Chinese Journal of Natural Medicines

MolP-PC: a multi-view fusion and multi-task learning framework for drug ADMET property prediction

Sishu Li, Jing Fan, Haiyang He, Ruifeng Zhou, Jun Liao

Citation: Sishu Li, Jing Fan, Haiyang He, Ruifeng Zhou, Jun Liao, MolP-PC: a multi-view fusion and multi-task learning framework for drug ADMET property prediction, Chinese Journal of Natural Medicines, 2025, 23(11), 1293-1300. doi: 10.1016/S1875-5364(25)60945-9.

View online: https://doi.org/10.1016/S1875-5364(25)60945-9

Chinese Journal of Natural Medicines 23 (2025) 1293-1300

Contents lists available at ScienceDirect

Chinese Journal of Natural Medicines

journal homepage: www.cjnmcpu.com/

Original article

MolP-PC: a multi-view fusion and multi-task learning framework for drug ADMET property prediction

Sishu Li[Δ], Jing Fan[Δ], Haiyang He, Ruifeng Zhou, Jun Liao[*]

School of Science, China Pharmaceutical University, Nanjing 211198, China

ARTICLE INFO

Article history:
Received 2 March 2025
Revised 21 April 2025
Accepted 20 June 2025
Available online 20 November 2025

Keywords:
Molecular ADMET prediction
Multi-view fusion
Attention mechanism
Multi-task deep learning.

ABSTRACT

The accurate prediction of drug absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties represents a crucial step in early drug development for reducing failure risk. Current deep learning approaches face challenges with data sparsity and information loss due to single-molecule representation limitations and isolated predictive tasks. This research proposes molecular properties prediction with parallel-view and collaborative learning (MolP-PC), a multi-view fusion and multi-task deep learning framework that integrates 1D molecular fingerprints (MFs), 2D molecular graphs, and 3D geometric representations, incorporating an attention-gated fusion mechanism and multi-task adaptive learning strategy for precise ADMET property predictions. Experimental results demonstrate that MolP-PC achieves optimal performance in 27 of 54 tasks, with its multi-task learning (MTL) mechanism significantly enhancing predictive performance on small-scale datasets and surpassing single-task models in 41 of 54 tasks. Additional ablation studies and interpretability analyses confirm the significance of multi-view fusion in capturing multi-dimensional molecular information and enhancing model generalization. A case study examining the anticancer compound Oroxylin A demonstrates MolP-PC's effective generalization in predicting key pharmacokinetic parameters such as half-life (T0.5) and clearance (CL), indicating its practical utility in drug modeling. However, the model exhibits a tendency to underestimate volume of distribution (VD), indicating potential for improvement in analyzing compounds with high tissue distribution. This study presents an efficient and interpretable approach for ADMET property prediction, establishing a novel framework for molecular optimization and risk assessment in drug development.

1. Introduction

Drug discovery represents a time-consuming, high-risk, costly, and complex process, wherein approximately 10% of candidate drugs successfully pass clinical trials and receive marketing approval [1,2]. A primary factor contributing to clinical trial failure involves deficiencies in drug absorption, distribution, metabolism, excretion, and toxicity (ADMET) [1]. Research indicates that accurate prediction of ADMET properties during early drug development stages substantially enhances research efficiency and minimizes late-stage failure risk [3]. Consequently, early assessment and optimization of ADMET properties constitute critical steps in new drug development.

Conventional ADMET evaluation methodologies predominantly rely on *in vitro* and *in vivo* experiments, which present time, cost, and ethical challenges [3,4]. The rapid progression of computational science and accumulation of ADMET experimental data have enabled computer-based prediction methods to emerge as cost-effective alternatives for ADMET research [5]. Among these approaches, quantitative structure-activity relationship (QSAR) models represent the most established computational methods. However, traditional QSAR models heavily rely on complex manual features, including molecular fingerprints (MFs) and descriptors, and require expert knowledge, potentially resulting in information loss and limited model adaptability [6]. Recent advances in deep learning methods, enabling data-driven automatic learning of molecular features, demonstrate potential to exceed traditional QSAR model performance, though their effectiveness remains dependent on molecular representation quality [7].

Current molecular representation methodologies broadly encompass 1D sequence representations, 2D graph-based representations, and 3D geometric representations. In 1D sequence representations, molecules are typically represented as chemical symbol sequences, such as the simplified molecular input line entry system (SMILES) [8,9] and molecular fingerprints (MFs) [10–12]. For instance, Pires et al. [13] developed pkCSM, integrating graph descriptors and pharmacophore fingerprints within a random forest model for multi-task ADMET prediction. Wu et al. [14] introduced a bidirectional long short-term memory (BiLSTM) network incorporating multi-step attention mechanisms, extracting latent features from SMILES strings to predict 11 molecular property tasks. 2D graph-based representations explicitly encode atomic (nodes) and chemical bond (edges) topological relationships to reflect molecular characteristics more directly [15]. For example, the MolGIN model by Peng et al. [16] achieved end-to-end ADMET property prediction through differential processing of

* Corresponding author.
E-mail address: liaojun@cpu.edu.cn
Δ These authors contributed equally to this work.

https://doi.org/10.1016/S1875-5364(25)60945-9

bond properties and atomic neighborhood effects. Xiong et al. [17] implemented a graph convolutional network (GCN) to construct fingerprint maps based on molecular graph representation, achieving state-of-the-art predictive performance in drug discovery tasks. 3D geometric representation depicts molecular structure through three-dimensional spatial coordinates, enabling more accurate stereochemical property description. Notable examples include the Drug3D-Net [18] model, which utilizes a spatiotemporal gated attention mechanism for processing 3D molecular representations, demonstrating effectiveness on the MoleculeNet benchmark dataset.

Contemporary deep learning methods for molecular property prediction predominantly utilize a single type of molecular representation, which constrains the completeness of captured molecular information. Different molecular representations characterize molecular information from distinct perspectives. 1D representations capture key functional group features [19], while 2D and 3D representations provide detailed topological and spatial geometric information, respectively [7,20]. Consequently, integrating multiple molecular representations enhances the extraction of comprehensive molecular features. Several studies have investigated combinations of multiple molecular representations to enhance model performance [21-25]. For example, the FP-GNN model developed by Cai et al. [24] integrated two-dimensional molecular graphs with three one-dimensional MFs and demonstrated superior performance in 16 experiments on 13 public datasets. Huang et al. [25] proposed MolMVC, a multi-view contrastive learning framework that employs attention-guided data augmentation and introduces an adaptive multi-view contrastive loss to effectively capture the 1D, 2D, and 3D structural features of molecules, significantly improving predictive performance across various downstream tasks, including molecular property prediction, drug-target binding affinity prediction, and cancer drug response prediction.

Furthermore, compared with large-scale datasets in computer vision and natural language processing, datasets related to drug ADMET properties remain relatively limited. The availability of ADMET data is constrained by expensive and time-consuming experimental procedures, creating a data bottleneck that affects the generalization ability and predictive performance of deep learning models. Multi-task learning (MTL) effectively mitigates overfitting caused by data scarcity by sharing information and extracting features across different tasks, thereby improving model robustness and predictive accuracy [26].

In light of these challenges, this study proposes a multi-view fusion and multi-task deep learning framework, molecular properties prediction with parallel-view and collaborative learning (MolP-PC), which integrates 1D MFs, 2D molecular graphs, and 3D geometric representations to comprehensively capture molecular features. Through the incorporation of an attention-gated fusion mechanism and a multi-task adaptive learning strategy, the framework achieves accurate prediction of drug ADMET properties.

2. Dataset preparation

The ADMET data utilized by the MolP-PC model were integrated from multiple authoritative public sources, including MoleculeNet [27], the Therapeutic Data Commons (TDC) database [28], ADMETNet [29], Tox21 [30], and peer-reviewed literature. To ensure data quality, consistency, and reliability, several preprocessing steps were implemented: standardized SMILES strings were uniformly used as input formats to eliminate isomorphic ambiguity, remove duplicate entries, remove molecules containing metal atoms, and remove mixture compounds that do not conform to drug-like properties. Following these preprocessing steps, a multi-task dataset was constructed, comprising 274 192 molecules covering 54 ADMET properties. The detailed statistics of the datasets and data partitioning strategies are provided in Supplementary Table S1.

3. Methods

Fig. 1 presents the overall architecture of MolP-PC, which consists of three main modules: multi-view feature extraction, multi-view feature fusion, and multi-task prediction. Specifically,

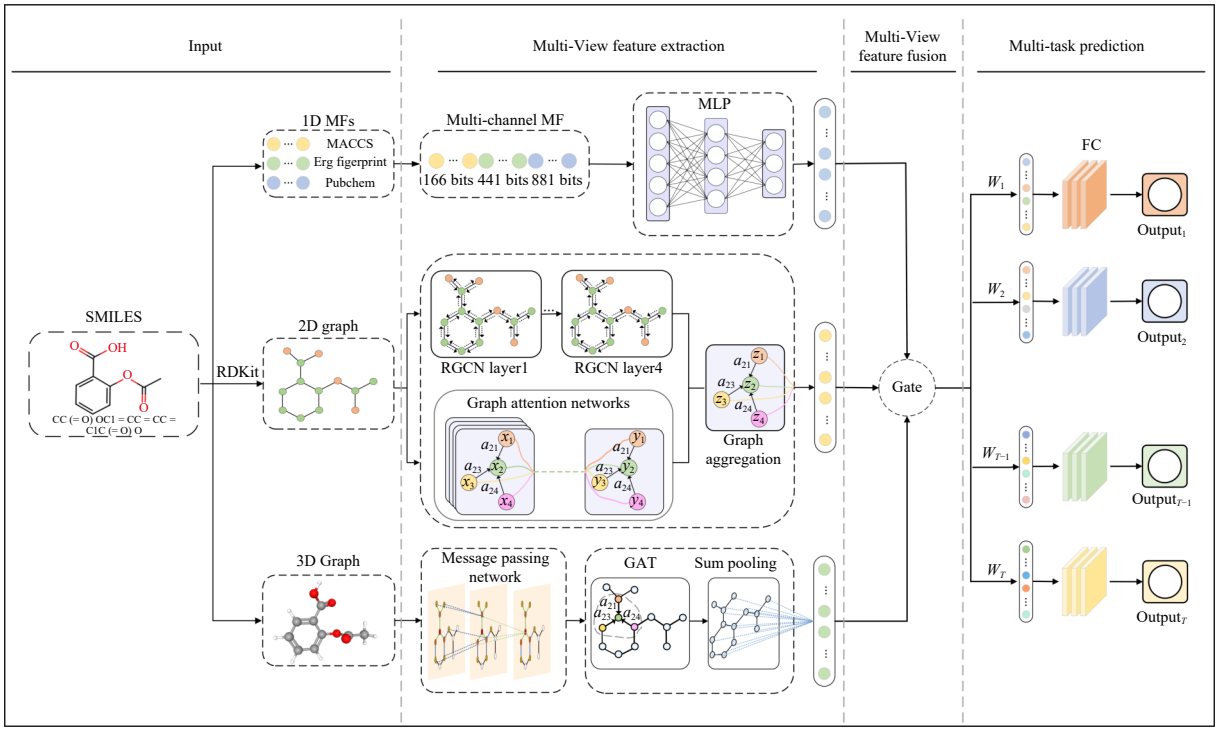

Fig. 1 The overall workflow of the MolP-PC architecture. Starting from SMILES, 1D fingerprints, 2D, and 3D features are sequentially extracted and fused via an attention-gated mechanism, with adaptive weighting generating task-specific representations for ADMET prediction, followed by multi-task prediction using an FC network.

the multi-view feature extraction module consists of a multi-layer perceptron (MLP), relational graph convolution (RGCN), geometric message passing networks, and 2D/3D graph attention networks (GAT). It extracts molecular features from chemical properties, topological structures, and stereo conformations through 1D MFs, 2D molecular graphs, and 3D geometric representations, respectively, and fuses the feature information from different views through the attention-gated fusion mechanism of the multi-view feature fusion module. The multi-task prediction module employs an adaptive weight allocation mechanism to generate specific features for each ADMET property prediction task and outputs the predicted values through fully connected layers (FC).

3.1. Feature extraction module based on 1D molecular fingerprint

MFs characterize molecular structural features by encoding key substructures, topological structures, paths, or circular features of molecules into fixed-length binary vectors. A single MF cannot capture all important structures or characteristics of a molecule [31]. This study employs a combination of MACCS fingerprints [10], Pharmacophore ErG fingerprints [32], and PubChem fingerprints [33], as they complement each other and holistically represent molecular features [34], thus enabling more comprehensive learning of molecular structural features and enhancing molecular property prediction accuracy. MACCS fingerprints represent molecules with 166 predefined functional groups/chemical bonds/chemical structures, using binary encoding to indicate the presence or absence of substructures. Pharmacophore ErG fingerprints utilize 2D pharmacophore fingerprint maps that combine chemical and structural information and employ pharmacophore-type nodes to encode relevant molecular features. PubChem fingerprints, with 881 predefined chemical structures, capture information on topology, bonds, and substructures more extensively.

Three specific fingerprint-encoding methods were employed to obtain MFs: MACCS, Pharmacophore ErG, and PubChem. These three fingerprint vectors are concatenated to create a hybrid fingerprint representation FP. To analyze complex nonlinear relationships within FP, the data were processed through an MLP network, generating the 1D feature representation H_{1D}, as expressed in the following equation:

$$H_{1D} = MLP(FP) = W \cdot \sigma([FP_{MACCS}||FP_{ErG}||FP_{PubChem}]) + b \quad (1)$$

Where W and b are learnable parameters, σ represents the $ReLU$ activation function, and $||$ denotes vector concatenation.

3.2. Feature extraction module based on 2D molecular graph representation

In this study, molecular graphs $G = (V, E)$ are generated from SMILES codes using the Deep Graph Library (DGL, https://www.dgl.ai/), with node sets $V = \{v_i\}_{i=1}^{n}$ representing atom sets and edge sets $E = \{e_{ij}\}_{i=1}^{m}$ representing chemical bond sets. Atomic features (such as atom type, hybridization, and aromaticity) and bond features (including bond type and conjugation) are extracted using the RDKit library, forming node feature matrices $X_v \in R^{n \times d_v}$ and edge feature matrices $X_e \in R^{m \times d_e}$, respectively, which are stored in the molecular graph.

For comprehensive capture of molecular graph structural characteristics, this study implements a parallel architecture combining RGCN and GAT, facilitating effective extraction of both atomic and bond-level information. RGCN enhances traditional GCN by incorporating different chemical bond types during message passing. It utilizes learnable relation-specific transformation matrices for each bond type, enabling more expressive representation learning in heterogeneous molecular graphs [35]. The implementation includes a 4-layer RGCN with inter-layer residual connections to stabilize information flow, producing updated node representations $H_{rgcn} \in R^{n \times d_h}$. The node update mechanism for the l layer of RGCN is expressed as:

$$h_i^{l+1} = \sigma\left(W_0^{(l)} h_i^{(l)} + \sum_{r \in R} \sum_{j \in N_i^r} W_r^{(l)} h_j^{(l)} / c_{i,r}\right) \quad (2)$$

Where h_i^{l+1} is the state vector of node v_i after l iterations, N_i^r is the set of neighboring nodes of node v_i connected by edges of type r, $c_{i,r} = |N_i^r|$ is a normalization coefficient, and $W_r^{(l)}$ is the learnable weight matrix for edges of type r.

GAT enhances the extraction of key structural information in molecular graphs by assigning different attention coefficients to neighboring nodes of each node, thereby reflecting the relative importance of neighboring nodes to the central node. For node v_i and its neighborhood N_i, the attention coefficient calculation proceeds as follows:

$$a_{ij} = \frac{\exp(LeakyReLU(a_T[Wh_i||Wh_j]))}{\sum_{l \in N_i} \exp(LeakyReLU(a_T[Wh_i||Wh_l]))} \quad (3)$$

Where a_{ij}^k is the attention coefficient of node v_i for its neighbor node v_j, W is a learnable weight matrix, and a_T is an attention vector. To ensure stable learning of more meaningful information, multi-head attention mechanisms are used in parallel, ultimately obtaining enhanced node representations $H_{gat} \in R^{n \times d_h}$:

$$H_{gat} = ||_{k=1}^{k}\left(\sigma\left(\sum_{j \in N_j} \alpha_{ij}^k W_k h_j\right)\right) \quad (4)$$

Where k is the number of attention heads, α_{ij}^k is the attention coefficient calculated by the k attention head, W_k is the weight matrix for the linear transformation of the k attention head, σ represents the activation function, and $||$ represents the concatenation operation.

The integration of RGCN and GAT leverages the complementary strengths of relational modeling and adaptive attention mechanisms, facilitating the extraction of more informative and comprehensive 2D molecular representations. The final 2D representation H_{2D} is obtained through the summation of node-level outputs from both models:

$$H_{2D} = Sum(H_{rgcn}, H_{gat}) \quad (5)$$

3.3. Feature extraction module based on 3D geometric representation

In this study, the 3D geometric representation is defined as $G_{3D} = (V, P)$, where $V = \{v_i\}_{i=1}^{n}$ is the node set representing the atom set, $\{p_i\}_{i=1}^{n}$ is the 3D atomic coordinate set. Each atom corresponds to a point in space, and can capture the covalent atomic connectivity and molecular conformation preference information [36]. To generate high-quality and representative 3D molecular structures, this study employs the RDKit library. Initially, 3D conformers are generated for each molecule using the ETKDG algorithm to ensure diversity in spatial configurations. These conformers are then optimized using the universal force field (UFF) to minimize their potential energy. Finally, the conformer with the lowest energy is selected as the final 3D structure to ensure geometric consistency and chemical plausibility.

To extract 3D structural features, a geometric message passing network enhanced with a multi-head attention mechanism was employed. This network aggregates information from neighboring atoms based on their features, bond attributes, and interatomic distances θ_{ij}, which function as rotation- and translation-invariant geometric descriptors, ensuring robustness to spatial transformations. The message passing procedure is defined as follows:

$$m_i^{(l)} = \frac{\sum_{j \in N_i} W_{ij} \cdot \phi_e \left(h_i^{(l)}, h_j^{(l)}, e_{ij}, \theta_{ij} \right)}{|N_i|} \quad (6)$$

$$h_i^{(l+1)} = MLP \left(h_i^{(l)} + m_i^{(l)} \right) \quad (7)$$

Where ϕ_e is an MLP constructing the message passing function, and W_{ij} is a learnable weight matrix. Residual connections are incorporated to maintain stable feature propagation. A multi-head attention mechanism is subsequently applied to enhance node representations by adaptively weighting contributions from neighboring atoms. The output comprises the 3D molecular feature representation H_{3D}.

3.4. Multi-view feature fusion module

To effectively utilize the complementary information across different views of molecular structure, this study implements a multi-view feature fusion module based on the self-attention mechanism, enabling a dynamic weighted combination of features from various structural perspectives. The molecular feature representations from the three distinct views are initially concatenated into a three-dimensional tensor:

$$Mol = [H_{1D}, H_{2D}, H_{3D}] \in R^{B \times N \times D} \quad (8)$$

Where B denotes the batch size, $N = 3$ represents the number of perspectives, and D corresponds to the feature dimension for each perspective.

To facilitate interactions among the features, learnable linear transformation parameters are applied to project the input features, generating query (Q), key (K), and value (V) matrices:

$$\mathrm{Q} = \mathrm{W_Q} \cdot \mathrm{Mol} \in R^{B \times N \times D}, \mathrm{K} = \mathrm{W_K} \cdot \mathrm{Mol} \in R^{B \times N \times D}, \mathrm{V} = \mathrm{W_V} \cdot \mathrm{Mol} \in R^{B \times N \times D} \quad (9)$$

The scaled dot-product attention mechanism calculates the correlation matrix between different perspectives:

$$A_{ij} = Softmax \left(\frac{\left(Q \cdot [K]^T \right)}{\sqrt{d_k} \cdot \tau} \right) \in R^{B \times 3 \times 3} \quad (10)$$

Where d_k represents the dimensionality of the key vector, which prevents gradient instability from large dot-product values. A temperature scaling factor $\tau = 2$ and batch normalization (BatchNorm) are implemented to adjust the attention distribution, enhancing model robustness and generalization capability.

Subsequently, the fused feature matrix $H_{fused} = A \cdot V$ is generated by weighting the value vectors V *via* the attention weights A_{ij}, and the final integrated molecular feature H_{mol} is obtained by performing mean pooling along the view dimension:

$$H_{mol} = Mean(H_{fused}, dim = 1) \quad (11)$$

The attention weights A_{ij} maintain global sharing across all tasks, utilizing the same attention mechanism throughout. This approach enables joint learning of molecular features in MTL, facilitating the sharing of beneficial feature information across different tasks.

3.5. Multi-task prediction module

An attention-based adaptive multi-task prediction module addresses heterogeneous tasks in ADMET prediction, encompassing both classification and regression. This module utilizes shared molecular representations while determining task-specific feature importance for individual task predictions.

For each task $t \in \{1, 2, \ldots, T\}$, a dedicated attention submodule calculates task-specific importance weights for the input features. This submodule assigns weights to feature dimensions based on their relevance to the current task. The attention weights are derived using a linear layer followed by a Sigmoid activation function:

$$\alpha^t = \sigma(W^t H_{mol}) \quad (12)$$

Where $W^t \in R^{D \times 1}$ denotes the learnable weight parameters for task t, and $\sigma(\cdot)$ represents the Sigmoid function.

The task-specific feature representation is calculated through element-wise multiplication of the attention weights with the fused molecular features:

$$H^t = \alpha^t \odot H_{mol} \quad (13)$$

Each task-specific representation H^t is subsequently processed through a dedicated fully connected layer FC^t to predict the corresponding ADMET property.

3.6. Loss function and evaluation

Prior research [37] has shown that for datasets containing fewer than 1000 compounds, structure-based splitting methods — such as scaffold splitting—can intensify data and label imbalance across folds, potentially resulting in insufficient label coverage and unstable model training. Given this consideration, and noting that 10 datasets in this study contain fewer than 1000 compounds, this work implements a stratified random split strategy. Each dataset is divided into training, validation, and test sets using an 8 : 1 : 1 ratio, ensuring balanced label distribution across all subsets.

As the prediction tasks encompass both classification and regression, appropriate loss functions are implemented: binary cross-entropy loss (BCEWithLogitsLoss) for classification tasks, and mean squared error loss (MSELoss) for regression tasks, defined as:

$$L_c = \sum_{c=1}^{C} \sum_{n=1}^{N} \left(- [p_c y_{n,c} log\sigma(x_{n,c}) + (1 - y_{n,c}) log(1 - \sigma(x_{n,c}))] \right) \quad (14)$$

$$L_r = \sum_{r=1}^{R} \sum_{n=1}^{N} (x_{n,r} - y_{n,r})^2 \quad (15)$$

Where C and R denote the number of classification and regression tasks respectively, N is the number of molecules, $x_{n,C}$ and $x_{n,r}$ are the predicted values, $y_{n,c}$ and $y_{n,r}$ are the corresponding ground truths, σ is the sigmoid activation function, and p_c represents the positive sample weighting factor designed to mitigate the impact of class imbalance in classification tasks. The final loss is computed as a combination of L_c and L_r, enabling joint optimization and effective multi-task learning.

Model performance evaluation utilized the area under the curve (AUC) and R-Square (coefficient of determination). The receiver operating characteristic (ROC) curve plots the false positive rate (FPR) against the true positive rate (TPR) at various classification thresholds. A larger area under the ROC curve indicates superior classification model performance. R-Square measures the fit quality through data variation analysis. R-Square values range from 0 to 1, with values approaching 1 indicating a stronger model-data fit, while values closer to 0 suggest a weaker fit.

4. Results

4.1. Comparison with baseline models

This study conducted a comprehensive evaluation of the MolP-PC framework across 54 ADMET prediction tasks, compris-

ing 42 classification and 12 regression tasks. As presented in Table 1, MolP-PC exhibited robust overall performance, attaining an average ROC-AUC of 0.8888 for classification tasks and an average R^2 of 0.6825 for regression tasks. The framework surpassed all baseline models in 27 tasks (23 classification and 4 regression), with comprehensive results available in Supplementary Table S2. MolP-PC demonstrated particularly strong performance in metabolism-related classification tasks, achieving exceptional ROC-AUC scores in predicting inhibitors of crucial enzymes, including CYP1A2, CYP2C19, and CYP3A4. In regression tasks, MolP-PC exhibited notable superiority in ecotoxicity-related predictions, specifically for LC50 and BCF, demonstrating its effectiveness in modeling continuous-valued endpoints.

In comparison with the single-task version MolP-PC_single, the MTL version of MolP-PC demonstrated superior performance in 41 out of 54 tasks, particularly in tasks with limited data availability. For instance, in the CYP2C9 substrate prediction task, the multi-task version achieved a ROC-AUC of 0.8735, representing a 21.6% improvement over the single-task version. These results substantiate the effectiveness of MTL in identifying task correlations and addressing data sparsity issues. However, for specific tasks including blood-brain barrier (BBB) permeability, Pgp-substrate prediction, and Caco2 permeability, MolP-PC showed reduced predictive performance compared to the single-task version, potentially due to the distinct data distribution and specific chemical properties associated with these tasks.

In conclusion, the study findings confirm that the MolP-PC framework, through its effective MTL mechanism, successfully integrates correlation information across various ADMET tasks, enhancing prediction performance for small sample size tasks while maintaining model generalization capabilities.

4.2. Ablation study

To evaluate the efficacy of multi-view feature fusion comprehensively, ablation experiments were conducted comparing predictive performance across various tasks using individual molecular representations (1D, 2D, and 3D), partial fusion strategies (1D + 2D, 1D + 3D, and 2D + 3D), and the full fusion strategy (1D + 2D + 3D, denoted as MolP-PC). As illustrated in Fig. 2, the MolP-PC model outperformed other strategies in 23 tasks (15 classification and 8 regression tasks), underscoring the benefits of integrating multi-dimensional molecular information. These results indicate that multi-view fusion effectively captures complementary structural and physicochemical features, resulting in enhanced predictive accuracy compared to single-view or partially fused approaches. Detailed performance metrics are available in Supplementary Table S3.

4.3. Analysis of fusion strategy

MolP-PC utilizes an attention-gated mechanism for dynamic integration of features extracted from different molecular dimensions. This study visualized the average attention weights across all molecules (Fig. 3), where attention flows from rows to columns, indicating each feature's influence on others during feature fusion. The heatmap reveals that while the model considers all three molecular representations (1D, 2D, and 3D), it predominantly relies on 1D features. The superior performance achieved through the fusion of 1D, 2D, and 3D molecular representations suggests that structural information from 2D and 3D features effectively complements 1D features, ultimately improving overall predictive performance. For detailed fusion weights of the multi-view fusion, refer to Fig. S1 in the supplementary material.

4.4. Model interpretability analysis

This study utilized the integrated gradients algorithm and attention weight analysis to examine the MolP-PC multi-view feature extraction module's capability in identifying key molecular substructures. Using inactive molecules from the BBB dataset as an example, the analysis results (Fig. 4) demonstrate that different molecular representations capture distinct chemical properties affecting BBB permeability: The 1D molecular sequence representation, encoded *via* substructure patterns, identifies chemically relevant local polar substructures containing nitrogen and oxygen. The 2D molecular graph, through modeling atomic connectivity, captures topological features such as sugar rings and ether bonds connecting to aromatic rings. In contrast, the 3D molecular representation, which incorporates spatial coordinates and interatomic distances, enables the analysis of stereochemical information, including hydroxyl and amino groups, as well as the spatial distribution of functional groups. These findings validate that the multi-view feature fusion strategy leverages the comple-

Table 1 Average performance of each model on the five major ADMET task categories. Best performance per row is highlighted in bold.

Category	Metric	MolP-PC	MolP-PC_single	Chemprop_rdkit [38]	FP-GNN [24]	MGA [39]	MolMVC [25]
Absorption	Avg ROC-AUC	0.8674	0.8243	0.8431	0.8183	0.8547	**0.8892**
	Avg R2	0.7353	**0.7636**	0.7311	0.7451	0.6918	0.7042
Distribution	Avg ROC-AUC	0.8471	0.9499	0.945	**0.961**	0.897	0.9558
	Avg R2	0.4494	0.4399	0.4731	0.4789	0.4098	**0.4835**
Excretion	Avg ROC-AUC	0.785	0.6047	**0.8033**	0.7573	0.7126	0.7492
	Avg R2	**0.6968**	0.6917	0.5833	0.5852	0.6013	0.5634
Metabolism	Avg ROC-AUC	**0.7604**	0.6918	0.7152	0.7013	0.7055	0.7332
Toxicology	Avg ROC-AUC	**0.8903**	0.8649	0.8689	0.8404	0.8532	0.8849
	Avg R2	**0.801**	0.7287	0.7616	0.7463	0.716	0.712
Overall	Avg ROC-AUC	**0.8888**	0.8466	0.8629	0.8384	0.8495	0.8827
	Avg R2	**0.6825**	0.6651	0.6644	0.6656	0.6218	0.6399
	Best Tasks (Classification/Regression)	**27(23/4)**	5(3/2)	7(4/3)	7(6/1)	2(2/0)	6(4/2)

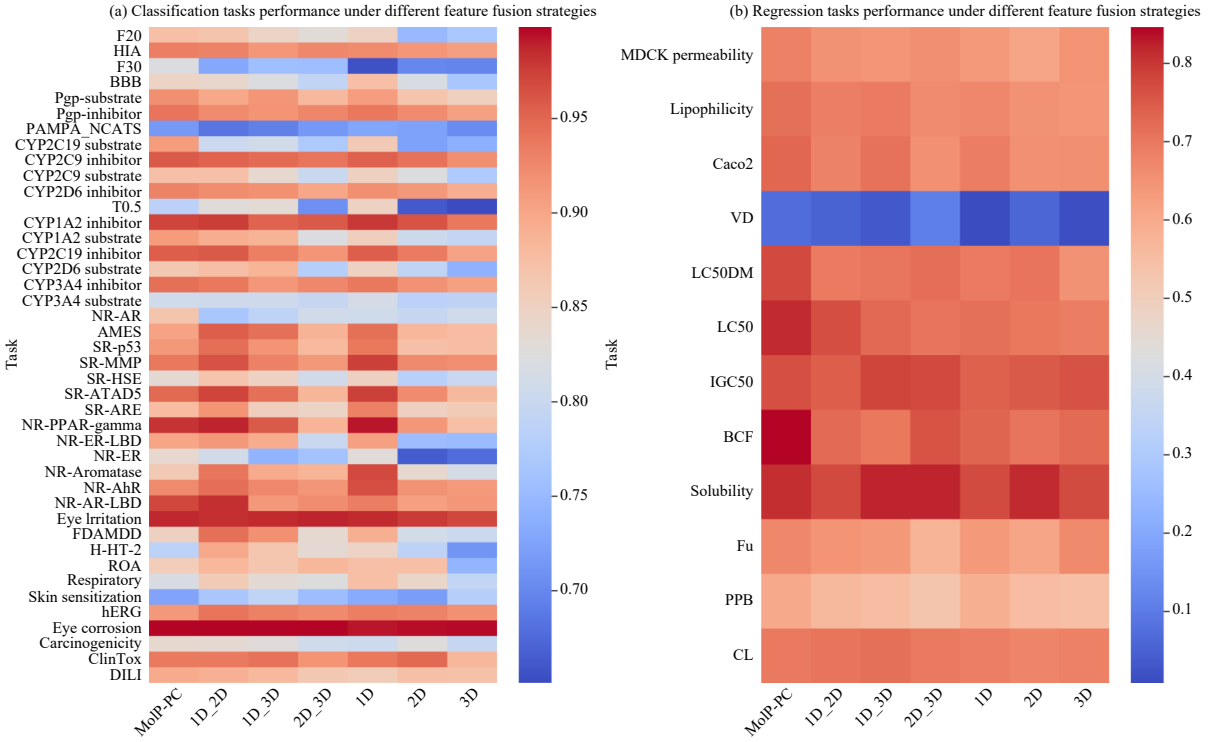

Fig. 2 A comparison of the performance of the model on classification and regression tasks under different ablation strategies.

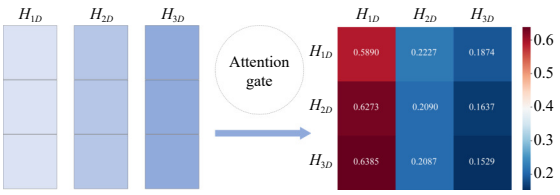

Fig. 3 Visualization of attention weights in feature fusion across three dimensions in the MolP-PC framework.

mentary advantages of different molecular representations, enabling accurate predictions of complex molecular properties and explaining the molecule's inability to cross the BBB.

4.5. Experimental validation

To further verify the reliability of the proposed model's predictive performance, this study selected experimentally determined ADMET properties of Oroxylin A for evaluation. Oroxylin A is an active flavonoid compound with potent anticancer effects. It can inhibit the IL-6/STAT3 pathway and NF-κB signaling, suppress cell proliferation, and induce apoptosis. Additionally, Oroxylin A has been shown to inhibit colitis-associated carcinogenesis [40-44]. In this study, LC-MS/MS analysis was performed using a SHIMADZU HPLC system coupled with an AB SCIEX mass spectrometer in negative ion mode, with chlorzoxazone as the internal standard. After oral administration of Oroxylin A (10 mg·kg^{-1}) to SD rats, plasma and tissue samples (heart, liver, kidney, etc.) were collected at multiple time points (0−72h). Samples were extracted using ethyl acetate and analyzed with WinNonlin 6.4 to calculate the pharmacokinetic parameters T0.5, clearance (CL), and volume of distribution (VD).

As demonstrated in Table 2, the MolP-PC model effectively captured the moderate metabolic stability of Oroxylin A (T0.5 > 3 h), and the predicted CL value approximated the experimental value, indicating robust generalization capability. Regarding VD, although MolP-PC substantially underestimated the experimental value, this limitation is consistently observed across baseline models, suggesting that the current model requires further refinement in modeling compounds with high tissue distribution. Subsequent research could enhance the model by incorporating additional structural and physicochemical property features to improve prediction accuracy for complex pharmacokinetic parameters.

5. Conclusion

This study introduces a multi-view fusion multi-task deep learning framework, MolP-PC, which combines 1D MFs, 2D molecular graphs, and 3D geometric representations. Through the implementation of an attention-gated fusion mechanism and a multi-task adaptive learning strategy, the framework effectively addresses data sparsity challenges. Experimental results reveal that MolP-PC achieves superior performance in 27 of 54 tasks, with MTL mitigating data sparsity and surpassing single-task models in 41 of 54 tasks, particularly in small-scale datasets. Subsequent ablation studies and interpretability analyses confirm the efficacy of the multi-view feature fusion strategy, demonstrating that this approach optimally utilizes the complementary advantages of chemical properties, topological relationships, and stereochemical configurations of molecular structures. This integration enhances the model's capacity to analyze complex molecular characteristics. A case study on Oroxylin A indicates that MolP-PC demonstrates strong generalization ability in predicting pharmacokinetic parameters such as T0.5 and CL, highlighting its practical applications in molecular modeling. However, the model exhibits limitations in specific tasks, including BBB, Pgp-substrate prediction, and Caco-2 permeability, potentially due to complex, task-specific mechanisms or data distribution patterns. Moreover, a significant bias is observed in VD prediction. The model's performance is additionally constrained by static 3D conformations and class imbalance limitations. Future research directions may include exploring dynamic conformation ensembles, domain-specific inductive biases, advanced data balancing techniques, and refined task modeling strategies to enhance model performance and generalization capability. The code for the MolP-

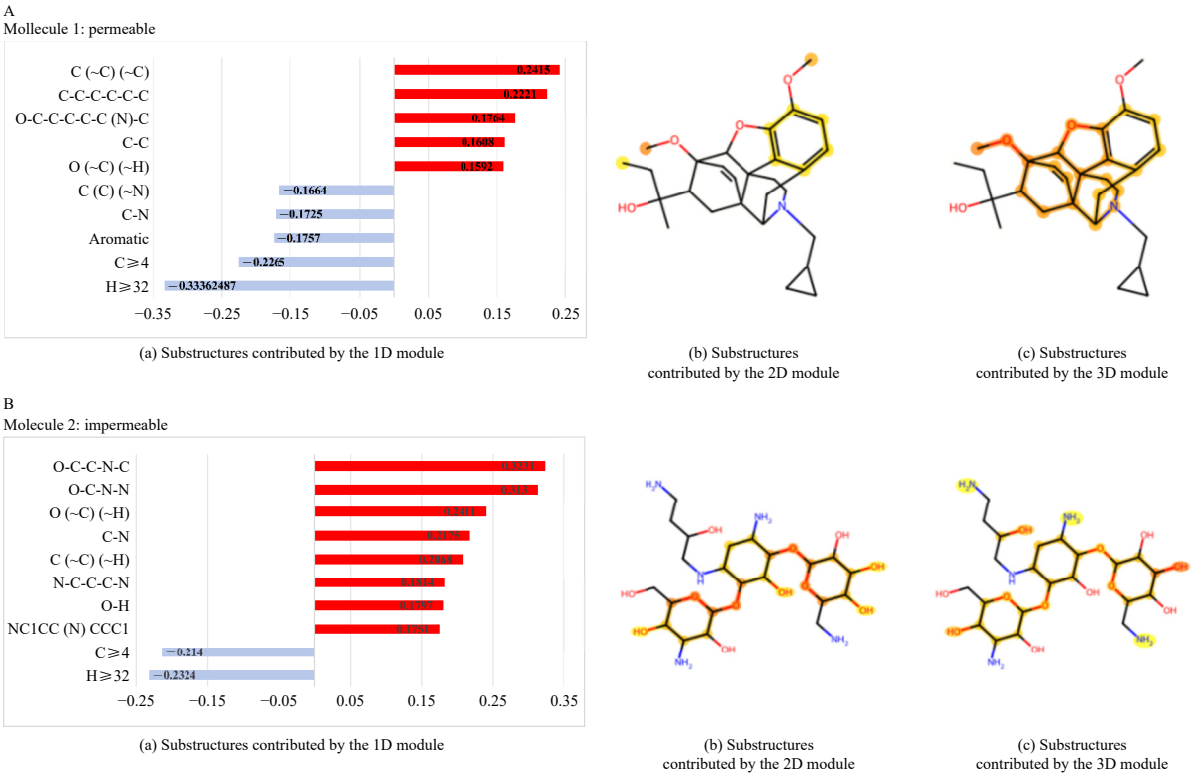

Fig. 4 The Multi-view feature extraction module in MolP-PC learns important molecular structures through a fusion strategy based on 1D, 2D, and 3D representations.

Table 2 Comparison of experimental values and predicted results for oroxylin A pharmacokinetic parameters across different models.

Property	Experimental Value	MolP-PC	Chemprop_rdkit [38]	FP-GNN [24]	MGA [39]	MolMVC [25]
$t_{1/2}$ (h)	7.69 ± 0.84	> 3	< 3	< 3	> 3	> 3
CL (mL·min^{-1}·kg^{-1})	14.48 ± 1	10.60	9.80	6.28	4.25	6.11
VD (L·kg^{-1})	9.70 ± 1.74	2.16	2.68	1.51	2.50	1.85

PC model is publicly available at: https://github.com/molxxp/MolP_PC.

Funding

This work was supported by the research on key technologies for monitoring and identifying drug abuse of anesthetic drugs and psychotropic drugs, and intervention for addiction (No. 2023YFC3304200), the program of a study on the diagnosis of addiction to synthetic cannabinoids and methods of assessing the risk of abuse (No.2022YFC3300905), and the program of Ab initio design and generation of AI models for small molecule ligands based on target structures (No. 2022PE0AC03) and ZHIJIANG LAB.

Acknowledgements

We thank the High Performance Computing Center, China Pharmaceutical University, and all those who have helped with this work

Declaration of competing interest

These authors have no conflict of interest to declare.

References

1 Pammolli F, Magazzini L, Riccaboni M, et al. The productivity crisis in pharmaceutical R&D. *Nat Rev Drug Discov*. 2011;10(6):428-438. https://doi.org/10.1038/nrd3405.
2 Sun D, Gao W, Hu H, et al. Why 90% of clinical drug development fails and how to improve it? *Acta Pharm Sin B*. 2022;12(7):3049-3062. https://doi.org/10.1016/j.apsb.2022.02.002.
3 Wang Y, Xing J, Xu Y, et al. In silico ADME/T modelling for rational drug design. *Q Rev Biophys*. 2015;48(4):488-515. https://doi.org/10.1017/S0033583515000190.
4 Wei Y, Li S, Li Z, et al. Interpretable-ADMET: a web service for ADMET prediction and optimization based on deep neural representation. *Bioinformatics*. 2022;38(10):2863-2871. https://doi.org/10.1093/bioinformatics/btac192.
5 Jia CY, Li JY, Hao GF, et al. A drug-likeness toolbox facilitates ADMET study in drug discovery. *Drug Discov Today*. 2020;25(1):248-258. https://doi.org/10.1016/j.drudis.2019.10.014.
6 Cherkasov A, Muratov EN, Fourches D, et al. QSAR modeling: where have you been? where are you going to? *J Med Chem*. 2014;57(12):4977-5010. https://doi.org/10.1021/jm4004285.
7 Li Z, Jiang M, Wang S, et al. Deep learning methods for molecular representation and property prediction. *Drug Discov Today*. 2022;27(12):103373. https://doi.org/10.1016/j.drudis.2022.103373.
8 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. https://doi.org/10.1021/ci00057a005.
9 Weininger D, Weininger A, Weininger JL, et al. SMILES. 2. Algorithm for generation of unique SMILES notation. *J Chem Inf Comput Sci*, 1989;29(2):97-101. https://doi.org/10.1021/ci00062a008.
10 Durant JL, Leland BA, Henry DR, et al. Reoptimization of MDL keys for use in drug discovery. *J Chem Inf Comput Sci*. 2002;42(6):1273-1280. https://doi.org/10.1021/ci010132r.
11 Cereto-Massagué A, Ojeda MJ, Valls C, et al. Molecular fingerprint similarity search in virtual screening. *Methods*. 2015;71:58-63. https://doi.org/10.1016/j.ymeth.2014.08.005.
12 Steffen A, Kogej T, Tyrchan C, et al. Comparison of molecular fingerprint methods on the basis of biological profile data. *J Chem Inf Model*. 2009;49(2):338-347. https://doi.org/10.1021/ci800326z.
13 Pires DEV, Blundell TL, Ascher DB, et al. pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J Med*

Chem, 2015;58(9):4066-4072. https://doi.org/10.1021/acs.jmedchem.5b00104.
14 Wu CK, Zhang XC, Yang ZJ, et al. Learning to SMILES: BAN-based strategies to improve latent representation learning from molecules. *Brief Bioinform*. 2021;22(6):bbab327. https://doi.org/10.1093/bib/bbab327.
15 Fang Y, Pan X, Shen HB. *De novo* drug design by iterative multiobjective deep reinforcement learning with graph-based molecular quality assessment. *Bioinformatics*. 2023;39(4):btad157. https://doi.org/10.1093/bioinformatics/btad157.
16 Peng Y, Lin Y, Jing XY, et al. Enhanced graph isomorphism network for molecular ADMET properties prediction. *IEEE Access*. 2020;8:168344-168360. https://doi.org/10.1109/ACCESS.2020.3022850.
17 Xiong Z, Wang D, Liu X, et al. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *J Med Chem*. 2020;63(16):8749-8760. https://doi.org/10.1021/acs.jmedchem.9b00959.
18 Li C, Wang J, Niu Z, et al. A spatial-temporal gated attention module for molecular property prediction based on molecular geometry. *Brief Bioinform*. 2021;22(5):bbab078. https://doi.org/10.1093/bib/bbab078.
19 Lin X, Dai L, Zhou Y, et al. Comprehensive evaluation of deep and graph learning on drug–drug interactions prediction. *Brief Bioinform*. 2023;24(4):bbad235. https://doi.org/10.1093/bib/bbad235.
20 Zhong S, Zhang K, Wang D, et al. Shedding light on "Black Box" machine learning models for predicting the reactivity of HO radicals toward organic compounds. *Chem Eng J*. 2021;405:126627. https://doi.org/10.1016/j.cej.2020.126627.
21 Tang Q, Nie F, Zhao Q, et al. A merged molecular representation deep learning method for blood–brain barrier permeability prediction. *Brief Bioinform*. 2022;23(5):bbac357. https://doi.org/10.1093/bib/bbac357.
22 Karim A, Riahi V, Mishra A, et al. Quantitative toxicity prediction *via* meta ensembling of multitask deep learning models. *ACS Omega*. 2021;6(18):12306-12317. https://doi.org/10.1021/acsomega.1c01247.
23 Guo Z, Yu W, Zhang C, et al. GraSeq: graph and sequence fusion learning for molecular property prediction. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020;435-443. https://doi.org/10.1145/3340531.3411981.
24 Cai H, Zhang H, Zhao D, et al. FP-GNN: a versatile deep learning architecture for enhanced molecular property prediction. *Brief Bioinform*. 2022;23(6):bbac408. https://doi.org/10.1093/bib/bbac408.
25 Huang Z, Fan Z, Shen S, et al. MolMVC: enhancing molecular representations for drug-related tasks through multi-view contrastive learning. *Bioinformatics*. 2024;40(Suppl 2):ii190-ii197. https://doi.org/10.1093/bioinformatics/btae386.
26 Zhang Y, Yang Q. An overview of multi-task learning. *Natl Sci Rev*. 2018;5(1):30-43. https://doi.org/10.1093/nsr/nwx105.
27 Wu Z, Ramsundar B, Feinberg EN, et al. MoleculeNet: a benchmark for molecular machine learning. *Chem Sci*. 2018;9(2):513-530. https://doi.org/10.1039/C7SC02664A.
28 Huang K, Fu T, Gao W, et al. Therapeutics data commons: machine learning datasets and tasks for drug discovery and development. In: NeurIPS Datasets and Benchmarks Track. 2021. https://zitniklab.hms.harvard.edu/publications/papers/TDC-neurips21-main.pdf.
29 Xu Q, Liu K, Lin X, et al. ADMETNet: the knowledge base of pharmacokinetics and toxicology network. *J Genet Genomics*. 2017;44(5):273-276. https://doi.org/10.1016/j.jgg.2017.04.005.
30 Richard AM, Huang R, Waidyanatha S, et al. The Tox21 10K compound library: collaborative chemistry advancing toxicology. *Chem Res Toxicol*. 2021;34(2):189-216. https://doi.org/10.1021/acs.chemrestox.0c00264.
31 Yang J, Cai Y, Zhao K, et al. Concepts and applications of chemical fingerprint for hit and lead screening. *Drug Discov Today*. 2022;27(11):103356. https://doi.org/10.1016/j.drudis.2022.103356.
32 Stiefl N, Watson IA, Baumann K, et al. ErG: 2D pharmacophore descriptions for scaffold hopping. *J Chem Inf Model*. 2006;46(1):208-220. https://doi.org/10.1021/ci050457y.
33 Shen WX, Zeng X, Zhu F, et al. Out-of-the-box deep learning prediction of pharmaceutical properties by broadly learned knowledge-based molecular representations. *Nat Mach Intell*. 2021;3(4):334-343. https://doi.org/10.1038/s42256-021-00301-6.
34 Bolton EE, Wang Y, Thiessen PA, et al. Chapter 12-PubChem: integrated platform of small molecules and biological activities. *Annu Rep Comput Chem. Amsterdam: Elsevier*. 2008;4:217-241. https://doi.org/10.1016/S1574-1400(08)00012-1.
35 Schlichtkrull M, Kipf TN, Bloem P, et al. Modeling relational data with graph convolutional networks. Lecture Notes in Computer Science. London: Springer Nature. 2018: 593-607. https://doi.org/10.1007/978-3-319-93417-4_38.
36 Sunseri J, Koes DR. libmolgrid: graphics processing unit accelerated molecular gridding for deep learning applications. *J Chem Inf Model*. 2020;60(3): 1079-1084. https://doi.org/10.1021/acs.jcim.9b01145.
37 Simm J, Humbeck L, Zalewski A, et al. Splitting chemical structure data sets for federated privacy-preserving machine learning. *J Cheminform*. 2021;13(1):96. https://doi.org/10.1186/s13321-021-00576-2.
38 Heid E, Greenman KP, Chung Y, et al. Chemprop: a machine learning package for chemical property prediction. *J Chem Inf Model*. 2024;64(1):9-17. https://doi.org/10.1021/acs.jcim.3c01250.
39 Xiong G, Wu Z, Yi J, et al. ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties . *Nucleic Acids Res*. 2021;49(W1):W5-W14. https://doi.org/10.1093/nar/gkab255.
40 Zhao K, Zhou Y, Qiao C, et al. Oroxylin A promotes PTEN-mediated negative regulation of MDM2 transcription *via* SIRT3-mediated deacetylation to stabilize p53 and inhibit glycolysis in wt-p53 cancer cells. *J Hematol Oncol*. 2015;8(1):41. https://doi.org/10.1186/s13045-015-0137-1.
41 Dai Q, Yin Q, Wei L, et al. Oroxylin A regulates glucose metabolism in response to hypoxic stress with the involvement of Hypoxia-inducible factor-1 in human hepatoma HepG2 cells. *Mol Carcinog*. 2016;55(8):1275-1289. https://doi.org/10.1002/mc.22369.
42 Wei L, Zhou Y, Qiao C, et al. Oroxylin A inhibits glycolysis-dependent proliferation of human breast cancer *via* promoting SIRT3-mediated SOD2 transcription and HIF1α destabilization. *Cell Death Dis*. 2015;6(4):e1714. https://doi.org/10.1038/cddis.2015.86.
43 Sun X, Chang X, Wang Y, et al. Oroxylin A suppresses the cell proliferation, migration, and EMT *via* NF- κB signaling pathway in human breast cancer cell. *BioMed Res Int*. 2019;2019:1-10. https://doi.org/10.1155/2019/9241769.
44 Yang X, Zhang F, Wang Y, et al. Oroxylin A inhibits colitis-associated carcinogenesis through modulating the IL-6/STAT3 signaling pathway. *Inflamm Bowel Dis*. 2013;19(1):1990-2000. https://doi.org/10.1097/MIB.0b013e318293c5e0.