate organic reactivity prediction. *Nature Machine Intelligence*, 4(9):772–780, 2022.

[11] S. Chen and Y. Jung. Assessing the extrapolation capability of template-free retrosynthesis models. *arXiv preprint arXiv:2403.03960*, 2024.

[12] Y. P. Chin, N. W. See, I. D. Jenkins, and E. H. Krenske. Computational discoveries of reaction mechanisms: recent highlights and emerging challenges. *Organic & Biomolecular Chemistry*, 20(10):2028–2042, 2022.

[13] C. W. Coley, R. Barzilay, T. S. Jaakkola, W. H. Green, and K. F. Jensen. Prediction of organic reaction outcomes using machine learning. *ACS central science*, 3(5):434–443, 2017.

[14] R. Dorel, C. P. Grugel, and A. M. Haydl. The buchwald–hartwig amination after 25 years. *Angewandte Chemie International Edition*, 58 (48):17118–17129, 2019.

[15] D. Flam-Shepherd, K. Zhu, and A. Aspuru-Guzik. Language models can learn complex molecular distributions. *Nature Communications*, 13 (1):3293, 2022.

[16] D. Fooshee, A. Mood, E. Gutman, M. Tavakoli, G. Urban, F. Liu, N. Huynh, D. Van Vranken, and P. Baldi. Deep learning for chemical reaction prediction. *Molecular Systems Design & Engineering*, 3 (3):442–452, 2018.

[17] S. Genheden, A. Thakkar, V. Chadimová, J.-L. Reymond, O. Engkvist, and E. Bjerrum. Aizynthfinder: a fast, robust and flexible open-source software for retrosynthetic planning. *Journal of cheminformatics*, 12 (1):70, 2020.

[18] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pages 1263–1272. PMLR, 2017.

[19] E. Goldstein, B. Beno, and K. Houk. Density functional theory prediction of the relative energies and isotope effects for the concerted and stepwise mechanisms of the diels- alder reaction of butadiene and ethylene. *Journal of the American Chemical Society*, 118(25):6036–6043, 1996.

[20] E. Heid, J. Liu, A. Aude, and W. H. Green. Influence of template size, canonicalization, and exclusivity for retrosynthesis and reaction prediction applications. *Journal of Chemical Information and Modeling*, 62 (1):16–26, 2021.

[21] A. Hoque, M. Das, M. Baranwal, and R. B. Sunoj. Reactaivate: A deep learning approach to predicting reaction mechanisms and unmasking reactivity hotspots, 2024. URL https://arxiv.org/abs/2407.10090.

[22] X. Huang, K. W. Anderson, D. Zim, L. Jiang, A. Klapars, and S. L. Buchwald. Expanding pd-catalyzed c- n bond-forming processes: the first amidation of aryl sulfonates, aqueous amination, and complementarity with cu-catalyzed reactions. *Journal of the American Chemical Society*, 125(22):6653–6655, 2003.

[23] W. Jin, C. Coley, R. Barzilay, and T. Jaakkola. Predicting organic reaction outcomes with weisfeiler-lehman network. *Advances in neural information processing systems*, 30, 2017.

[24] E. Kim, D. Lee, Y. Kwon, M. S. Park, and Y.-S. Choi. Valid, plausible, and diverse retrosynthesis using tied two-way transformers with latent variables. *Journal of Chemical Information and Modeling*, 61(1):123–133, 2021.

[25] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[26] H. Kurosawa and A. Yamamoto. *Fundamentals of molecular catalysis*. Elsevier, 2003.

[27] G. Landrum. Rdkit documentation. *Release*, 1(1-79):4, 2013.

[28] K. Lin, Y. Xu, J. Pei, and L. Lai. Automatic retrosynthetic route planning using template-free models. *Chemical science*, 11(12):3355–3364, 2020.

[29] J. Lu and Y. Zhang. Unified deep learning model for multitask reaction predictions with explanation. *Journal of Chemical Information and Modeling*, 62(6):1376–1387, 2022.

[30] R. Martin and S. L. Buchwald. Palladium-catalyzed suzuki- miyaura cross-coupling reactions employing dialkylbiaryl phosphine ligands. *Accounts of chemical research*, 41(11):1461–1473, 2008.

[31] Z. Meng, P. Zhao, Y. Yu, and I. King. Doubly stochastic graph-based non-autoregressive reaction prediction. *In IJCAI*, 2023.

[32] Z. Meng, P. Zhao, Y. Yu, and I. King. A unified view of deep learning for reaction and retrosynthesis prediction: Current status and future challenges. *In IJCAI*, 2023.

[33] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[34] S. Riniker and G. A. Landrum. Open-source platform to benchmark fingerprints for ligand-based virtual screening. *Journal of cheminformatics*, 5(1):26, 2013.

[35] P. Ruiz-Castillo and S. L. Buchwald. Applications of palladium-catalyzed c–n cross-coupling reactions. *Chemical reviews*, 116(19): 12564–12649, 2016.

[36] A. M. Rush. The annotated transformer. In *Proceedings of workshop for NLP open source software (NLP-OSS)*, pages 52–60, 2018.

[37] M. Sakai, M. Kaneshige, and K. Yasuda. Learning organo-transition metal catalyzed reactions by graph neural networks. *Journal of Computational Chemistry*, 2023.

[38] F. Sandfort, F. Strieth-Kalthoff, M. Kühnemund, C. Beecks, and F. Glorius. A structure-based platform for predicting chemical reactivity. *Chem*, 6(6):1379–1390, 2020.

[39] N. Schneider, N. Stiefl, and G. A. Landrum. What's what: The (nearly) definitive guide to reaction role assignment. *Journal of chemical information and modeling*, 56(12):2336–2346, 2016.

[40] M. Schupp, I. Saridakis, D. Kaiser, and N. Maulide. Chemical synthesis as a discovery platform in immunosuppression and determination of mode of action. *Nature Synthesis*, pages 1–13, 2024.

[41] P. Schwaller, T. Laino, T. Gaudin, P. Bolgar, C. A. Hunter, C. Bekas, and A. A. Lee. Molecular transformer: a model for uncertainty-calibrated chemical reaction prediction. *ACS central science*, 5(9):1572–1583, 2019.

[42] M. H. Segler and M. P. Waller. Modelling chemical reasoning to predict and invent reactions. *Chemistry–A European Journal*, 23(25):6118–6128, 2017.

[43] M. H. Segler and M. P. Waller. Neural-symbolic machine learning for retrosynthesis and reaction prediction. *Chemistry–A European Journal*, 23(25):5966–5971, 2017.

[44] M. H. Segler, M. Preuss, and M. P. Waller. Planning chemical syntheses with deep neural networks and symbolic ai. *Nature*, 555(7698):604–610, 2018.

[45] M. Tavakoli, Y. T. T. Chiu, A. Shmakov, A. M. Carlton, D. Van Vranken, and P. Baldi. Ai for interpretable chemistry: Predicting radical mechanistic pathways via contrastive learning. *arXiv preprint arXiv:2311.01118*, 2023.

[46] U. V. Ucak, I. Ashyrmamatov, J. Ko, and J. Lee. Retrosynthetic reaction pathway prediction through neural machine translation of atomic environments. *Nature communications*, 13(1):1186, 2022.

[47] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[48] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

[49] Y. Wan, B. Liao, C.-Y. Hsieh, and S. Zhang. Retroformer: Pushing the limits of interpretable end-to-end retrosynthesis transformer. *arXiv preprint arXiv:2201.12475*, 2022.

[50] Y. Wang, Y. Gu, C. Lou, Y. Gong, Z. Wu, W. Li, Y. Tang, and G. Liu. A multitask gnn-based interpretable model for discovery of selective jak inhibitors. *Journal of cheminformatics*, 14(1):16, 2022.

[51] J. N. Wei, D. Duvenaud, and A. Aspuru-Guzik. Neural networks for the prediction of organic chemistry reactions. *ACS central science*, 2(10): 725–732, 2016.

[52] J. P. Wolfe and S. L. Buchwald. Improved functional group compatibility in the palladium-catalyzed amination of aryl bromides. *Tetrahedron Letters*, 38(36):6359–6362, 1997.

[53] Z. Xiong, D. Wang, X. Liu, F. Zhong, X. Wan, X. Li, Z. Li, X. Luo, K. Chen, H. Jiang, et al. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *Journal of medicinal chemistry*, 63(16):8749–8760, 2019.

[54] S. Zheng, J. Rao, Z. Zhang, J. Xu, and Y. Yang. Predicting retrosynthetic reactions using self-corrected transformer neural networks. *Journal of chemical information and modeling*, 60(1):47–55, 2019.