CJC Chinese Journal of Chemistry 中国化学 - An International Journal

For submission: https://wiley.atyponrex.com/journal/CJOC
For published articles: https://onlinelibrary.wiley.com/journal/16147065

Concise Report

Cite this paper: *Chin. J. Chem.* **2025**, *43*, 1230—1238. **DOI: 10.1002/cjoc.202401186**

Multi-modal Homogeneous Chemical Reaction Performance Prediction with Graph and Chemical Language Information

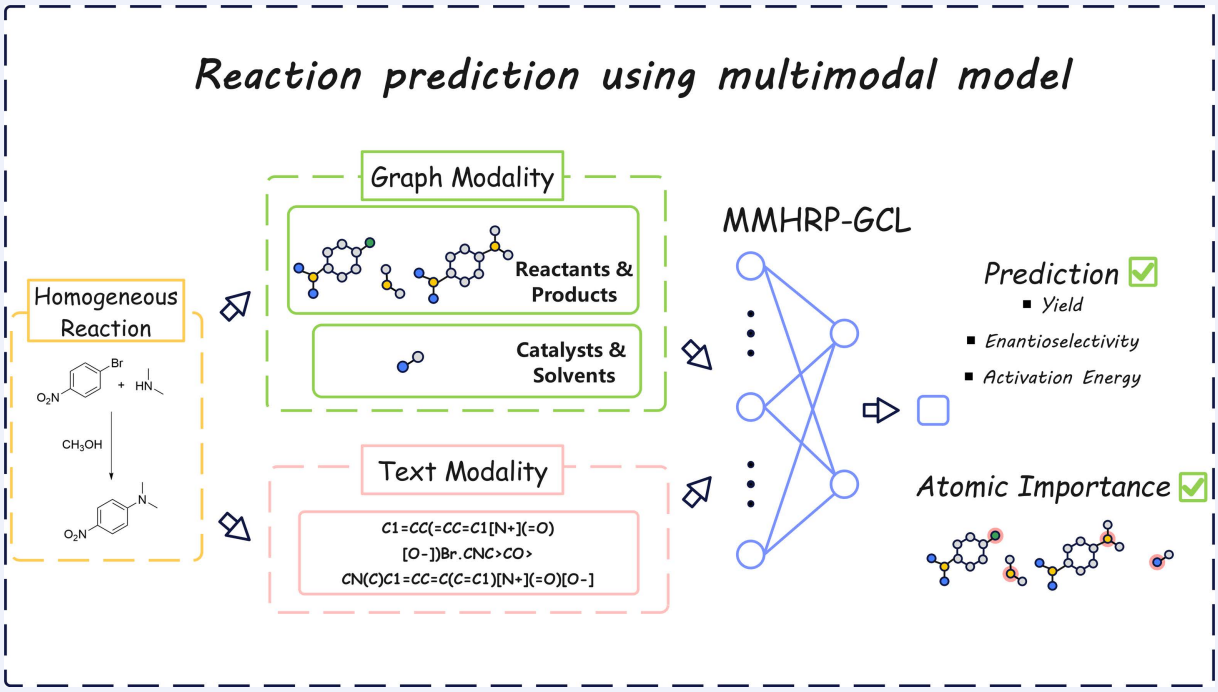
Shen Wang,[a,b] Weiren Zhao,[a,c] Yining Liu,[a,c] and Yang Li*[,a,c]

[a] *State Key Laboratory of Fine Chemicals, Dalian University of Technology, Dalian, Liaoning 116024, China*
[b] *Leicester International Institute, Dalian University of Technology, Panjin, Liaoning 124221, China*
[c] *School of Chemical Engineering, Ocean and Life Sciences, Dalian University of Technology, Panjin, Liaoning 124221, China*

Keywords

Reaction prediction | Deep learning | Multimodal model | Interpretable | Computational chemistry | Molecular modeling | Homogeneous catalysis

Comprehensive Summary



Accurate prediction for chemical reaction performance offers optimal direction for synthetic development. To this end, we present a novel multi-modal model called MMHRP-GCL to achieve the prediction of homogeneous chemical reaction yield, enantioselectivity, and activation energy by fusing the information from the text and graph modalities, requiring only 8 simple descriptors and Reaction SMILES obtained without high-cost DFT computation, and capable of managing reactions involving a fluctuating number of molecules. Experimental results on 4 datasets show that MMHRP-GCL outperforms at least 7 generalized SOTA methods. Ablation study confirms the critical roles of the complementation of graph and text modalities, as well as the significance of modality alignment and atomic features in prediction. Albeit there is still room for improvement in the interpretation of atomic relationships, the model has a remarkable ability to identify important atoms. A statistically interpretable study of the feature importance and a test on challenging dataset further demonstrates the utility and potential of the model. As a high-accuracy, low-cost, interpretable, and general multi-modal model, MMHRP-GCL provides valuable guidance on the design of forward predictors for homogeneous catalytic reactions.

*E-mail: chyangli@dlut.edu.cn

Background and Originality Content

Homogeneous chemical reactions, an important branch of chemical science, greatly influence the economic and environmental sustainability of chemical manufacturing through efficiency, selectivity, and controllability.[1] Accurately predicting the performances of homogeneous chemical reactions — specifically the relationship between reactants, catalysts, and their effects on yield, enantioselectivity, and reaction activation energy — has far-reaching impacts on advancing catalyst development, optimizing drug synthesis pathways, and enhancing environmental management strategies.[2-4] Traditional experimental techniques investigating chemical reaction performance are labor-intensive, costly, and time-consuming. Despite the increased efficiency of reaction optimization enabled by automated reaction systems and optimization algorithms, data-driven prediction of homogeneous reaction performance is a promising alternative for its labor-free, cost-effective, and high-efficiency. However, homogeneous reactions cover intricate interactions between compounds, while the complexity of chemical space and various factors from experimental data make the task of data-driven reaction prediction even more challenging.[5]

Developing efficient techniques for chemical reaction representation is essential for the advancement of traditional machine learning methodologies since they largely rely on feature engineering. Thus, tremendous progress has been made in establishing different descriptors.[6-7] Molecular fingerprints, which are one-dimensional (1D) vectors based on molecular structures, can capture key differences between compounds and predict reaction performance when combined at the molecular level.[8-12] As an instance, the MFF proposed by Glorius *et al*. achieved an excellent performance on three chemical reaction prediction tasks using a random forest (RF) machine learning model.[13] Whereas the wider utilization of molecular fingerprints is limited due to their inability to make predictions for homogeneous reactions containing a fluctuating number of molecules. Density functional theory (DFT) parameters are also commonly employed for a wide range of reaction prediction tasks for their exceptional description of electronic characteristics of chemical compounds.[14-22] Associated works consist of Gao *et al*., who also used the RF model to predict the enantiomeric excess value of asymmetric transfer hydrogenation by combining spatial characteristics with DFT parameters to describe the electronic properties of the substrates and catalysts.[23] Still, the massive computational cost of DFT parameters limits the large-scale exploration of novel catalysts.[24]

The advancement of deep learning techniques has spurred the rise of reaction prediction. Among them, graph neural networks have been vigorously advocated for their ability to emulate chemical topology information.[25-31] Graph modality portrays the physical and chemical properties of the chemical compound, also exhibiting the characteristics of special atoms in compounds as two-dimensional (2D) features, which express the local information of chemical reactions. Motivated by this, Hong *et al*. introduced SEMG-MIGNN, a graph-based model considering the interaction between reactants, and successfully predicted both yield and enantiomeric excess value of reactions with an explanation of essential groups.[32] Nevertheless, akin to most existing efforts, their model focused on the interpretation of key chemical groups, neglecting an in-depth study of the reaction mechanisms. Additionally, due to the flexibility in handling chemical reactions with indeterminate molecular numbers, the simplified molecular input line entry system (SMILES) in text modal is also frequently utilized for chemical reaction prediction.[33-35] While SMILES emphasizes the potential interrelationships of different atoms in a reaction and describes the 1D features of the reaction to fully represent the global information of chemical reactions. A typical example is Yield-BERT accomplished reaction yield prediction as it is not restricted by a fixed reaction template.[36] The multi-modal model is gaining increasing attention for the benefit of modality complementariness.[37-38] Yield-GNN proposed by Saebi *et al*. predicted chemical reaction yield with high precision on high-throughput experimentation (HTE) datasets by integrating graph information and molecular descriptors but limited on a large-scale dataset.[39]

Chemical reaction prediction is a challenging field that requires not only an excellent representation of the chemical compound features but also an efficient capacity to extract their associations, solely utilizing graph modal or text modal can only extract reaction features from one perspective. Therefore, the combination of the two modalities to glean maximum insights into chemical reaction knowledge aids in the precise prediction of chemical reaction performance and also illuminates the complicated interaction between reaction information and phenomena.

Along this line, we proposed a Multi-Modal Homogeneous Reaction Predictor using Graph and Chemical Language information (MMHRP-GCL). From a topological and sequence viewpoint, we designed a chemical reaction encoder that integrates text and graph modalities to facilitate homogeneous chemical reaction performance prediction at low computational cost. In three reaction performance prediction tasks — yield prediction, enantioselectivity prediction, and reaction activation energy prediction — our model outperforms existing state-of-the-art (SOTA) methods by evaluating the coefficient of determination (R^2), root mean squared error (RMSE) and mean absolute error (MAE) values comprehensively and comparatively. Subsequent ablation studies demonstrate the successful design idea of information complementariness of graph and chemical language information fusion and also emphasize the significance of modality alignment in multi-modal model development as well as the atomic features for graph data. Via masking techniques and attention matrix analysis, we assessed the ability of the model interpretation of atomic importance and potential associations. Afterwards, we evaluated the performance of the model on a challenging dataset and discussed the prospects for its application. MMHRP-GCL, a generalizable and interpretable model with high prediction precision and low computational cost, as a novel alternative to homogeneous chemical reaction performance prediction, offers valuable insights into the model design for reaction forward prediction.

Results and Discussion

Datasets descriptions

In this work, we chose 4 datasets: Buchwald-Hartwig reaction (B-H) HTE dataset,[17] Suzuki-Miyaura reaction (S-M) HTE dataset,[40] asymmetric thiol addition reaction (A-T) HTE dataset,[19] and nucleophilic aromatic substitution reaction (S_NAr) literature-based dataset[15] for testing the performance of MMHRP-GCL on 3 kinds of homogeneous reaction prediction tasks. Figure 1 presents the details of 4 datasets:

- B-H HTE dataset: Yield dataset in % unit obtained by HTE approach with 3955 reactions, including 15 aryl halides, 22 additives, 4 Pd catalysts, and 3 bases.
- S-M HTE dataset: Yield dataset in % unit comprising 7 quinolines, 4 indazoles, 12 ligands, 8 bases, 6 solvents, and a total of 5760 reactions were acquired by the HTE platform.
- A-T HTE dataset: The exponential relationship that exists between enantiomeric excess value and free energy differential ($\Delta\Delta G$) between the transition structures leading to each enantiomer. Therefore, the dataset records the $\Delta\Delta G$ value in kcal/mol unit of 1075 reactions acquired via HTE approach, containing 5 imines, 5 thiols, and 43 chiral phosphoric acid (CPA) catalysts, to test the enantioselectivity prediction ability of the model.
- S_NAr literature-based dataset: A collection of activation energy in kcal/mol unit from 37 cases of literature including 443 reactions involving 119 substrates, 83 nucleophiles, and 17 solvents.

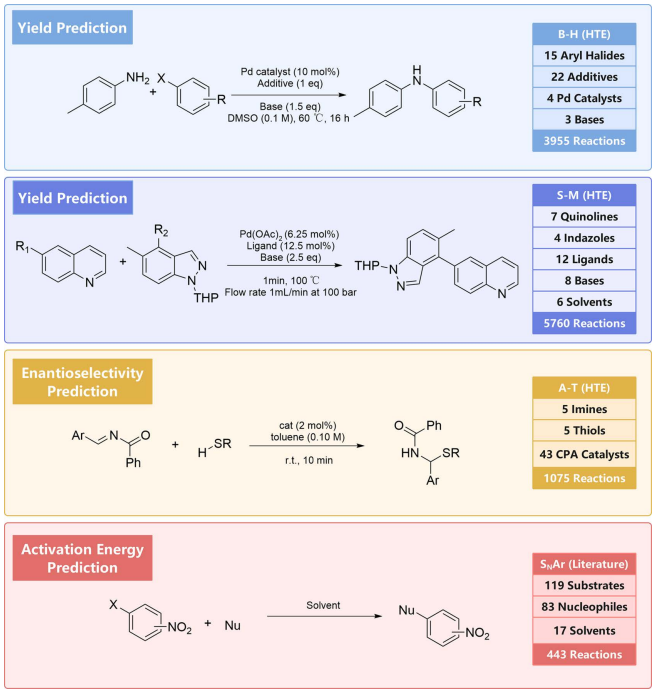


Figure 1 From top to bottom lists the detailed information about B-H HTE dataset, S-M HTE dataset, A-T HTE dataset and S_NAr literature-based dataset, respectively.

Overview of MMHRP-GCL

MMHRP-GCL, which is a two-modal, three-channel model, achieving reaction performance prediction by inputting graph data and text data representing chemical reactions, consists of an encoder for graph data with two input channels, an encoder for text data with one input channel, and a decoder with modality alignment module, as depicted in Figure 2A.

For the data in graph modality, we utilized atomic features in Table 1 as the features for nodes and connected each node in accordance with the way molecules bond to represent compounds in a reaction as the graph data to be fed into the graph encoder. To describe the electronic effect, we used atomic number, former charge, explicit valence, and Gasteiger charge contribution,[42] and the number of bonded hydrogen atoms (Hs), bonding number, whether the atom is in the aromatic system, and whether the atom is in a ring was employed to offer structural information about the surroundings of an atom. All these parameters can be quickly generated by RDKit Python package,[43] saving a huge computational cost comparing to DFT calculation parameters. The 2 channels mentioned above are taken up by the graph encoder, which receives Compounds Graphs, the graph that includes multiple subgraphs with atomic features of compounds. Two GAT units (GATU), each learning the properties of the reactants & products and catalysts & solvents, make up the graph encoder. Figure 2B depicts the structure of GATU. By the graph encoder, the input Compounds Graphs will be embedded as 2 vectors and then spliced together to represent the information of the main reaction and reaction conditions of the chemical reaction.

Computers frequently save chemicals in SMILES format in text modality, for it makes storing molecular structures simple and effective. Chemical reactions involve multiple compounds, thus the use of "." to split different chemicals, and ">" to indicate the direction in which the reaction is proceeding, that is, "REACTANTS > REAGENTS > PRODUCTS" in Reaction SMILES format, which is also shown in Table 1. Since strings cannot be directly utilized as model inputs by neural networks, we converted strings into vectors using the Smi2Vec[44] and input into the text encoder subsequently. The text encoder, aiming at capturing the global information of the chemical reactions, is made up of a transformer layer,[45-46] as illustrated in Figure 2C, a Bi-GRU layer, which is frequently used in compound feature extraction,[38,44,47] as well as a linear layer.

Table 1 Description of the input features

Modality	Feature	Description	Type
Graph	Atomic number	The atomic number of this atom, which shows the type of an atom.	int
	Former charge	Integer electronic charge assigned to atom.	int
	Hs	Number of bonded hydrogen atoms.	int
	Explicit valence	The valance of this atom.	int
	Bonds	Number of bonds the atom is involved in.	int
	Aromatic	Whether this atom is part of an aromatic system.	bool
	Ring	Whether this atom is in a ring.	bool
	Gasteiger charge contribution[42]	Atomic charge distribution descriptors derived from algorithms.	float
Text	Reaction SMILES	SMILES for chemical reactions, with the form "REACTANTS > REAGENTS > PRODUCTS".	string

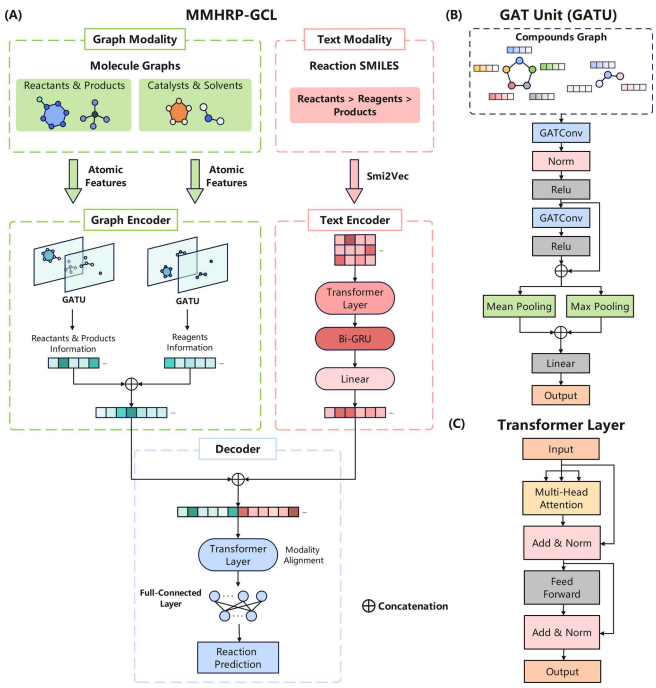


Figure 2 Overview of MMHRP-GCL. (A) Architecture of MMHRP-GCL. (B) Framework of GAT unit (GATU). (C) Framework of the transformer layer.

The vectors generated by the 2 encoders representing the corresponding modal information are subsequently transmitted into the model decoder after a concatenation operation. Aiming at achieving a better complementarity between modal information, we utilized the transformer layer for implicit modality alignment.[41] The modal-aligned vectors are subsequently passed through a 5-layer fully-connected layer to achieve a final prediction of reaction performance. A more concise description of the model is shown in ESI Section 2.

Evaluation methods

To comprehensively evaluate the prediction performance, we assessed the capability of the model to fit or predict the data using R^2, RMSE, and MAE values. See Section 3 in ESI for their specific formulas. For every dataset, we separated it into a train set and a test set. The model was trained on the train set to learn its features, and model performance was evaluated on the test set. The model performs better at prediction on the test set the higher its R^2 values and the lower its MAE and RMSE values. Although k-fold cross-validation allows for a better assessment of model prediction performance, in order to compare with related work, 5 times random splits between the train and test sets were done for each dataset. To facilitate comparisons with the original work for the datasets, we employed distinct splitting for various datasets: train/test sets equal to 70%/30% for the B-H HTE dataset and the S-M HTE dataset, 600/475 for the A-T HTE dataset, and 80%/20% for the S_NAr literature-based dataset.

To assess the models' ability, we replicated and compared the performance of two graph neural network models, GCN[48] and GAT,[49] and two chemical reaction fingerprints proposed based on the SMILES sequences: RXNFP, generated by the BERT model through unsupervised learning,[33] and DRFP, which borrows the creation of circular substructures from a molecule and the subsequent hashing of their SMILES,[11] to be used in RF and XGBoost machine learning models to test their performance on each dataset. Detailed parameter settings and regression performances for the replicated models are provided in Section 2.3 and 4.3 in ESI. Also, we collected some SOTA models and descriptors proposed in related work to compare with our proposed model. The testing approach of the models proposed in these related works differs from ours only in the repeated test number of randomly splitting the train and test set.

Predictions

Yield prediction. MMHRP-GCL is able to predict chemical reaction yields accurately after training on the B-H HTE dataset and the S-M HTE dataset. Table 2 displays the test performance of MMHRP-GCL compared to other SOTA models for the yield prediction task on these 2 datasets. For the single-modal models, we additionally collected the test performance of RF using DFT parameters calculated by Saebi *et al.*,[39] as well as the performance of Yield-BERT.[36] To compare with multi-modal models, we show the test performance data on both datasets for Multimodal BERT[37] and Yield-GNN,[39] which also uses their calculated DFT parameters as input.

It is obvious that the multi-modal models generally outperform the single-modal models, except for Multimodal BERT in the S-M HTE dataset. On the B-H HTE dataset, the best performance of the single-modal models is R^2 = 0.951, while the worst performance of the multi-modal model is R^2 = 0.959, and on the S-M HTE dataset, R^2 = 0.841 represents the best single-modal model performance, while R^2 = 0.855 represents the worst multi-modal model performance, with the exception of Multimodal BERT, demonstrating that the superior performance of multi-modal models due to its ability to integrate data in multi-view. In the B-H HTE dataset, the MAE of the SOTA model is in the range of 0.04—0.13 and in the S-M HTE dataset, the MAE is ranging from 0.08—0.12. Remarkably, MMHRP-GCL outperforms other SOTA

Table 2 Comparison of MMHRP-GCL with other SOTA models on the yield prediction task

Dataset	Model	Features	Evaluation Metrics		
			R^2	RMSE	MAE
		Single-Modal Model			
B-H HTE dataset (Yield Prediction)	RF	DFT[39]	0.915 ± 0.007	0.080 ± 0.003	0.054 ± 0.002
	RF	RXNFP	0.615 ± 0.019	0.169 ± 0.005	0.128 ± 0.004
	RF	DRFP	0.928 ± 0.002	0.073 ± 0.001	0.049 ± 0.001
	XGBoost	RXNFP	0.631 ± 0.023	0.165 ± 0.005	0.119 ± 0.003
	XGBoost	DRFP	0.946 ± 0.005	0.063 ± 0.003	0.042 ± 0.001
	GCN	RDKit	0.802 ± 0.020	0.121 ± 0.007	0.087 ± 0.007
	GAT	RDKit	0.911 ± 0.036	0.080 ± 0.018	0.056 ± 0.011
	Yield-BERT[36]	SMILES	0.951 ± 0.005	0.058 ± 0.004	0.054 ± 0.003
		Multi-Modal Model			
	Yield-GNN[39]	DFT + RDKit	0.961 ± 0.005	—	0.040 ± 0.002
	Multimodal BERT[37]	Chemical Descriptors + SMILES	0.959 ± 0.005	0.055 ± 0.003	—
	MMHRP-GCL (This work)	RDKit + SMILES	**0.968 ± 0.002**	**0.049 ± 0.002**	**0.034 ± 0.001**
		Single-Modal Model			
S-M HTE dataset (Yield Prediction)	RF	DFT + RDKit[39]	0.832 ± 0.011	0.118 ± 0.004	0.081 ± 0.002
	RF	RXNFP	0.662 ± 0.017	0.164 ± 0.005	0.122 ± 0.004
	RF	DRFP	0.824 ± 0.006	0.118 ± 0.002	0.081 ± 0.001
	XGBoost	RXNFP	0.642 ± 0.015	0.169 ± 0.005	0.121 ± 0.003
	XGBoost	DRFP	0.841 ± 0.010	0.112 ± 0.004	0.077 ± 0.002
	GCN	RDKit	0.749 ± 0.013	0.141 ± 0.004	0.103 ± 0.002
	GAT	RDKit	0.826 ± 0.006	0.118 ± 0.002	0.082 ± 0.002
	Yield-BERT[36]	SMILES	0.815 ± 0.013	0.121 ± 0.005	0.081 ± 0.003
		Multi-Modal Model			
	Yield-GNN[39]	DFT + RDKit	0.855 ± 0.013	—	0.083 ± 0.001
	Multimodal BERT[37]	Chemical Descriptors + SMILES	0.833 ± 0.010	0.115 ± 0.003	—
	MMHRP-GCL (This work)	RDKit + SMILES	**0.872 ± 0.011**	**0.101 ± 0.004**	**0.067 ± 0.002**

models in both datasets without the need for DFT calculation with R^2 by 0.968, RMSE by 0.049 and MAE by 0.034 on the B-H HTE dataset, and R^2 by 0.872, RMSE by 0.101 and MAE by 0.067 on the S-M HTE dataset, indicative of its capability of achieving higher accuracy in yield prediction. Although Han *et al*. used a pre-trained graph neural network to achieve higher performance than our model on both the B-H HTE and S-M HTE datasets, it is worth noting that they used an independent GNN model for each reactant in graph data which implies that they were unable to handle chemical reactions with an unfixed number of reactants.[26] MMHRP-GCL uses Compounds Graphs and Reaction SMILES as inputs, indicating its ability in handling chemical reactions with a flexible number of molecules reacted.

Enantioselectivity prediction. Training on the A-T HTE dataset enabled MMHRP-GCL to predict the enantioselectivity values of chemical reactions. The tested results of the models are presented in Table 3. At an R^2 of 0.910, RMSE of 0.203 kcal·mol^{-1}, and MAE of 0.143 kcal·mol^{-1}, the evaluation metrics of MMHRP-GCL confirm its highest accuracy in reaction enantioselectivity prediction. However, the original work on the A-T HTE dataset achieved an R^2 of 0.917 on the test set, using DFT to calculate multiple parameters to describe the electronic and spatial properties of the molecules with the supporting vector machine to learn the reaction features, even though the MAE of this method is 0.152 kcal·mol^{-1}, which is higher than that of our proposed model.[19] Consequently, even only employing basic descriptors, our model can still compete with the DFT parameters that are capable of accurately describing chemical reactions.

Activation energy prediction. We also investigated the reaction activation energy prediction ability of MMHRP-GCL. Test results of our proposed model, together with other SOTA models, are shown in Table 4. Obviously, the MMHRP-GCL shows marginal improvement in activation energy prediction. Apart from our replicated models, we also compared the prediction performance with the original work based on the Gaussian processing regression (GPR) model using DFT parameters or pre-trained BERT parameters.[15] The original work achieved high accuracy prediction of chemical reaction activation energy using DFT calculation parameters with MAE of 0.80 kcal·mol^{-1}, the smallest error among many models, while our proposed MMHRP-GCL reaches the highest accuracy among R^2 and RMSE, which are 0.88 and 1.26 kcal·mol^{-1}, respectively, indicating that our model is another accurate alternative method for reaction activation energy prediction.

Ablation study

The ablation study is conducted by removing parts of the model structure to assess the importance of the deleted parts to the model. To support the model design, we carried out an ablation study for MMHRP-GCL. Table S15 lists the test performance on the four datasets for the baseline model, the model without the graph encoder, the model without the text encoder, the model without the modality alignment, and the model without the atomic features in graph modality. Obviously, the baseline model performs best across all datasets. Compared to the models without graph encoder or text encoder, the result proves that the graph and text modality can complement one another's information to achieve a higher accuracy prediction for reaction performance prediction tasks than a single modality. Moreover, in Figure S1, a dimensionality reduction analysis towards the latent vectors of each modality was performed to further prove the complement of information between the two modalities, even though graph and SMILES representations are essentially interconvertible. The graph encoder appears to be a key component of

Table 3 Comparison of MMHRP-GCL with other SOTA models on the enantioselectivity prediction task

Dataset	Model	Features	Evaluation Metrics		
			R^2	RMSE	MAE
			Single Modality Model		
A-T HTE dataset (ΔΔ*G* Prediction)	RF	RXNFP	0.446 ± 0.027	0.502 ± 0.014	0.374 ± 0.010
	RF	DRFP	0.907 ± 0.009	0.206 ± 0.012	0.143 ± 0.009
	XGBoost	RXNFP	0.419 ± 0.055	0.514 ± 0.016	0.388 ± 0.010
	XGBoost	DRFP	0.887 ± 0.012	0.226 ± 0.013	0.155 ± 0.008
	GCN	RDKit	0.880 ± 0.007	0.234 ± 0.009	0.173 ± 0.006
	GAT	RDKit	0.872 ± 0.017	0.241 ± 0.015	0.175 ± 0.011
			Multi-Modal Model		
	MMHRP-GCL (This work)	RDKit Features + SMILES	**0.910 ± 0.007**	**0.203 ± 0.008**	**0.143 ± 0.002**

Note: The units of RMSE and MAE are kcal·mol^{-1}.

Table 4 Comparison of MMHRP-GCL with other SOTA models on the activation energy prediction task

Dataset	Model	Features	Evaluation Metrics		
			R^2	RMSE	MAE
			Single Modality Model		
S_NAr dataset (Activation Energy Prediction)	GPR	DFT[15]	0.87 ± 0.02	1.41 ± 0.14	**0.80 ± 0.06**
	GPR	BERT (pre-trained)[15]	0.85 ± 0.02	1.51 ± 0.10	1.03 ± 0.05
	RF	RXNFP	0.68 ± 0.07	2.05 ± 0.17	1.52 ± 0.15
	RF	DRFP	0.69 ± 0.06	2.04 ± 0.26	1.30 ± 0.09
	XGBoost	RXNFP	0.66 ± 0.10	2.08 ± 0.23	1.52 ± 0.16
	XGBoost	DRFP	0.71 ± 0.10	1.95 ± 0.36	1.17 ± 0.11
	GCN	RDKit	0.77 ± 0.05	1.73 ± 0.18	1.32 ± 0.12
	GAT	RDKit	0.82 ± 0.07	1.51 ± 0.32	1.06 ± 0.22
			Multi-Modal Model		
	MMHRP-GCL (This work)	RDKit Features + SMILES	**0.88 ± 0.04**	**1.26 ± 0.18**	0.93 ± 0.15

Note: The units of RMSE and MAE are kcal·mol^{-1}. Reaction activation energy is retained here to 2 decimal places as they can't be obtained accurately.

the MMHRP-GCL model, as evidenced by the significantly inferior performance of the model without graph encoder compared to the model without text encoder, with the R^2 value higher by 0.024—0.107. Hence, the lower bound on the prediction ability of the model is determined by its graph encoder. This may be due to the fact that the data in graph modality is better suited to capturing the differences in the substituents for different compounds.

We designed a modality alignment layer based on the transformer layer for the model to reduce the performance degradation brought on by the redundant information in multi-modal data.[41] According to the ablation study, the model without modality alignment is marginally worse than the baseline model, whose R^2 value is lower by 0.001—0.015, suggesting that modality alignment layer design indeed significantly enhanced the model performance.

We subsequently evaluated the importance of atomic features by the ablation study. The performance of MMHRP-GCL without atomic features is much lower than that of the benchmark model, with R^2 values lower by 0.022—0.116 on the 4 datasets, indicating that the atomic features are essential to the graph encoder. It is worth noting that the model without atomic features outperforms the model without graph encoder in all but the A-T HTE dataset, with R^2 values higher by 0.003—0.020, showing that in situations where the descriptors in graph data are unable to characterize the compound properties, the model can still rely on the text data to predict the reaction performance with a relatively high degree of accuracy, which proves the superiority of multi-modal modeling.

Interpretability

Deep learning has always been challenging due to its black-box nature.[50-51] Numerous comprehensible methods have been devised to tackle this problem, and attention matrices analysis and perturbation-based approaches are often employed among them.[32-33,52] In this work, MMHRP-GCL will use these two methods respectively, to reveal the implicit relationship of atoms, which is beneficial for the exploration of the reaction mechanism, and to identify the important atoms to assist in the design of new chemocatalytic conditions. The methods for model interpretation and more interpretation results are shown in Section 5 in ESI.

Figure 3 shows the results of model interpretation for one case in B-H HTE reaction and S-M HTE reaction. In Figure 3F, the model finds a potential relationship between "[Pd+2]" and "I", representing an association between the Pd(II) ion in $Pd(CH_3COO)_2$ and -I in 6-iodoquinoline, which is consistent with the insertion reaction of the Pd atom into the C—I bond in the S-M reaction.[53] Nevertheless, in Figures 3B, the model shows the implicit relationship of all "c" and "o" with "P", which are not chemically meaningful, indicating the model has limited ability to explain potential relationships of atoms. Strikingly, however, MMHRP-GCL has good explanatory ability for atomic importance. In Figures 3C, 3G, the model emphasizes the importance of -Cl in 3-chloropyridine, -I in 6-iodoquinoline, demonstrating that the model is able to capture the electronic effects of the heteroatoms. The model also pays attention to the O and N atoms of isoxazole in Figure 3D, which is the insertion sites of Pd metal during the side reaction of B-H reaction,[54] and the -OH and =O groups in the CPA reagent from Figure S2D, which form hydrogen bonds to control the reaction stereoselectivity.[55] These outcomes further confirm the ability of the model to focus on the key atoms in the compound and hence find potential reaction sites accordingly.

With MMHRP-GCL, we obtained the importance of each atomic feature in the graph modality, which further demonstrated the utility of the model, the results of which are depicted in Figure 4. It is evident that every atomic feature contributed to the reaction prediction of MMHRP-GCL. Among them, "Formal Charge" is of less significance in all the datasets, due to the fact that no atom experienced a change in formal charge before or after the reaction. The B-H HTE dataset has similar feature importance to the S-M HTE dataset and low similarity to the other two datasets, indicating that similar feature information can be utilized for the same reaction prediction task. In the S_NAr dataset, the "Gasteiger Charge Contribution" is the second-most important feature, which is in line with the phenomenon that the S_NAr reaction is strongly influenced by the electronic effects of substituents, suggesting that the interpretation result towards feature importance is chemically meaningful.

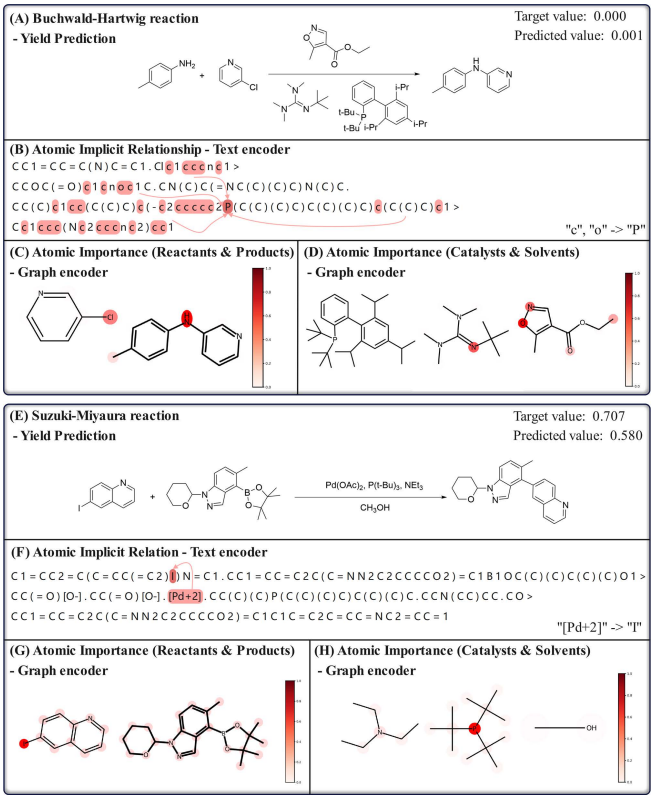


Figure 3 Interpretability results of MMHRP-GCL on a case of B-H HTE reaction and S-M HTE reaction. (A) & (E) prediction result. (B) & (F) Atomic implicit relationship. (C) & (G) Atomic importance of reactants & products. (D) & (H) Atomic importance of catalysts & solvents.

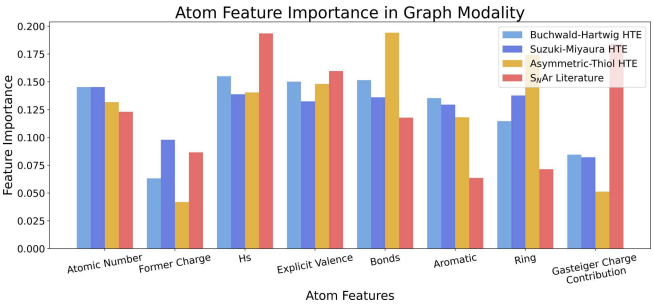


Figure 4 Atomic feature importance on 4 datasets produced by MMHRP-GCL.

Towards challenging dataset

To evaluate the prediction performance of MMHRP-GCL on a challenging task, we tested the model on the ELN dataset,[39] a data-scarce chemical reaction yield dataset. As elaborated in Section 6 of the ESI, the model achieves an R^2 of 0.194. This outperforms the Yield-GNN[39] and GNN models with a context pre-trained strategy.[56] While the performance of MMHRP-GCL is not as good as that of two traditional machine learning models, it

is important to note that these machine learning methods require high-cost DFT parameters for predictions and become more cumbersome and expensive to operate when applied to new reaction predictions. Therefore, we contend that MMHRP-GCL strikes a balance between computational cost and predictive capability.

Possible applications

As an emerging tool for exploring real chemical systems, chemical reaction performance prediction models have played a significant role in recent years. The development of forward predictor for chemical reactions underpins the advancement of automated synthetic robot.[57] An AI-driven automated synthetic robotic platform relies on forward prediction models to forecast reaction outcomes, thereby assessing the feasibility of proposed synthetic routes.[58]

Inspired by these prior works, we posit that MMHRP-GCL, as an effective forward predictor, holds promise for application in real chemical systems. In the Section 7 in ESI, we have outlined an initial framework illustrating the practicality of the MMHRP-GCL model for small laboratories to synthesis target organic compounds at low cost and high yield. This comprehensive exploration marks one of our future goals, and we are eager to collaborate with organic chemists to bring this vision to fruition.

Conclusions

In this issue, we designed MMHRP-GCL, a novel model that combined the information of the data in graph and text modalities to achieve low-cost, high-precision prediction of homogeneous chemical reaction performance. We also designed a modality alignment method based on the transformer encoder to cope with the negative impact of modal data redundancy on model prediction. By parsing the attention matrix in the text modal encoder, we interpreted the relationship between atoms, though sometimes this explanatory capability is limited. Moreover, the perturbation method for graph data highlights significant atoms in the reaction, which can provide insights into the design of chemical reaction conditions. To further emphasize the potential value of our technique, we also employed a statistical interpretability study towards the significance of atomic features for graph data and a test of the model on a data-scarce dataset. Ultimately, the application prospect of MMHRP-GCL was discussed.

MMHRP-GCL was tested for reaction yield prediction on the B-H HTE dataset, S-M HTE dataset with R^2 values of 0.968 and 0.872, respectively, on the test set in 70%/30% splitting of train and test set, reaction enantioselectivity prediction on the A-T HTE dataset with R^2 value by 0.910 on the test set in 600/475 splitting of train and test set, and reaction activation energy prediction on the S_NAr dataset with R^2 value by 0.877 on the test set in 80%/20% splitting of train and test set. Comprehensively comparing R^2, RMSE, MAE values on the test set for 4 datasets indicates that MMHRP-GCL outperforms at least 7 generalized SOTA methods in homogeneous reaction performance prediction requiring only 8 simple descriptors and Reaction SMILES, which significantly saves computing resources compared to the DFT parameters. Also, the ability to handle chemical reactions containing a variable number of molecules is one of the great strengths of the model. Ablation study of MMHRP-GCL proves that both graph and text modalities play a role in reaction performance prediction, and a latent vector analysis for each modality further confirms this. It also validates the importance of the design for modality alignment as well as the atomic features. Cases study shows the reliable identification of key atoms despite the weak interpretability of MMHRP-GCL in atomic implicit relationships. The statistical result towards the significance of atomic features reveals that similar features in graph data are utilized for the same reaction prediction task. Additionally, MMHRP-GCL tested on a challenging yield dataset illustrates its balance between computational cost and prediction accuracy.

Experimental

All data are described in the Datasets Descriptions section of the article. All the training results have been provided in the Electronic Supplementary Information (ESI). All the code and software files have been deposited on GitHub (https://github.com/AIChemShenWang/MMHRP-GCL-Code).

Supporting Information

Electronic Supplementary Information to this article is available on the WWW under https://doi.org/10.1002/cjoc.202401186.

Acknowledgement

This research was supported by the National Science Foundation of China (21903010) and the Fundamental Research Funds for the Central Universities (DUT24BK047).

References

[1] Hinshelwood, C. N. Homogeneous Reactions. *Chem. Rev.* **1926**, *3*, 227–256.
[2] Voinarovska, V.; Kabeshov, M.; Dudenko, D.; Genheden, S.; Tetko, I. V. When Yield Prediction Does Not Yield Prediction: An Overview of the Current Challenges. *J. Chem. Inf. Model.* **2024**, *64*, 42–56.
[3] Zahrt, A. F.; Athavale, S. V.; Denmark, S. E. Quantitative Structure–Selectivity Relationships in Enantioselective Catalysis: Past, Present, and Future. *Chem. Rev.* **2020**, *120*, 1620–1689.
[4] Lewis–Atwell, T.; Townsend, P. A.; Grayson, M. N. Machine Learning Activation Energies of Chemical Reactions. *WIREs Comput. Mol. Sci.* **2022**, *12*, e1593.
[5] Stocker, S.; Csányi, G.; Reuter, K.; Margraf, J. T. Machine Learning in Chemical Reaction Space. *Nat. Commun.* **2020**, *11*, 5505.
[6] Yang, Q.; Liu, Y.; Cheng, J.; Li, Y.; Liu, S.; Duan, Y.; Zhang, L.; Luo, S. An Ensemble Structure and Physicochemical (SPOC) Descriptor for Machine-Learning Prediction of Chemical Reaction and Molecular Properties. *ChemPhysChem* **2022**, *23*, e202200255.
[7] Hong, X.; Yang, Q.; Liao, K.; Pei, J.; Chen, M.; Mo, F.; Lu, H.; Zhang, W.-B.; Zhou, H.; Chen, J.; Su, L.; Zhang, S.-Q.; Liu, S.; Huang, X.; Sun, Y.-Z.; Wang, Y.; Zhang, Z.; Yu, Z.; Luo, S.; Fu, X.-F.; You, S.-L. AI for Organic and Polymer Synthesis. *Sci. China Chem.* **2024**, *67*, 2461–2496.
[8] Haywood, A. L.; Redshaw, J.; Hanson-Heine, M. W. D.; Taylor, A.; Brown, A.; Mason, A. M.; Gärtner, T.; Hirst, J. D. Kernel Methods for Predicting Yields of Chemical Reactions. *J. Chem. Inf. Model.* **2022**, *62*, 2077–2092.
[9] Asahara, R.; Miyao, T. Extended Connectivity Fingerprints as a Chemical Reaction Representation for Enantioselective Organophosphorus-Catalyzed Asymmetric Reaction Prediction. *ACS Omega* **2022**, *7*, 26952–26964.
[10] Segler, M. H. S.; Waller, M. P. Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction. *Chem.-Eur. J.* **2017**, *23*, 5966–5971.
[11] Probst, D.; Schwaller, P.; Reymond, J.-L. Reaction Classification and Yield Prediction Using the Differential Reaction Fingerprint DRFP. *Digital Discovery* **2022**, *1*, 91–97.
[12] Gao, B.; Cai, L.; Zhang, Y.; Huang, H.; Li, Y.; Xue, X.-S. A Machine Learning Model for Predicting Enantioselectivity in Hypervalent Iodine(III) Catalyzed Asymmetric Phenolic Dearomatizations. *CCS Chem.* **2024**, *6*, 2515–2528.
[13] Sandfort, F.; Strieth-Kalthoff, F.; Kühnemund, M.; Beecks, C.; Glorius, F. A Structure-Based Platform for Predicting Chemical Reactivity. *Chem* **2020**, *6*, 1379–1390.
[14] Choi, S.; Kim, Y.; Kim, J. W.; Kim, Z.; Kim, W. Y. Feasibility of Activa-

tion Energy Prediction of Gas-Phase Reactions by Machine Learning. *Chem.-Eur. J.* **2018**, *24*, 12354–12358.
[15] Jorner, K.; Brinck, T.; Norrby, P.-O.; Buttar, D. Machine Learning Meets Mechanistic Modelling for Accurate Prediction of Experimental Activation Energies. *Chem. Sci.* **2021**, *12*, 1163–1175.
[16] Schleinitz, J.; Langevin, M.; Smail, Y.; Wehnert, B.; Grimaud, L.; Vuilleumier, R. Machine Learning Yield Prediction from NiCOlit, a Small-Size Literature Data Set of Nickel Catalyzed C–O Couplings. *J. Am. Chem. Soc.* **2022**, *144*, 14722–14730.
[17] Ahneman, D. T.; Estrada, J. G.; Lin, S.; Dreher, S. D.; Doyle, A. G. Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning. *Science* **2018**, *360*, 186–190.
[18] Zuranski, A. M.; Martinez Alvarado, J. I.; Shields, B. J.; Doyle, A. G. Predicting Reaction Yields via Supervised Learning. *Acc. Chem. Res.* **2021**, *54*, 1856–1865.
[19] Zahrt, A. F.; Henle, J. J.; Rose, B. T.; Wang, Y.; Darrow, W. T.; Denmark, S. E. Prediction of Higher-Selectivity Catalysts by Computer-Driven Workflow and Machine Learning. *Science* **2019**, *363*, eaau5631.
[20] Liu, Y.; Li, Y.; Yang, Q.; Yang, J.; Zhang, L.; Luo, S. Prediction of Bond Dissociation Energy for Organic Molecules Based on a Machine-Learning Approach. *Chin. J. Chem.* **2024**, *42*, 1967–1974.
[21] Zhu, X.; Ran, C.; Wen, M.; Guo, G.; Liu, Y.; Liao, L.; Li, Y.; Li, M.; Yu, D. Prediction of Multicomponent Reaction Yields Using Machine Learning. *Chin. J. Chem.* **2021**, *39*, 3231–3237.
[22] Xu, Y.; Gao, Y.; Su, L.; Wu, H.; Tian, H.; Zeng, M.; Xu, C.; Zhu, X.; Liao, K. High-Throughput Experimentation and Machine Learning-Assisted Optimization of Iridium-Catalyzed Cross-Dimerization of Sulfoxonium Ylides. *Angew. Chem. Int. Ed.* **2023**, *62*, e202313638.
[23] Gao, B.; Chang, Y.; Tang, W. Prediction of the Enantiomeric Excess Value for Asymmetric Transfer Hydrogenation Based on Machine Learning. *Org. Chem. Front.* **2023**, *10*, 1456–1462.
[24] Ding, Y.; Qiang, B.; Chen, Q.; Liu, Y.; Zhang, L.; Liu, Z. Exploring Chemical Reaction Space with Machine Learning Models: Representation and Feature Perspective. *J. Chem. Inf. Model.* **2024**, *64*, 2955–2970.
[25] Wen, M.; Blau, S. M.; Xie, X.; Dwaraknath, S.; Persson, K. A. Improving Machine Learning Performance on Small Chemical Reaction Data with Unsupervised Contrastive Pretraining. *Chem. Sci.* **2022**, *13*, 1446–1458.
[26] Han, J.; Kwon, Y.; Choi, Y.-S.; Kang, S. Improving Chemical Reaction Yield Prediction Using Pre-Trained Graph Neural Networks. *J. Cheminform.* **2024**, *16*, 25.
[27] Kwon, Y.; Lee, D.; Choi, Y.-S.; Kang, S. Uncertainty-Aware Prediction of Chemical Reaction Yields with Graph Neural Networks. *J. Cheminform.* **2022**, *14*, 2.
[28] Bi, H.; Wang, H.; Shi, C.; Coley, C.; Tang, J.; Guo, H. Non-Autoregressive Electron Redistribution Modeling for Reaction Prediction. *arXiv* **2021**, DOI: 10.48550/arXiv.2106.07801.
[29] Heid, E.; Green, W. H. Machine Learning of Reaction Properties via Learned Representations of the Condensed Graph of Reaction. *J. Chem. Inf. Model.* **2022**, *62*, 2101–2110.
[30] Wang, H.; Li, W.; Jin, X.; Cho, K.; Ji, H.; Han, J.; Burke, M. D. Chemical-Reaction-Aware Molecule Representation Learning. *arXiv* **2021**, DOI: 10.48550/arXiv.2109.09888.
[31] Yarish, D.; Garkot, S.; Grygorenko, O. O.; Radchenko, D. S.; Moroz, Y. S.; Gurbych, O. Advancing Molecular Graphs with Descriptors for the Prediction of Chemical Reaction Yields. *J. Comput. Chem.* **2023**, *44*, 76–92.
[32] Li, S.-W.; Xu, L.-C.; Zhang, C.; Zhang, S.-Q.; Hong, X. Reaction Performance Prediction with an Extrapolative and Interpretable Graph Model Based on Chemical Knowledge. *Nat. Commun.* **2023**, *14*, 3569.
[33] Schwaller, P.; Probst, D.; Vaucher, A. C.; Nair, V. H.; Kreutter, D.; Laino, T.; Reymond, J.-L. Mapping the Space of Chemical Reactions Using Attention-Based Neural Networks. *Nat. Mach. Intell.* **2021**, *3*, 144–152.
[34] Sagawa, T.; Kojima, R. ReactionT5: A Large-Scale Pre-Trained Model towards Application of Limited Reaction Data. *arXiv* **2023**, DOI: 10.48550/arXiv.2311.06708.
[35] Jiang, S.; Zhang, Z.; Zhao, H.; Li, J.; Yang, Y.; Lu, B.-L.; Xia, N. When SMILES Smiles, Practicality Judgment and Yield Prediction of Chemical Reaction via Deep Chemical Language Processing. *IEEE Access* **2021**, *9*, 85071–85083.
[36] 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.
[37] Baraka, S.; Kerdawy, A. M. E. Multimodal Transformer-Based Model for Buchwald-Hartwig and Suzuki-Miyaura Reaction Yield Prediction. *arXiv* **2022**, DOI: 10.48550/arXiv.2204.14062.
[38] Shi, R.; Yu, G.; Huo, X.; Yang, Y. Prediction of Chemical Reaction Yields with Large-Scale Multi-View Pre-Training. *J. Cheminform.* **2024**, *16*, 22.
[39] Saebi, M.; Nan, B.; Herr, J. E.; Wahlers, J.; Guo, Z.; Zurański, A. M.; Kogej, T.; Norrby, P.-O.; Doyle, A. G.; Chawla, N. V.; Wiest, O. On the Use of Real-World Datasets for Reaction Yield Prediction. *Chem. Sci.* **2023**, *14*, 4997–5005.
[40] Perera, D.; Tucker, J. W.; Brahmbhatt, S.; Helal, C. J.; Chong, A.; Farrell, W.; Richardson, P.; Sach, N. W. A Platform for Automated Nanomole-Scale Reaction Screening and Micromole-Scale Synthesis in Flow. *Science* **2018**, *359*, 429–434.
[41] Baltrušaitis, T.; Ahuja, C.; Morency, L.-P. Multimodal Machine Learning: A Survey and Taxonomy. *arXiv* **2017**, DOI: 10.48550/arXiv.1705.09406.
[42] Geidl, S.; Bouchal, T.; Raček, T.; Svobodová Vařeková, R.; Hejret, V.; Křenek, A.; Abagyan, R.; Koča, J. High-Quality and Universal Empirical Atomic Charges for Chemoinformatics Applications. *J. Cheminform.* **2015**, *7*, 59.
[43] RDKit: Open-source chemoinformatics and machine learning. http://www.rdkit.org.
[44] Lin, X.; Quan, Z.; Wang, Z.-J.; Huang, H.; Zeng, X. A Novel Molecular Representation with BiGRU Neural Networks for Learning Atom. *Brief. Bioinform.* **2020**, *21*, 2099–2111.
[45] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2023**, DOI: 10.48550/arXiv.1706.03762.
[46] Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2019**, DOI: 10.48550/arXiv.1810.04805.
[47] Du, B.-X.; Long, Y.; Li, X.; Wu, M.; Shi, J.-Y. CMMS-GCL: Cross-Modality Metabolic Stability Prediction with Graph Contrastive Learning. *Bioinformatics* **2023**, *39*, btad503.
[48] Kipf, T. N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv* **2017**, DOI: 10.48550/arXiv.1609.02907.
[49] Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. *arXiv* **2018**, DOI: 10.48550/arXiv.1710.10903.
[50] Liu, X.; An, H.; Cai, W.; Shao, X. Deep Learning in Spectral Analysis: Modeling and Imaging. *TrAC-Trend Anal. Chem.* **2024**, *172*, 117612.
[51] Duan, C.; Liu, X.; Cai, W.; Shao, X. Interpretable Perturbator for Variable Selection in Near-Infrared Spectral Analysis. *J. Chem. Inf. Model.* **2024**, *64*, 2508–2514.
[52] Ying, R.; Bourgeois, D.; You, J.; Zitnik, M.; Leskovec, J. GNNExplainer: Generating Explanations for Graph Neural Networks. *arXiv* **2019**, DOI: 10.48550/arXiv.1903.03894.
[53] Bourouina, A.; Meille, V.; De Bellefon, C. About Solid Phase vs. Liquid Phase in Suzuki-Miyaura Reaction. *Catalysts* **2019**, *9*, 60.
[54] Estrada, J. G.; Ahneman, D. T.; Sheridan, R. P.; Dreher, S. D.; Doyle, A. G. Response to Comment on "Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning." *Science* **2018**, *362*, eaat8763.
[55] Liu, C.; Han, P.; Wu, X.; Tang, M. The Mechanism Investigation of Chiral Phosphoric Acid-Catalyzed Friedel–Crafts Reactions–How the Chiral Phosphoric Acid Regains the Proton. *Comput. Theor. Chem.* **2014**, *1050*, 39–45.
[56] Hu, W.; Liu, B.; Gomes, J.; Zitnik, M.; Liang P.; Pande V.; Leskovec, J. Strategies for Pre-training Graph Neural Networks. *arXiv*, **2019**, DOI: 10.48550/arXiv.1905.12265.

[57] Liu, Y.-D.; Qi, Y.; Li, Y.; Zhang, L.; Luo, S. Z. Application of Machine Learning in Organic Chemistry. *Chin. J. Org. Chem.* **2020**, *40*, 3812–3827.
[58] Coley, C. W.; Thomas D. A.; Lummiss, J. A. M.; Jaworski, J. N.; Breen, C. P.; Schultz, V.; Hart, T.; Fishman, J. S.; Rogers, L.; Gao, H.; Hicklin, R. W.; Plehiers, P. P.; Byington, J.; Piotti, J. S.; Green, W. H.; Hart, A. J.; Jamison, T. F.; Jensen, K, F. A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* **2019**, *365*, eaax1566.

Manuscript received: November 21, 2024
Manuscript revised: January 4, 2025
Manuscript accepted: January 27, 2025
Version of record online: February 20, 2025

The Authors

Left to Right: Shen Wang, Weiren Zhao, Yining Liu and Yang Li

