ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

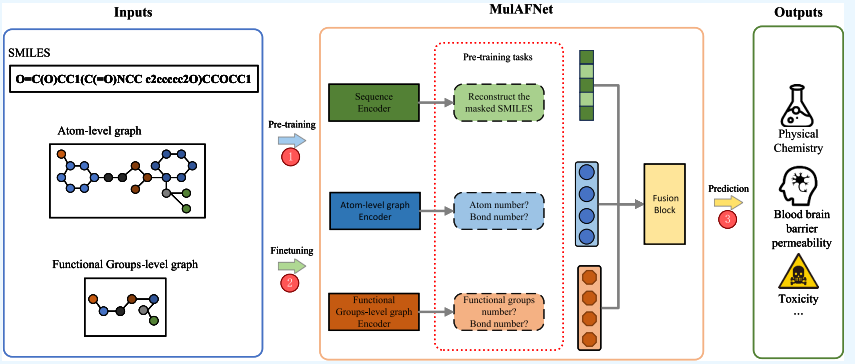
MulAFNet: Integrating Multiple Molecular Representations for Enhanced Property Prediction

Lei Ci, Beilei Li, Jiahao Xu, Sihua Peng, Linhua Jiang, and Wei Long*

Cite This: *ACS Omega* 2025, 10, 12043−12053

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: In computer-aided drug design, molecular representation plays a crucial role. Most existing multimodal approaches primarily perform simple concatenation of various feature representations, without adequately emphasizing effective integration among these features. To address this issue, this study proposes a network framework that integrates multimodal representations using a multihead attention flow (MulAFNet). MulAFNet utilizes SMILES string representation and two levels of molecular graph representations: atom-level and functional group-level graph structure. Pretraining tasks are established for each of these three representations, which are then fused in downstream tasks to predict molecular properties. The experiments were conducted on six classification data sets and three regression data sets, demonstrating that the use of multiple molecular representations as input has a significant impact on the results. In particular, the excellent performance of our fusion method in molecular property prediction outperforms other state-of-the-art methods, proving its superiority. Additionally, comparative experiments on fusion methods and ablation studies, further validate the effectiveness of MulAFNet. The results demonstrate that multiple molecular feature representations provide a more comprehensive molecular understanding, and appropriate pretraining tasks enhance molecular property prediction.

1. INTRODUCTION

With the rapid development of artificial intelligence, deep learning methods for predicting molecular properties have become essential in drug discovery.[1] Accurate prediction of molecular properties allows for the screening of lead compounds with desirable properties before synthesizing new drugs, thereby reducing the probability of drug development failures and significantly accelerating the pace of drug discovery.[2] Molecular representation is one of the most critical steps in applying machine learning methods to predicting molecular properties.[3] The feature information obtained from molecular representations varies across different representations. Ensuring efficient feature information integration during the fusion process has been a prominent research focus and challenge.[4]

Natural language processing (NLP)[5] has become a promising field within deep learning. Inspired by NLP, some researchers have applied methods from this field to molecular sequence representations like SMILES.[6] Classic NLP algorithms, including RNN,[7] GRU,[8] and Transformer,[9] have been successfully applied to molecular property prediction. However, molecular sequence representations like SMILES do not capture molecular topology,[10] which limits their effectiveness for property prediction.[11] Consequently, there is growing interest in graph representations of molecules,[12,13] where atoms are treated as nodes in the graph, and bonds are treated as edges in the graph.[14] Graph neural networks aggregate feature information from adjacent atoms and bonds, layer by layer, to generate a feature vector for the molecule,[15] which is then used for training specific downstream tasks. However, due to the generality of graph neural networks, the specific functional

Received: October 31, 2024
Revised: February 12, 2025
Accepted: February 28, 2025
Published: March 19, 2025

group structures within chemical molecules are often overlooked.[16] Recent approaches focus on extracting functional groups from molecules, dividing the original atom-level graph into structures composed of functional groups, which are then combined into what we refer to as FG-level graphs.[17] On the other hand, in the field of chemical information, labeled data sets of molecules are scarce and expensive to obtain.[18] Exploring the infinite chemical representation space using limited labeled molecular data sets is highly challenging.[19] However, the number of unlabeled molecules is significantly large in comparison.[20,21] Addressing this issue effectively involves using self-supervised learning methods with auxiliary tasks during pretraining.

Several relevant methods have been proposed to address the challenges associated with molecular property prediction. FP-GNN combines molecular fingerprint representations with molecular graph representations to learn molecular features, resulting in improved performance compared to individual molecular representations.[22] MMRLFN employs dual network architectures to simultaneously learn and integrate molecular graph representations and SMILES sequences for drug molecule characterization.[23] MMFRL explores the optimal stages for fusing different modalities in molecular property prediction tasks.[24] Although these methods utilize multimodal molecular representations as model inputs, their feature fusion strategies have certain limitations. Specifically, they rely on simple concatenation of feature vectors and fail to fully exploit the effective integration of molecular representations. Studies have shown that using a single molecular representation limits the capability of molecular property prediction.[25] However, efficiently integrating multiple molecular representations to enhance the predictive performance of molecular properties remains a critical challenge that requires further investigation.

In response to the aforementioned challenges mentioned above, we propose in this study a pretrained network called MulAFNet that integrates multiple feature representations using multihead attention flow. MulAFNet incorporates three molecular feature representations: atom-level graph representation, functional group-level graph representation, and molecular sequence representation. Subsequently, each of these three molecular feature representations is subjected to separate pretraining tasks for self-supervised learning, thereby enhancing the capability of molecular feature representation to improve performance in downstream tasks. Finally, the three types of feature information are fused using a multihead attention flow mechanism, and the fused representation is then fed into downstream tasks for molecular property prediction. In summary, our contributions are as follows:

(1) We utilized molecular sequence representation, atom-level graph representation, and functional group-level graph representation, providing rich feature representations right from the input stage.

(2) Pretraining tasks were designed separately for the three molecular representations, aiming to enhance the encoders' ability to capture more chemical semantics and feature information for each representation.

(3) The multihead attention flow fusion module was employed to integrate the information from the three molecular feature representations, facilitating the subsequent tasks.

2. MATERIAL AND METHODS

2.1. Materials. The ZINC15 unlabeled data set, consisting of approximately 250,000 unlabeled molecular data points, was used for the self-supervised pretraining tasks.[26] During the fine-tuning stage, we employed six classification data sets and three regression data sets from MoleculeNet.[27] These data sets cover a wide range of molecular properties including physical chemistry, biophysics, and physiology, such as BACE,[28] BBBP,[29] Clintox,[30,31] Tox21,[32] ToxCast,[27] and Sider[33] for classification, and ESOL,[27] Freesolv,[34] and Lipophilicity[27] for regression. For comparison with other methods, we used the scaffold method to partition all data sets into training, validation, and test sets at an 8:1:1 ratio. Detailed information about the data sets is presented in Table 1. Further descriptions are provided in the Supporting Information.

Table 1. Information about the Pretraining and Downstream Task Data Sets used in MulAFNet

data sets	size	tasks	task type	metrics
BACE	1522	1	classification	ROC-AUC
BBBP	2053	1	classification	ROC-AUC
Tox21	8014	12	classification	ROC-AUC
ToxCast	8615	617	classification	ROC-AUC
SIDER	1427	27	classification	ROC-AUC
ClinTox	1491	2	classification	ROC-AUC
ESOL	1128	1	regression	RMSE
FreeSolv	643	1	regression	RMSE
Lipophilicity	4200	1	regression	RMSE
ZINC15	249456			

2.2. MulAFNet Framework. This section provides a detailed overview of the MulAFNet framework, with the model architecture depicted in Figure 1. First, we introduce the concept of different molecular feature representations and how these representations are constructed. Next, we describe the pretraining tasks designed for each molecular feature representation. Finally, we elucidate the multihead attention flow fusion module and the prediction module of MulAFNet.

2.2.1. Molecular Feature Representations. MulAFNet includes three molecular feature representations: the molecular SMILES sequence representation, atom-level graph representation, and functional group-level graph representation. SMILES encoding is a common method in chemical informatics, widely used for molecular representation in various research methods. It represents atoms, bonds, and ring information in molecules using character strings, forming a sequential representation.

Atom-level graph representation of molecules. Molecules can naturally be viewed as graph structures, where atoms of the molecule are considered as nodes in the graph, and the chemical bonds of the molecule are considered as edges in the graph.

An atom-level graph can be represented as $G_{atom} = \{V_{atom}, E_{atom}\}$.In a molecule, $Vatom = \{u_{a,i} \mid i \in [1,\cdots,N_a]\}$ represents the set of atoms in the molecule, where $u_{a,i}$ denotes the i-th atom. $E_{atom} = \{e_{a,ij} \mid i,j \in [1,\cdots,N_a]\}$ represents the set of bonds in the molecule, where $e_{a,ij}$ denotes the bond between the i-th and j-th atoms. N_a represents the number of atoms in the molecule. The initial atom features are denoted as $u_{a,i} \in R^{d_n}$, where d_n represents the dimensionality of atom features. Similarly, the initial bond features are denoted as $e_{a,ij} \in R^{d_e}$, where d_e represents the dimensionality of bond features. Using rdkit,[35] atomic feature information and bond formation can be extracted from SMILES to construct the molecular atom-level

https://doi.org/10.1021/acsomega.4c09884
ACS Omega 2025, 10, 12043−12053

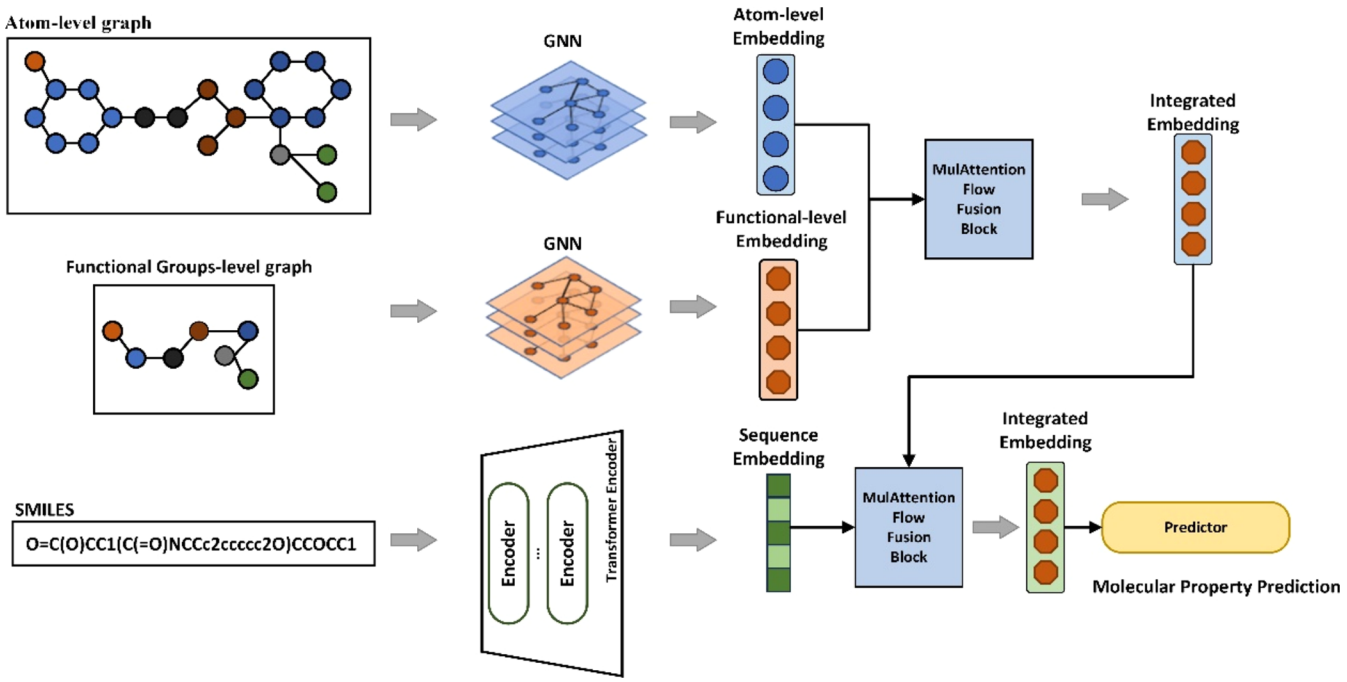


Figure 1. Overview of MulAFNet. MulAFNet utilizes three types of molecular representations as inputs: atomic-level graph representations, functional group-level graph representations, and SMILES string representations. Each representation is processed by its respective pretrained network model to obtain corresponding embeddings. The atomic-level embedding and functional group-level embedding are then fused through the MulAttention Flow Fusion Block to obtain an integrated vector. This fused feature vector is further combined with the sequence embedding using another MulAttention Flow Fusion Block. The final fused embedding is then used for downstream molecular property prediction.

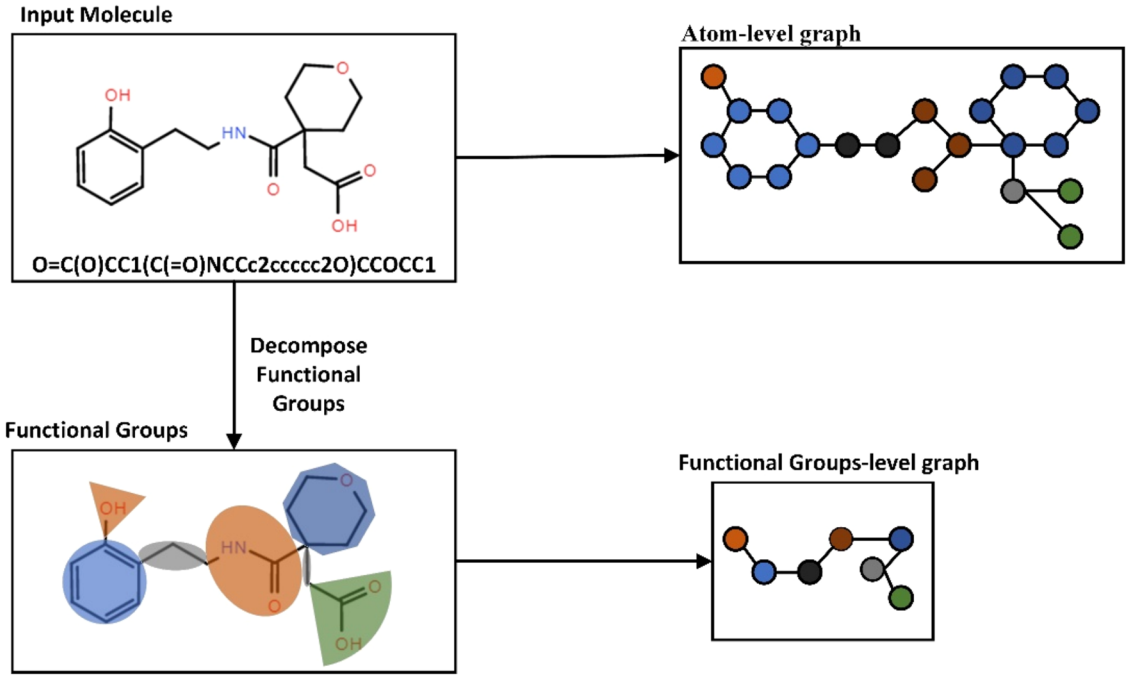


Figure 2. Process of constructing atomic-level and functional groups-level graphs. The SMILES string representation of molecules is first obtained from the data set. The functional groups of the molecule are then identified and split using the FG splitting algorithm, forming a functional group-level molecular graph. On the other hand, the SMILES string is converted into an atomic-level graph representation.

graph structure. Node feature information and bond feature information for the atom-level graph are shown in Table S1.

Functional group-level graph representation of molecules. There are many chemical structures and semantic information in molecules that cannot be expressed solely by atom-level graphs processed through graph neural networks. For example, the structural configuration of functional groups in chemical molecules is closely linked to molecular properties. Therefore, the FG splitting algorithm[17] is adopted to partition the atom-level graph into functional groups, which are then combined to form the functional group-level graph. This can be represented as $G_{fg} = \{V_{fg}, E_{fg}\}$. In the functional group-level graph, each

https://doi.org/10.1021/acsomega.4c09884
ACS Omega 2025, 10, 12043−12053

functional group is treated as a node in the graph. $V_{fg} = \{u_{fg,i} | i \in [1,\cdots,N_m]\}$ represents the set of nodes in the functional group, where $u_{fg,i}$ denotes the i-th node in the functional group. $E_{fg} = \{e_{fg,ij} | i, j \in [1,\cdots,N_m]\}$ represents the set of edges between functional groups, where $e_{fg,ij}$ denotes the edge between the i-th and j-th functional groups. N_m denotes the number of functional groups in the molecule. The initial functional group features are denoted as $u_{fg,i} \in R^{d_m}$, and the bond features are denoted as $e_{fg,ij} \in R^{d_e}$,where d_m and d_e represent the dimensions of the functional group features and bond features, respectively. The node feature information in the functional group-level graph is shown in Table S2. The process of constructing atom-level and functional group-level graphs is illustrated in Figure 2.

2.2.2. Pretraining of Different Molecular Representations. In this framework, a Transformer architecture is employed to extract molecular feature information from SMILES strings. The SMILES representation of a molecule is a linear sequence, and the chemical properties and interactions within it often involve long-range dependencies. The Transformer architecture, with its self-attention mechanism, can efficiently capture these long-range dependencies in the input sequence, especially excelling in handling long SMILES strings. Many studies have shown that Transformer-based models achieve outstanding performance in tasks such as molecular property prediction and chemical reaction prediction. Before the SMILES strings are input into the Transformer, they need to be tokenized. We use the SMILES Tokenization algorithm to first tokenize the SMILES strings, obtaining a Token List. Next, the Token List is converted into vector representations. The dictionary for converting the Token List into vectors is provided in the Table S4.[36] For the SMILES sequence representation of molecules, this study employs unsupervised masked learning. This involves randomly masking portions of the tokenized SMILES encoding, using the masked sequence as input to the model to obtain an output. The original, unmasked tokens are then used as pseudolabels. By calculating the loss between the masked output and the original tokens and gradually minimizing this difference, unsupervised learning is achieved. The cross-entropy loss function is used to calculate the loss between the original and masked tokens, as shown in eq 1. Here, C represents the number of token types, y_i represents the values of the initial token list, and $\hat{y}_i$ represents the values output by the model. Figure 3 illustrates the pretraining process for the SMILES string molecular representation.

$$L_{seq} = -\sum_{i=0}^{C} y_i \log \hat{y}_i \tag{1}$$

For the atom-level and functional group-level graph representations, we use the GIN proposed by Xu et al. [15, 36] as the network backbone. GIN introduces a powerful aggregation function that allows it to distinguish between different graph structures, providing stronger expressive power. In molecular graphs, even small changes in molecular structure can lead to entirely different chemical properties, and GIN, through its designed aggregation strategy, is better able to learn and represent these subtle structural differences. Numerous experiments and studies have shown that GIN, as the backbone of graph neural networks, performs excellently in molecular graph learning tasks. During the self-supervised pretraining process, the number of atoms and bonds is used as pretasks for the atom-level graph, while the number of functional groups and bonds is used as pretasks for the functional group-level graph. The cross-entropy loss function is used to optimize these self-supervised learning tasks. A 3-layer GIN is employed for the atom-level graph, and a 2-layer GIN is used for the functional group-level graph to extract feature information. Feature extraction is achieved by aggregating the features of adjacent atoms.

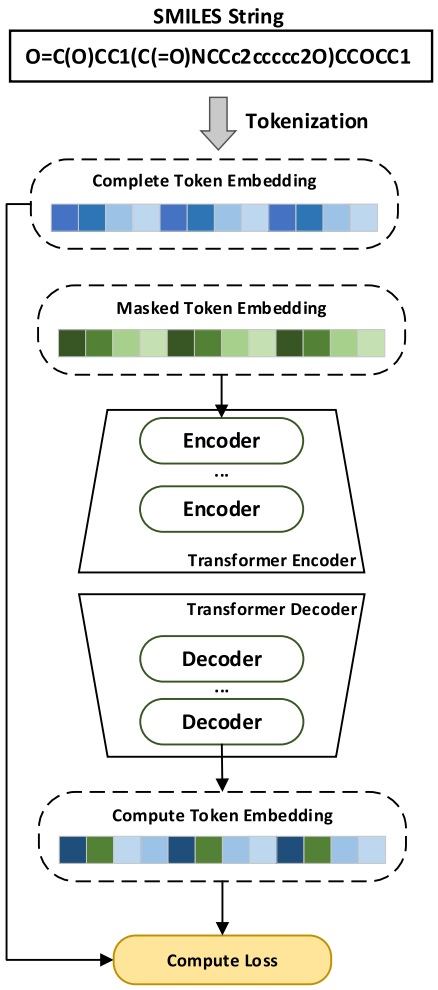


Figure 3. Pretraining process for SMILES string molecular representation. The transformer is used to train the masked SMILES tokens, obtaining the computed token embeddings. These embeddings are then compared with the original, unmasked token embeddings to calculate the loss, thereby achieving the pretraining objective.

$$a_j^k = AGGREGATE^{(k)}(\{h_j^{k-1}, h_i^{k-1}, e_{ij} | i \in N(j)\}) \tag{2}$$

$$h_j^k = COMBINE^{(k)}(h_j^{k-1}, a_j^k) \tag{3}$$

Here, h_j^k represents the node $u_{a,j}$ at the k-th layer in the atom-level graph, e_{ij} represents the bond between atoms $u_{a,i}$ and $u_{a,j}$, and $N(j)$ denotes the set of adjacent nodes to node $u_{a,j}$ in the molecule. $AGGREGATE(k)$ denotes the aggregation function at the k-th layer, and $COMBINE(k)$ denotes the combination function at the k-th layer. h_j^0 represents the initial feature of the atom $u_{a,j}$.

After K propagations, each node captures its K-hop structural information and encodes it as h_j^K. Following this, a readout function is added to pool the information from the nodes in the graph, such as max pooling or average pooling, to obtain the atom-level graph representation h_a. The formula is shown in eq 4. The same graph neural network structure is used for the

https://doi.org/10.1021/acsomega.4c09884
ACS Omega 2025, 10, 12043−12053

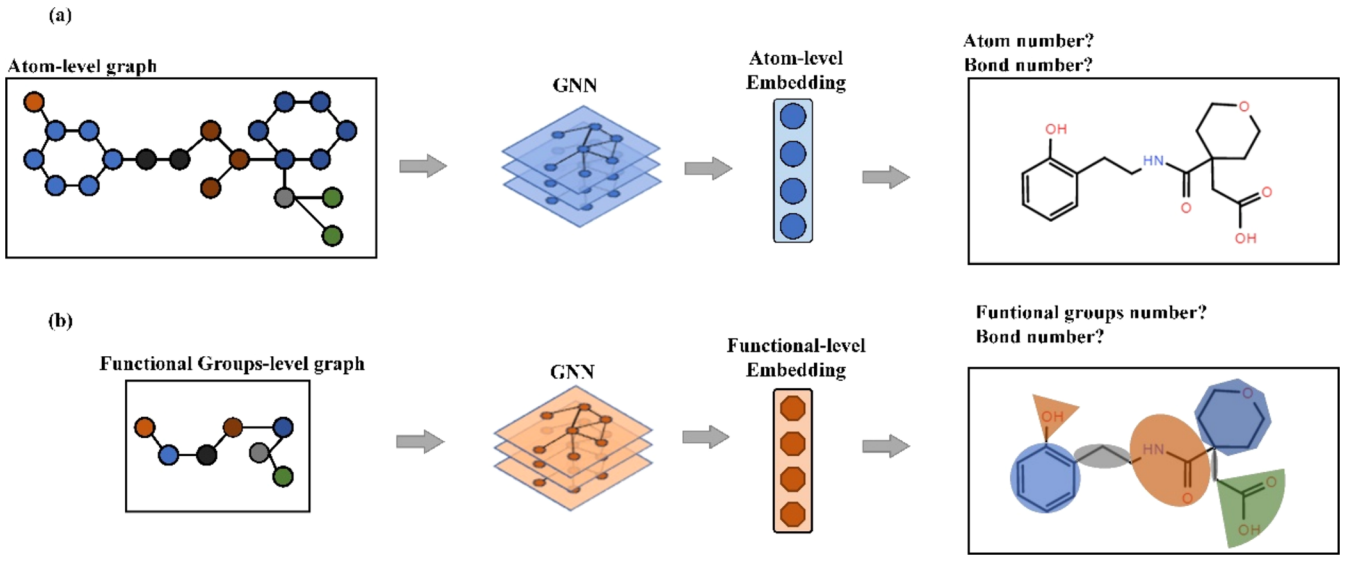


Figure 4. Setup of pretraining tasks for atomic-level and functional group-level graph representations. (a) Illustrates the pretraining process for atomic-level graphs using graph neural networks. (b) Illustrates the pretraining process for functional group-level graphs using graph neural networks.

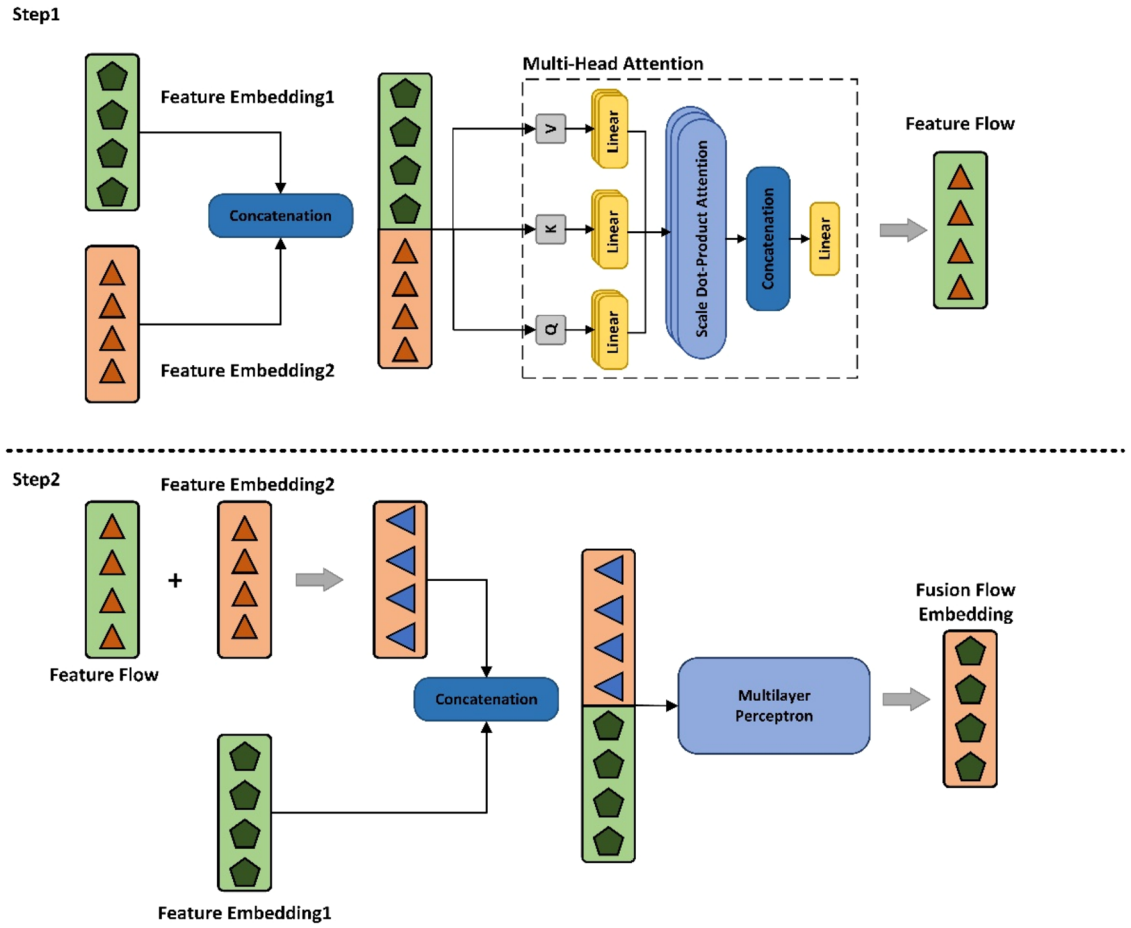


Figure 5. Multihead attention flow fusion module. This module consists of two steps: the first step integrates the input of two feature vectors using multihead attention mechanism to obtain a feature flow field, while the second step further processes the feature flow field and feature vectors to obtain a fused flow field vector.

functional group-level graph, but with a different number of layers.

$$h_a = READOUT(\{h_i^K | i \in [1, ..., N_a]\}) \quad (4)$$

For different molecular representations, different pretraining tasks are designed. For the atomic-level graph representation, the pretraining tasks involve predicting the number of atoms and bonds within the molecule. For the functional group-level graph

https://doi.org/10.1021/acsomega.4c09884
ACS Omega 2025, 10, 12043−12053

representation, the tasks focus on predicting the number of functional groups and bonds within the molecule. SmoothL1 loss is utilized to compute the loss in the prediction tasks, with its formulas provided in eqs 5 and 6.

$$L_{atom\_num} = \begin{cases} 0.5 \times (y_a - \hat{y}_a)^2, & \text{if } |(y_a - \hat{y}_a)| < 1 \\ |y_a - \hat{y}_a| - 0.5, & \text{if } |y_a - \hat{y}_a| \geq 1 \end{cases} \quad (5)$$

$$L_{bond\_num} = \begin{cases} 0.5 \times (y_b - \hat{y}_b)^2, & \text{if } |y_b - \hat{y}_b| < 1 \\ |y_b - \hat{y}_b| - 0.5, & \text{if } |y_b - \hat{y}_b| \geq 1 \end{cases} \quad (6)$$

Here, y_a and y_b respectively denote the true values of the number of atoms and bonds in a molecule, while $\hat{y}_a$ and $\hat{y}_b$ denote the predicted values of atoms and bonds in a molecule. For the pretraining loss function of the atomic-level graph representation, as shown in eq 7, where α_1 and α_2 represent trainable weights.

$$L_{atom} = \alpha_1 L_{atom\_num} + \alpha_2 L_{bond\_num} \quad (7)$$

For the pretraining formulas for functional group-level graph representation as shown in eqs 8 and 9, y_m and y_n respectively denote the true values of the number of functional groups and the number of bonds in a functional group-level graph. $\hat{y}_m$ and $\hat{y}_n$ represent the predicted values of the number of functional groups and the number of bonds in a functional group-level graph.

$$L_{fg\_num} = \begin{cases} 0.5 \times (y_m - \hat{y}_m)^2, & \text{if } |(y_m - \hat{y}_m)| < 1 \\ |y_m - \hat{y}_m| - 0.5, & \text{if } |y_m - \hat{y}_m| \geq 1 \end{cases} \quad (8)$$

$$L_{fg\_bond\_num} = \begin{cases} 0.5 \times (y_n - \hat{y}_n)^2, & \text{if } |y_n - \hat{y}_n| < 1 \\ |y_n - \hat{y}_n| - 0.5, & \text{if } |y_n - \hat{y}_n| \geq 1 \end{cases} \quad (9)$$

The loss function formula for the functional group-level graph representation is shown in eq10, where α_3 and α_4 represent trainable weights. Figure 4 illustrates the setup of the pretraining tasks for both atomic-level graph and functional group-level graph representations.

$$L_{fg} = \alpha_3 L_{fg\_num} + \alpha_4 L_{fg\_bond\_num} \quad (10)$$

2.2.3. The Module for Multihead Attention Flow Field Fusion. The main purpose of designing the multihead attention flow field fusion module is to integrate different molecular feature representations to enhance robustness and stability, as illustrated in Figure 5. We illustrate this with the fusion of atomic-level graph features and functional group-level graph features. First, the atomic-level feature information and functional group-level graph feature information are concatenated to form a combined feature vector, as shown in eq 11.

$$h_{com1} = h_{atom} \oplus h_{fg} \quad (11)$$

The input to the multihead attention mechanism module results in feature information that includes both atomic-level graph features and functional group-level graph features, collectively referred to as the flow field. The formula is as follows.

$$Q = h_{com1} W^Q,\ K = h_{com1} W^K,\ V = h_{com1} W^V \quad (12)$$

$$Attention(Q, K, V) = soft\max\left(\frac{QK^T}{\sqrt{d_K}}\right)V \quad (13)$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (14)$$

$$h_{flow} = Concat(head_1, ..., head_k)V \quad (15)$$

The flow field and functional group-level feature information belong to different spaces. By performing an addition operation, the flow field vector is added to the functional group-level feature information, resulting in the mapping of functional group-level feature information onto atomic-level feature information. This mapped information is then concatenated with atomic-level feature information and passed through a Multi-Layer Perceptron (MLP), resulting in a feature vector that integrates atomic-level graph feature information and functional group-level graph feature information. The formula is shown in (16).

$$h_{FT} = MLP((h_{flow} + h_{fg}) \oplus h_{atom}) \quad (16)$$

where h_{atom} and h_{fg} represent the atomic-level graph representation and functional group-level graph representation, respectively. After concatenating the two, h_{com1} is obtained as the combined vector. h_{com1} is then input into a multihead attention mechanism to compute the flow field vector, h_{flow}. Finally, h_{flow} is added to h_{fg} and concatenated with h_{atom}. This concatenated vector is then fed into a multilayer perceptron (MLP) to obtain the fused feature vector, h_{FT}.

2.2.4. MulAFNet Predicts Molecular Properties. In downstream tasks, pretrained weights are loaded and fine-tuned to obtain embedding vectors for each representation. The atomic-level graph representation, functional group-level graph representation, and sequence vector representation are sequentially fused to obtain a blended feature vector. This blended vector is then fed into a predictor for molecular property prediction. In this study, we evaluated the effectiveness of our proposed framework using six classification and three regression data sets from MoleculeNet for molecular property prediction. For classification tasks, we calculate the binary cross-entropy loss formula as shown in (17).

$$L_{BCE} = -\frac{1}{N}\sum_{i}^{N}(\hat{y}_i \log(p_i) + (1 - \hat{y}_i)\log(1 - p_i)) \quad (17)$$

Here, $\hat{y}_i$ represents the true value of the training sample, and p_i represents the predicted value of the training sample performance. For regression tasks, we calculate the mean squared error, as shown in eq 18, where $\hat{y}_i$ and y_i respectively denote the predicted and true values of the sample.

$$L_{MSE} = \frac{1}{N}\sum_{i}^{N}|\hat{y}_i - y_i|^2 \quad (18)$$

3. RESULTS AND DISCUSSION

In this section, we will discuss the experimental setup during the training process, model training parameters, and the methods proposed in this study compared with other baselines. To demonstrate the effectiveness of our proposed model, we conducted experiments using different molecular feature representations as individual inputs. We performed ablation experiments on the proposed model, including the designed

https://doi.org/10.1021/acsomega.4c09884
ACS Omega 2025, 10, 12043−12053

Table 2. Molecular Property Prediction Performance on Classification Benchmarks[a]

data sets	BACE	BBBP	Tox21	ToxCast	SIDER	ClinTox	avg.
RF	85.1 ± 0.0	71.6 ± 0.0	69.1 ± 0.0	63.48 ± 0.0	62.4 ± 0.0	69.87 ± 0.0	70.2
HiMol	84.3 ± 0.3	73.2 ± 0.8	76.2 ± 0.3	66.3 ± 0.4	61.3 ± 0.5	80.8 ± 1.4	73.7
LGGA	85.6 ± 1.9	72.7 ± 0.6	75.5 ± 0.4	64.5 ± 0.6	65.1 ± 0.7	84.9 ± 3.1	74.7
FG-Bert	84.5 ± 1.5	70.2 ± 0.9	78.4 ± 0.8	66.3 ± 0.8	64.0 ± 0.7	83.2 ± 1.6	74.3
Mole-BERT	80.8 ± 1.4	71.9 ± 1.6	76.8 ± 0.5	64.3 ± 0.2	62.8 ± 1.1	78.9 ± 3.0	74.0
GIT	81.08 ± 1.5	73.9 ± 0.6	75.9 ± 0.5	66.8 ± 0.5	63.4 ± 0.8	88.3 ± 1.2	74.9
MoMu	76.7 ± 2.1	70.5 ± 2.0	75.6 ± 0.3	63.4 ± 0.5	60.5 ± 0.9	79.9 ± 4.1	71.1
MolFM	83.9 ± 1.1	72.9 ± 0.1	77.2 ± 0.7	64.4 ± 0.2	64.2 ± 0.9	79.7 ± 1.6	73.7
MulAFNet	87.5 ± 0.6	75.1 ± 0.3	74.9 ± 0.5	66.4 ± 0.7	63.6 ± 0.2	93.9 ± 1.8	**76.9**

[a]The values in bold highlight the best performing results of each benchmark.

pretraining and feature modules, and compared various fusion methods to validate the effectiveness of our framework.

3.1. Experiments Setting. For the molecular sequence SMILES encoding representation, we employed a 4-layer Transformer architecture consisting of 4 layers each of Encoder and Decoder. Each layer includes a hidden dimension of 256 and utilizes 4 attention heads for information extraction. For pretraining the sequence modal representation, we utilized an Adam optimizer with a learning rate of 1e-4 for self-supervised learning. We set a 15% masking rate and used a cross-entropy loss function to calculate losses based on masked and unmasked outputs. The pretraining consisted of 50 epochs to refine the molecular representation in the sequence modal.

We employ graph isomorphism networks (GIN) as the backbone of graph neural networks for pretraining molecular atom-level and functional group-level graph representations. For atom-level and functional group-level graph representations, we utilize 3 layers of GIN and 2 layers of GIN as encoders respectively, with the ReLU function chosen as the activation function. Finally, both atom-level and functional group-level graphs are mapped to 128-dimensional embeddings. The pretraining task for atom-level graph representation involves predicting the number of atoms and bonds in a molecule. The pretraining task for functional group-level graphs involves predicting the number of functional groups and bonds in the graph. For the pretraining of the atom-level graph, we used the Adam optimizer for optimization. The loss function used is the SmoothL1Loss function. Additionally, trainable parameters are added to balance the losses from different pretraining tasks. We set the pretraining learning rate for graph modalities to 1e-3, with a batch size of 512, and conducted training for 100 epochs.

During the fine-tuning stage, we employ the pretrained model structure for all molecular feature representations by loading pretrained weight models. After passing through their respective encoders, the feature vectors for each molecular modality are obtained. These modality-specific feature representations are then input into a multihead attention flow field fusion module, and the output serves as input to downstream task classifiers for training. We utilize a two-layer MLP at the prediction layer. Dropout is a commonly used regularization technique in deep learning that effectively prevents model overfitting. We selected an appropriate Dropout probability through experimentation, typically testing several common values, such as {0, 0.2, 0.5}. Ultimately, depending on the data set, we chose the most suitable Dropout value. Based on existing studies, we set the parameter range to {5e-5, 5e-4, 1e-4, 1e-3}, and then used grid search and stepwise adjustment experiments to ultimately determine the most suitable learning rate for the model. More detailed information is presented in Table S3.

For classification data sets, we used the AUC-ROC value as the evaluation metric for downstream tasks. For regression data sets, we used RMSE as the evaluation metric. MulAFNet was implemented using the PyTorch-Geometric framework. All experiments were conducted on a Linux server equipped with an Nvidia A100 GPU and an Intel(R) Xeon(R) Silver 4110 CPU @ 2.10 GHz. The computational cost of the training phase is primarily influenced by the model architecture, data set size, and hyperparameter settings. For the pretraining of atom-level and functional group-level graph representations, based on the experimental settings mentioned above and using the ZINC15 data set, it takes approximately 6 h for each on a single Nvidia A100 GPU. Meanwhile, training the sequence representation requires around 4 h.

3.2. Baseline. We selected models with different molecular feature representations for comprehensive evaluation to demonstrate the superior performance of our model. We selected two types of models: those with a single molecular feature representation and those with multiple molecular feature representations. The models with multiple molecular feature representations simultaneously used different types of molecular feature information.

Single molecular feature representation: First, traditional machine learning methods, such as those combining Morgan fingerprints (count-based) with models like Random Forest (RF)[37] and Support Vector Regression (SVR),[38] should be considered as baseline methods. HiMol,[39] while also considering the influence of chemical structure information on molecular properties, combines its motif information with the original atomic information to form a single molecular feature representation. LGGA[40] considers 3D structural information, but this method only uses 3D information. Mole-Bert[41] uses an encoder from a VQ-VAE variant as a disambiguator to distinguish between common and rare atoms, involving only a single molecular feature representation. FG-Bert[42] focuses on functional group information but uses only functional group information for molecular property prediction.

Multiple molecular representations: GIT[43] uses different modalities, including graph modality, sequence modality, image modality, and supplementary chemical knowledge information. MoMu[44] employs a contrastive approach for pretraining, leveraging contrastive learning between molecular structures and text representations. MolFM[45] utilizes a molecular structure modality encoder, prior knowledge graphs, and textual descriptions as inputs to the model.

3.3. Experiment Result. To validate the effectiveness of our proposed model, we conducted classification and regression experiments for molecular property prediction using various data sets from MoleculeNet. Table 2 presents the ROC-AUC

https://doi.org/10.1021/acsomega.4c09884
ACS Omega 2025, 10, 12043−12053

values for the classification data sets, including the mean and standard deviation from five independent tests. Results from comparative experiments are sourced from previously published papers, and our framework uses the same data set partitions as those in the comparative experiments.

From Table 2, we observe that (1) Among the six data sets for molecular property prediction classification, our model achieves the best performance on four data sets. For the remaining data sets, our results are slightly behind the method that performs best specifically on those data sets; (2) On the classification data sets for molecular property prediction, MulAFNet shows a 2% improvement in average performance compared to the previously best models. This indicates that our proposed MulAFNet model can extract more comprehensive molecular representations and perform more accurate molecular property prediction tasks; (3) Compared to other methods that combine multiple representation features, our proposed approach continues to achieve the highest performance evaluation, demonstrating its effectiveness in integrating multiple molecular feature representations.

Table 3 displays the results for regression data sets, where RMSE is used as the metric for evaluation across three data sets

Table 3. Molecular Property Prediction Performance on Regression Benchmarks[a]

data sets	ESOL	FreeSolv	Lipo	Avg.
SVR	1.50	3.14	0.82	1.82
HimGNN	0.87	1.92	0.63	1.14
LGGA	1.13	2.42	0.77	1.44
HiMol	0.83	2.28	0.71	1.27
FG-BERT	0.94	1.75	0.65	1.11
MulAFNet	**0.51**	**0.79**	**0.59**	**0.63**

[a]The values in bold highlight the best performing results of each benchmark.

sourced from MoleculeNet. From the table, it is evident that our method achieves the best performance on all three regression data sets.

3.4. Comparison Experiment of Different Fusion Methods. To validate the effectiveness of our proposed Multi-Head Attention Flow Fusion (MulAFlow) module, we compared other methods with MulAFlow, including Concatenation (Concat), Mean, Max methods, and a direct fusion method of the three molecular representations, denoted as MulAFlow-DF. We used the Multi-Head Attention Flow Fusion method as the baseline and compared the results obtained from other experiments with it, as shown in Figure 6. From the results, we observe that the method using the Multi-Head Attention Flow Fusion module achieves better results compared to other methods on most data sets. This indicates that the Multi-Head Attention Flow Fusion module can better aggregate information from different molecular feature representations.

Compared to MulAFlow-DF, MulAFlow performs significantly better across all data sets. In our design, we hypothesized that atom-level and fragment-level feature representations have more direct structural and topological correlations. Therefore, fusing these representations first allows the model to better capture both local and global information from molecular graphs. Since SMILES-based sequence representations differ greatly from graph-based representations, integrating sequence representations after fusing the graph representations effectively reduces conflicts between modalities, thereby enhancing the

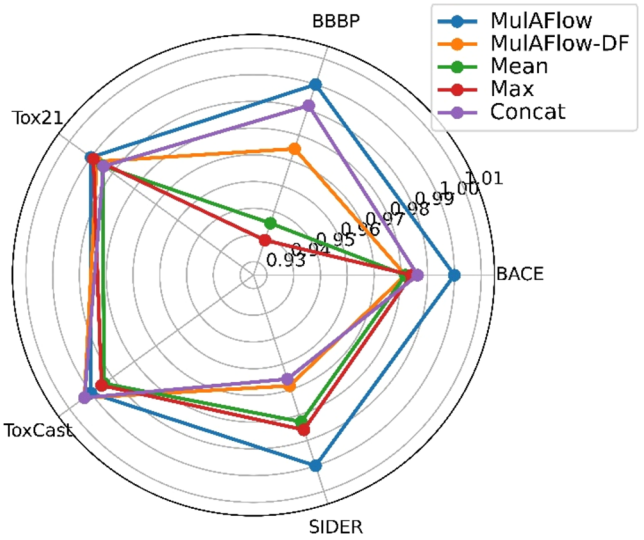


Figure 6. Comparison of different fusion methods. Feature fusion methods, including Mean, Max, Concat, MulAFlow-DF and MulAFlow, were evaluated across five data sets. A radar chart is used to compare the results of other methods with those of MulAFlow, providing a visual representation of the performance differences.

fusion process. Experimental results demonstrate that directly fusing the three feature representations using MulAFlow-DF shows weaker capabilities in modeling interactions among different modalities. This is particularly evident when handling large amounts of multimodal feature information, where redundancy may be introduced, leading to a decline in performance. In contrast, the stepwise fusion method enables a more detailed capture of relationships between different feature representations, significantly improving predictive performance.

On some individual data sets, due to inherent data set differences, the Multi-Head Attention Flow Fusion module may not perform as well as other methods. Data sets with poorer performance often contain more complex or heterogeneous molecular properties, such as multiple types of molecular reactions or toxicity indicators, which pose greater challenges for fusion models. Specifically, in the ToxCast data set, the imbalanced distribution of toxicity labels, along with significant variations in molecular structures and properties, may hinder the fusion method's ability to effectively capture the complex interactions between different molecular representations, leading to a performance decline. Additionally, for certain data sets, specific modalities may provide more complementary information, thereby enhancing the fusion performance. However, for other data sets, the complementarity between modalities may be weaker, resulting in suboptimal performance of the fusion method. Particularly in the Tox21 data set, the alignment between different molecular representations and the target prediction task is relatively low, preventing the fusion method from adequately capturing the most relevant features, which in turn affects overall performance. Overall, the Multi-Head Attention Flow Fusion module proves to be the superior fusion method.

3.5. Comparison Experiment of Different Single Molecular Feature Representations. In this section, we investigate the performance of different single molecular feature representations compared to the comprehensive use of multiple molecular feature representations in our proposed model

https://doi.org/10.1021/acsomega.4c09884
ACS Omega 2025, 10, 12043−12053

framework. Our model framework integrates three types of molecular feature representations: molecular sequence representation, atom-level graph representation, and functional group-level graph representation. We selected five classification data sets and conducted experiments using each of these three molecular feature representations individually for validation. As shown in Figure 7, MulAFNet-Seq uses only the molecular sequence representation, MulAFNet-Atom uses only the atom-level graph representation, MulAFNet-FG uses only the functional group-level graph representation, and MulAFNet represents our complete proposed method. From the figure, it is evident that individual molecular representations perform poorly compared to the comprehensive use of various molecular feature representations. Using single molecular representations can hinder the learning ability of molecular feature representation.

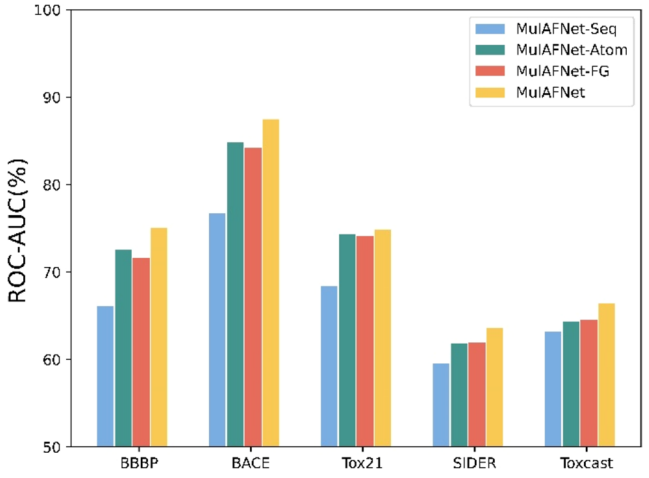


Figure 7. Performance results of different single molecule feature representations. The performance of molecular property prediction using each of the three molecular representations (atomic-level graph, functional group-level graph, and SMILES string) individually is compared with the performance when all three representations are combined. The comparison is conducted across five data sets.

3.6. Ablation Experiments. To further investigate the effectiveness of each module in our proposed model, we conducted a series of ablation experiments. The results of these experiments across different data sets are presented in Table 4, which includes the impact of pretraining on the model as well as the effect of the multiattention mechanism flow fusion module. The MulAFNet-WPT (without pretrain) model indicates the model architecture without the pretraining task. The MulAFNet-WMAFF (without MulAttention flow Fusion) model represents the model architecture where fusion is achieved by simply concatenating various molecular feature information together instead of using the Multi-Head Attention Flow Fusion module. Lastly, the MulAFNet-WA (without all) model removes all modules except for the multifeature molecular representation. Clearly, MulAFNet achieves the best performance across the majority of variants.

Table 4. Ablation Experiment[a]

data sets	BACE	BBBP	Tox21	Toxcast	SIDER
MulAFNet-WPT	87.3	71.9	74.2	65.9	62.1
MulAFNet-WMAFF	86.1	74.4	74.7	**66.6**	61.8
MulAFNet-WA	85.7	72.0	73.8	66.2	62.8
MulAFNet	**87.5**	**75.1**	**74.9**	66.4	**63.6**

[a]The values in bold highlight the best performing results of each benchmark.

3.7. Discussion. Handling the effective fusion of various molecular representation vectors is an inevitable issue, and several studies have addressed this problem.[44,46] Currently, most existing research adopts simple operations such as concatenation or averaging to fuse different molecular representations. While concatenation is commonly used in deep learning, it overlooks the potential shared semantics between different molecular representations. These methods often treat the features of each modality equally, without distinguishing their relevance to the task at hand. In contrast, our method provides a more nuanced and dynamic approach to handling the interactions between different input features. It assigns varying attention weights to each molecular representation based on its relevance to the downstream task. The multihead attention mechanism allows the model to focus on different parts of the input features simultaneously, capturing the complex relationships between molecular representations across multiple levels. This is particularly important in tasks like molecular property prediction, where different representations (e.g., SMILES, atom-level graphs, and functional group-level graphs) provide complementary information. Proper fusion of these representations is crucial to enhance predictive performance.

MulAFNet performs relatively poorly on the Tox21 and SIDER data sets, and its advantage on the ToxCast data set is not as pronounced. This may be due to the inherent complexity of these data sets compared to others, as they include a variety of different molecular reactions and toxicity indicators. The SIDER data set, in particular, suffers from label imbalance, where adverse reaction samples are relatively few, and the diversity of drugs is higher. This imbalance, especially when dealing with rare negative samples, may negatively affect MulAFNet's predictive performance.

From the results of comparing different fusion methods, variants without the Multi-Head Attention Flow Fusion module generally perform lower than MulAFNet across most data sets. Among the five data sets tested, only the variant without this module performs slightly lower than the concatenation operation variant on the toxcast data set, yet still outperforms other variants. Furthermore, single molecular feature representations often lack comprehensive information. A major issue with some models in comparative experiments lies in their relatively singular input information.[40,47] Despite these models showing outstanding performance in certain aspects, their overall performance still falls short compared to models using multiple feature molecular representations.

In the comparison experiments of different single molecular feature representations, we observed the advantages of SMILES string representation, atom-level graph representation, and functional group-level graph representation across different data sets. Integrating these diverse molecular representations allows for a more comprehensive expression of molecular information, which is beneficial for downstream tasks. Of course, there are other studies that partition molecular graphs into subgraphs and use both molecular graphs and the partitioned subgraphs as inputs. For example, GroupGAT[48] divides the molecular graph into subgraphs (groups/substructures), where each subgraph has encoded features. These are then used to characterize subgraphs in a tree-based model, and the features

https://doi.org/10.1021/acsomega.4c09884
ACS Omega 2025, 10, 12043−12053

are concatenated to form feature vectors for downstream tasks. We plan to explore this aspect in future research.

Finally, we conducted ablation experiments by removing the fusion module and pretraining task separately and comparing them with the complete model framework. Experimental results demonstrate that the pretraining task significantly improves model performance across all tested data sets, particularly with a notable enhancement on the BBBP data set. Appropriate pretraining tasks can effectively enhance molecular property prediction performance, whereas inappropriate ones may negatively impact or hinder downstream tasks.

4. CONCLUSIONS

In this study, we propose MulAFNet, a novel pretrained multirepresentation network that leverages Multi-Head Attention Flow Fusion for molecular property prediction. MulAFNet integrates multiple molecular feature representations, including atom-level graph representation, functional group-level graph representation, and molecular sequence representation. These diverse feature types are modeled individually and pretrained using a self-supervised learning approach, allowing the network to effectively capture complementary information from each feature type. After pretraining, the Multi-Head Attention Flow Fusion module combines these feature representations to generate comprehensive embedding vectors that capture rich, multilevel molecular information.

Experimental results demonstrate that MulAFNet outperforms models using single molecular feature representations and current state-of-the-art (SOTA) models in molecular property prediction tasks. By employing multiple feature representations, MulAFNet provides richer and more comprehensive feature input compared to methods using only single molecular representations. Additionally, the Multi-Head Attention Flow Fusion module effectively integrates different molecular feature representations. The study highlights that employing multiple molecular feature representations enhances molecular representations comprehensively, thereby improving performance across various downstream tasks.

Compared to other pretrained methods, MulAFNet not only utilizes multiple molecular feature representations as inputs but also incorporates a more efficient feature fusion module and pretrained tasks designed specifically for different molecular representations, achieving more comprehensive molecular representations. Extensive experimental results demonstrate that our proposed MulAFNet surpasses current state-of-the-art benchmark models on multiple benchmark data sets. Given the advantages of multilevel molecular representation, future research may explore expanding molecular feature information into three dimensions and integrating it into the final molecular feature representation, as well as expanding downstream tasks to apply multiple molecular feature representations in molecular generation and optimization tasks.

■ ASSOCIATED CONTENT

Data Availability Statement

In this paper, all the relevant data sets used are publicly available. The pretraining data set on ZINC15 can be downloaded from https://github.com/zaixizhang/MGSSL/tree/main/motif_based_pretrain/data/zinc. All downstream task data sets can be downloaded from the MoleculeNet Web site. All downstream task data sets can be downloaded from the MoleculeNet Web site. All codes used in this study are available at GitHub:https://github.com/cilei/MulAFNet.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.4c09884.

The details of the atomic-level graph representations and functional group-level graph representations are presented, along with information about the data sets used, parameter settings for the implementation, and the dictionary for converting SMILES (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Wei Long − *School of Information Engineering, Huzhou University, Huzhou 313000, China;* orcid.org/0000-0001-8162-0954; Email: lw@zjhu.edu.cn

Authors

Lei Ci − *School of Information Engineering, Huzhou University, Huzhou 313000, China*

Beilei Li − *Huzhou Fengshengwan Aquatic Products Co., Ltd, Huzhou 313000, China*

Jiahao Xu − *School of Information Engineering, Huzhou University, Huzhou 313000, China*

Sihua Peng − *College of Public Health, University of Georgia, Athens, Georgia 30602, United States*

Linhua Jiang − *School of Information Engineering, Huzhou University, Huzhou 313000, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c09884

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The research was partly supported by the National Natural Science Foundation of China (No. 62175037) and the funding of Zhejiang-French Digital Monitoring Lab for Aquatic Resources and Environment, Department of Science and Technology of Zhejiang Province. We would like to express our sincere gratitude to Ye Shixin and Ding Ying for their invaluable assistance in the revision of this manuscript.

■ REFERENCES

(1) Liu, W.; Wang, Z.; Liu, X.; Zeng, N.; Liu, Y.; Alsaadi, F. E. A survey of deep neural network architectures and their applications. *Neurocomputing* **2017**, *234*, 11−26.

(2) Song, C. M.; Lim, S. J.; Tong, J. C. Recent advances in computer-aided drug design. *Briefings Bioinf.* **2009**, *10* (5), 579−591.

(3) David, L.; Thakkar, A.; Mercado, R.; Engkvist, O. Molecular representations in AI-driven drug discovery: a review and practical guide. *J. Cheminf.* **2020**, *12* (1), 56.

(4) Eklund, M.; Norinder, U.; Boyer, S.; Carlsson, L. Choosing feature selection and learning algorithms in QSAR. *J. Chem. Inf. Model.* **2014**, *54* (3), 837−843.

(5) Hirschberg, J.; Manning, C. D. Advances in natural language processing. *Science* **2015**, *349* (6245), 261−266.

(6) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28* (1), 31−36.

(7) Elman, J. L. Finding structure in time. *Cognitive Sci.* **1990**, *14* (2), 179−211.

(8) Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling, 2014, arXiv:1412.3555. arXiv.org e-Print archive. https://arxiv.org/abs/1412.3555.

(9) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need *Adv. Neural Inf. Process. Syst.*, 2017; Vol. *30*.
(10) Kusner, M. J.; Paige, B.; Hernández-Lobato, J. M. Grammar variational autoencoder. In *International Conference on Machine Learning*; PMLR, 2017; pp 1945−1954.
(11) Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* **2018**, *4* (2), 268−276.
(12) Rong, Y.; Bian, Y.; Xu, T.; Xie, W.; Wei, Y.; Huang, W.; Huang, J. Self-supervised graph transformer on large-scale molecular data. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 12559−12571.
(13) Wang, H.; Li, W.; Jin, X.; Cho, K.; Ji, H.; Han, J.; Burke, M. D. Chemical-reaction-aware molecule representation learning, 2021, ':2109.09888. arXiv.org e-Print archive. https://arxiv.org/abs/2109.09888.
(14) Kipf, T. N.; Welling, M. Semi-supervised classification with graph convolutional networks, 2016, arXiv:1609.02907. arXiv.org e-Print archive. https://arxiv.org/abs/1609.02907.
(15) Xu, K.; Hu, W.; Leskovec, J.; Jegelka, S. How powerful are graph neural networks? 2018, arXiv:1810.00826. arXiv.org e-Print archive. ttps://arxiv.org/abs/1810.00826.
(16) Li, Q.; Han, Z.; Wu, X.-M. Deeper insights into graph convolutional networks for semi-supervised learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32, 2018.
(17) Ji, Z.; Shi, R.; Lu, J.; Li, F.; Yang, Y. ReLMole: Molecular Representation Learning Based on Two-Level Graph Similarities. *J. Chem. Inf. Model* **2022**, *62* (22), 5361−5372.
(18) Brown, N.; Fiscato, M.; Segler, M. H.; Vaucher, A. C. GuacaMol: benchmarking models for de novo molecular design. *J. Chem. Inf. Model.* **2019**, *59* (3), 1096−1108.
(19) Kirkpatrick, P.; Ellis, C. Chemical space. *Nature* **2004**, *432* (7019), 823−824.
(20) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2012**, *40* (D1), D1100−D1107.
(21) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; et al. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* **2019**, *47* (D1), D1102−D1109.
(22) Cai, H.; Zhang, H.; Zhao, D.; Wu, J.; Wang, L. FP-GNN: a versatile deep learning architecture for enhanced molecular property prediction. *Briefings Bioinf.* **2022**, *23* (6), No. bbac408.
(23) Wu, J.; Su, Y.; Yang, A.; Ren, J.; Xiang, Y. An improved multi-modal representation-learning model based on fusion networks for property prediction in drug discovery. *Comput. Biol. Med.* **2023**, *165*, No. 107452.
(24) Zhou, Z.; Xu, H.; Hong, P. Multimodal Fusion with Relational Learning for Molecular Property Prediction, 2024, arXiv:2410.12128. arXiv.org e-Print archive. https://arxiv.org/abs/2410.12128.
(25) Guo, Z.; Yu, W.; Zhang, C.; Jiang, M.; Chawla, N. V.; et al. Graseq: graph and sequence fusion learning for molecular property prediction. *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* **2020**, 435−443.
(26) Sterling, T.; Irwin, J. J. ZINC 15−ligand discovery for everyone. *J. Chem. Inf. Model.* **2015**, *55* (11), 2324−2337.
(27) Wu, Z.; Ramsundar, B.; Feinberg, E. N.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K.; Pande, V. MoleculeNet: a benchmark for molecular machine learning. *Chem. Sci.* **2018**, *9* (2), 513−530.
(28) Subramanian, G.; Ramsundar, B.; Pande, V.; Denny, R. A. Computational modeling of β-secretase 1 (BACE-1) inhibitors using ligand based approaches. *J. Chem. Inf. Model.* **2016**, *56* (10), 1936−1949.
(29) Martins, I. F.; Teixeira, A. L.; Pinheiro, L.; Falcao, A. O. A Bayesian approach to in silico blood-brain barrier penetration modeling. *J. Chem. Inf. Model.* **2012**, *52* (6), 1686−1697.
(30) Gayvert, K. M.; Madhukar, N. S.; Elemento, O. A data-driven approach to predicting successes and failures of clinical trials. *Cell chemical biology* **2016**, *23* (10), 1294−1301.
(31) Sharakhov, I. V.; Artemov, G. N.; Bondarenko, S. M.; Shirokova, V.; Stegniy, V. N. Spatial Organization of Chromosomes in Malaria Mosquitoes, 2016.
(32) Richard, A. M.; Huang, R.; Waidyanatha, S.; Shinn, P.; Collins, B. J.; Thillainadarajah, I.; Grulke, C. M.; Williams, A. J.; Lougee, R. R.; Judson, R. S.; et al. The Tox21 10K compound library: collaborative chemistry advancing toxicology. *Chem. Res. Toxicol.* **2021**, *34* (2), 189−216.
(33) Kuhn, M.; Letunic, I.; Jensen, L. J.; Bork, P. The SIDER database of drugs and side effects. *Nucleic Acids Res.* **2016**, *44* (D1), D1075−D1079.
(34) Mobley, D. L.; Guthrie, J. P. FreeSolv: a database of experimental and calculated hydration free energies, with input files. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 711−720.
(35) Bento, A. P.; Hersey, A.; Felix, E.; Landrum, G.; Gaulton, A.; Atkinson, F.; Bellis, L. J.; De Veij, M.; Leach, A. R. An open source chemical structure curation pipeline using RDKit. *J. Cheminform* **2020**, *12* (1), 51.
(36) Honda, S.; Shi, S.; Ueda, H. R. Smiles Transformer: Pre-trained Molecular Fingerprint for Low Data Drug Discovery, 2019, arXiv:1911.04738. arXiv.org e-Print archive. https://arxiv.org/abs/1911.04738.
(37) Ho, T. K. Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*. Vol. 1, IEEE, 1995; pp 278−282.
(38) Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273 DOI: 10.1007/BF00994018.
(39) Zang, X.; Zhao, X.; Tang, B. Hierarchical molecular graph self-supervised learning for property prediction. *Commun. Chem.* **2023**, *6* (1), 34.
(40) Song, L.; Zhu, H.; Wang, K.; Li, M. LGGA-MPP: Local Geometry-Guided Graph Attention for Molecular Property Prediction. *J. Chem. Inf. Model.* **2024**, *64* (8), 3105−3113.
(41) Xia, J.; Zhao, C.; Hu, B.; Gao, Z.; Tan, C.; Liu, Y.; Li, S.; Li, S. Z. Mole-bert: Rethinking pre-training graph neural networks for molecules. 2023.
(42) Li, B.; Lin, M.; Chen, T.; Wang, L. FG-BERT: a generalized and self-supervised functional group-based molecular representation learning framework for properties prediction. *Briefings Bioinf.* **2023**, *24* (6), No. bbad398.
(43) Liu, P.; Ren, Y.; Tao, J.; Ren, Z. Git-mol: A multi-modal large language model for molecular science with graph, image, and text. *Comput. Biol. Med.* **2024**, *171*, No. 108073.
(44) Su, B.; Du, D.; Yang, Z.; Zhou, Y.; Li, J.; Rao, A.; Sun, H.; Lu, Z.; Wen, J.-R. A molecular multimodal foundation model associating molecule graphs with natural language, 2022, arXiv:2209.05481. arXiv.org e-Print archive. https://arxiv.org/abs/2209.05481.
(45) Luo, Y.; Yang, K.; Hong, M.; Liu, X.; Nie, Z. Molfm: A multimodal molecular foundation model, 2023, arXiv:2307.09484. arXiv.org e-Print archive. https://arxiv.org/abs/2307.09484.
(46) Zhang, H.; Wu, J.; Liu, S.; Han, S. A pre-trained multi-representation fusion network for molecular property prediction. *Information Fusion* **2024**, *103*, No. 102092.
(47) Zang, X.; Zhao, X.; Tang, B. Hierarchical Molecular Graph Self-Supervised Learning for property prediction. *Commun. Chem.* **2023**, *6* (1), 34.
(48) Aouichaoui, A. R. N.; Fan, F.; Mansouri, S. S.; Abildskov, J.; Sin, Gr. Combining Group-Contribution concept and graph neural networks toward interpretable molecular property models. *J. Chem. Inf. Model.* **2023**, *63* (3), 725−744.

https://doi.org/10.1021/acsomega.4c09884
ACS Omega 2025, 10, 12043−12053