

RESEARCH

Open Access



MultiCycPermea: accurate and interpretable prediction of cyclic peptide permeability using a multimodal image-sequence model

Zixu Wang^{1†}, Yangyang Chen^{1†}, Yifan Shang², Xiulong Yang³, Wenqiong Pan⁴, Xiucai Ye^{1*}, Tetsuya Sakurai¹ and Xiangxiang Zeng^{2*}

Abstract

Background Cyclic peptides, known for their high binding affinity and low toxicity, show potential as innovative drugs for targeting “undruggable” proteins. However, their therapeutic efficacy is often hindered by poor membrane permeability. Over the past decade, the FDA has approved an average of one macrocyclic peptide drug per year, with romidepsin being the only one targeting an intracellular site. Biological experiments to measure permeability are time-consuming and labor-intensive. Rapid assessment of cyclic peptide permeability is crucial for their development.

Results In this work, we proposed a novel deep learning model, dubbed as MultiCycPermea, for predicting cyclic peptide permeability. MultiCycPermea extracts features from both the image information (2D structural information) and sequence information (1D structural information) of cyclic peptides. Additionally, we proposed a substructure-constrained feature alignment module to align the two types of features. MultiCycPermea has made a leap in predictive accuracy. In the in-distribution setting of the CycPeptMPDB dataset, MultiCycPermea reduced the mean squared error (MSE) by approximately 44.83% compared to the latest model Multi_CycGT (0.29 vs 0.16). By leveraging visual analysis tools, MultiCycPermea can reveal the relationship between peptide modification structures and membrane permeability, providing insights to improve the membrane permeability of cyclic peptides.

Conclusions MultiCycPermea provides an effective tool that accurately predicts the permeability of cyclic peptides, offering valuable insights for improving the membrane permeability of cyclic peptides. This work paves a new path for the application of artificial intelligence in assisting the design of membrane-permeable cyclic peptides.

Keywords Cyclic peptide, Peptide membrane permeability, Interpretable model, Deep learning

[†]Zixu Wang and Yangyang Chen contributed equally to this work.

*Correspondence:

Xiucai Ye

yexiucai@cs.tsukuba.ac.jp

Xiangxiang Zeng

xzeng@hnu.edu.cn

¹ Department of Computer Science, University of Tsukuba, Tsukuba, Ibaraki 3058577, Japan

² College of Computer Science and Electronic Engineering, Hunan University, Changsha, Hunan 410082, People's Republic of China

³ School of Computer Science, Central China Normal University, Wuhan 430079, People's Republic of China

⁴ Department of Clinical Pharmacy, Jilin University, Changchun, Jilin 130025, People's Republic of China



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

Membrane-permeable cyclic peptides have emerged as a promising class of therapeutic agents capable of accessing intracellular targets that are typically inaccessible to conventional small molecule drugs [1–3]. However, due to their structural limitations, cyclic peptides do not easily cross cell membranes. Over the past decade, FDA (the Food and Drug Administration) has granted approval to an average of one macrocyclic peptide drug per year [4]. Among these, romidepsin [5] approved for lymphoma treatment is the only that acts on the intracellular target. This limitation has been a major hurdle in advancing the use of macrocyclic peptides in therapeutics.

Researchers have developed various design strategies to assist in the membrane permeability of cyclic peptides. One prevalent approach to enhancing membrane permeability is to stabilize cyclic peptides in a “closed” conformation [6–8] within the cell membrane by protecting exposed NH groups via backbone N-methylation [9, 10]. Additionally, methods such as amide-to-ester substitution [11] and altering side chain structures to form intramolecular hydrogen bonds with the main chain have been investigated to further improve membrane permeability [12]. However, a universal rule regarding membrane permeability has not yet been discovered [13].

The elusive nature of cyclic peptide membrane permeability makes it challenging to assess permeability potential during the initial design stages. The biochemical assessment of the permeability, using methods such as parallel artificial membrane permeability assay [14] (PAMPA), is labor-intensive and time-consuming. This underscores the importance of developing an accurate *in silico* model for predicting membrane permeability. Early computational models for assessing membrane permeability often relied on molecular dynamics simulations [15–17]. However, the computational efficiency of these simulations is unsatisfactory.

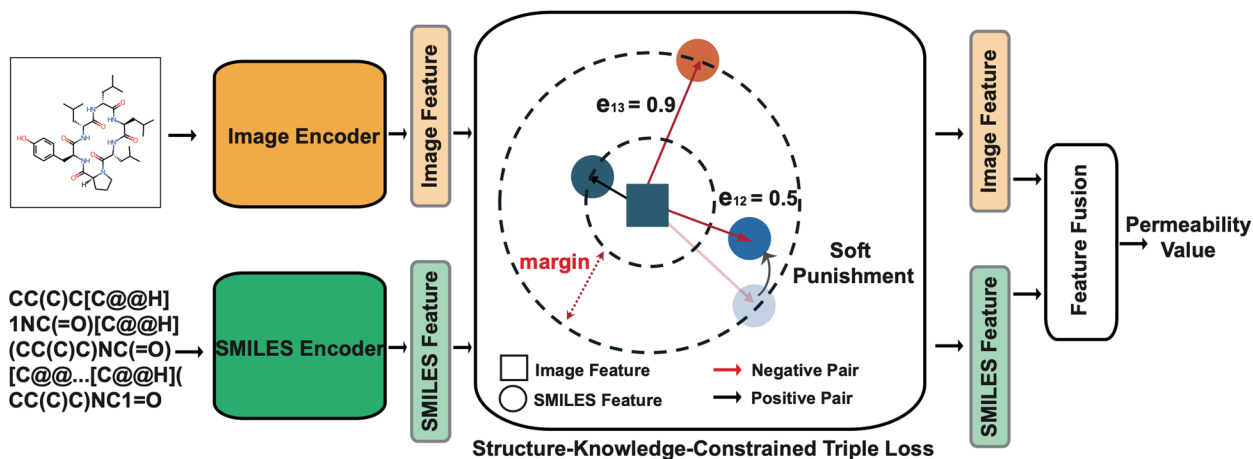
Quantitative structure–property relationship [18] (QSPR) models based on machine learning are rapidly gaining traction. Their rise in prominence stems from their precision in predicting intricate biological outcomes, reducing the need for costly experimental procedures [19–22]. Some studies have attempted to construct QSPR models for predicting the permeability of cyclic peptides. Poongavanam et al. pioneered a machine learning approach to categorize cyclic peptides as having either low-moderate or high permeability [23]. Digiesi et al. used various computational tools, ranging from physics-based methods to 3D-QSPR methods, to predict permeability [24]. Although these approaches accelerate the experimental process, their applicability remains limited due to the small amount of available data (only dozens of samples). Recently, a database of cyclic

peptide membrane permeability (CycPeptMPDB) [25] has injected new vitality into this field. It contains thousands of cyclic peptide structures and their associated properties, including experimentally measured membrane permeability. Benefiting from this dataset, previous works [26, 27] such as Multi_CycGT attempted to predict the membrane permeability of cyclic peptides using deep learning methods. These models are multimodal, based on the graph, sequence, and various molecular properties of cyclic peptides. However, we identified several issues which limited their performance: (1) The models do not leverage the knowledge of pretrained models [28]. Despite the further development of cyclic peptide datasets, the overall data volume is still relatively small, which limits the performance of deep learning models that require large amounts of data. (2) Their modality fusion methods are relatively simple and do not effectively harness the potential of integrating information from different modalities [29]. (3) The evaluation is limited to relatively simple data partitioning methods, such as random partitioning, and does not further challenge the performance of models in real-world scenarios.

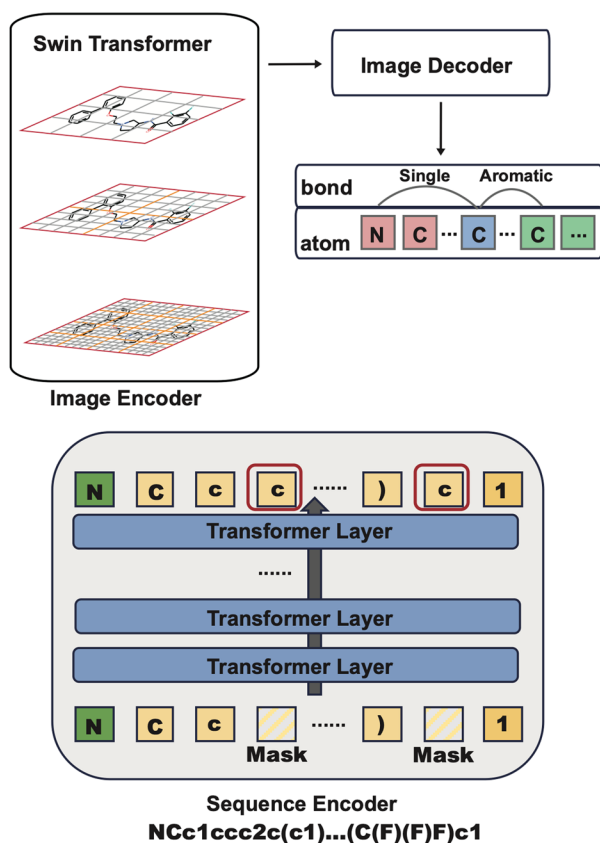
In this work, we proposed a novel cyclic peptide permeability prediction model, dubbed as MultiCycPermea (Fig. 1). MultiCycPermea integrates both the molecular image representation and the sequence representation of cyclic peptides, incorporating three key modules. One cyclic peptide image encoding module utilizes the Swin Transformer [30] to capture the two-dimensional structural features of cyclic peptide images, subsequently converting the image information into vectors. Another cyclic peptide sequence encoding module uses multiple transformer layers to encode the SMILES [31] sequence (one-dimensional) features of cyclic peptides into vectors. Notably, these two encoding modules are not trained from scratch but are initialized with models pretrained on small molecule data. Finally, a structure-constrained triple loss module is used to align features from these two modalities. It constructs a graph using substructure similarity between cyclic peptides to constrain feature alignment. In the experimental section, unlike previous works that only evaluated models under simple in-distribution (random partitioning) conditions, we simulated out-of-distribution data partitioning to test on unseen cyclic peptide structures. Furthermore, to investigate the ability of models to capture drastic changes in membrane permeability due to subtle structural variations in cyclic peptides, we also evaluated the model using the permeability cliff partitioning method. Primary contributions of this work can be summarized as follows:

- 1) We proposed a novel multimodal method to predict cyclic peptide permeability, named MultiCycPer-

a Model Structure



b Encoder Pre-training



c Substructure Knowledge Graph

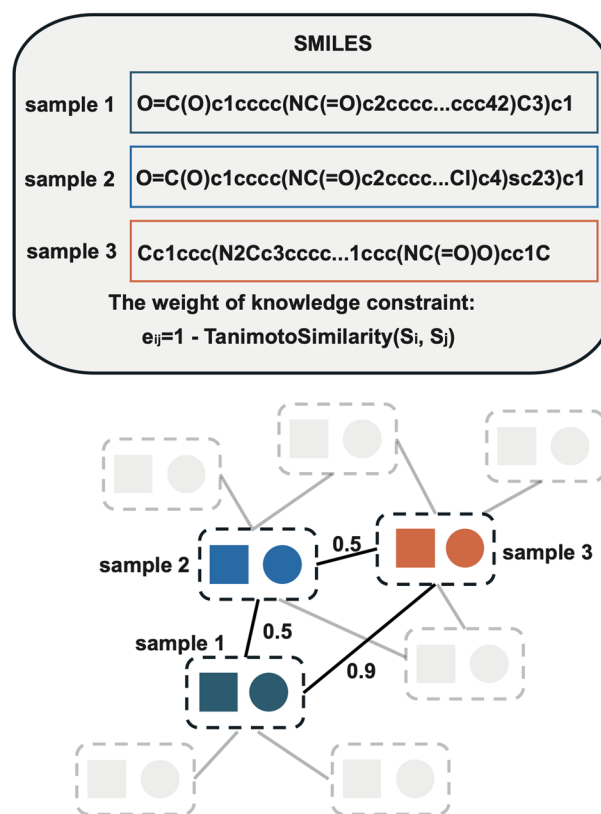


Fig. 1 The architecture of MultiCycPermea. **a** MultiCycPermea incorporates one image encoder for 2D structure feature extraction and one sequence encoder for 1D structure feature extraction. The information from these two modalities is aligned using a triple loss adjusted based on substructure similarity. **b** The cyclic peptide image encoder is first pretrained on a molecular OCR recognition task, while the cyclic peptide sequence encoder is pretrained on a molecular "MASK" prediction task. **c** The logic of substructure knowledge network construction: We use Tanimoto similarity between cyclic peptides to calculate substructure differences, which are then used to constrain the triple loss for aligning the two modalities

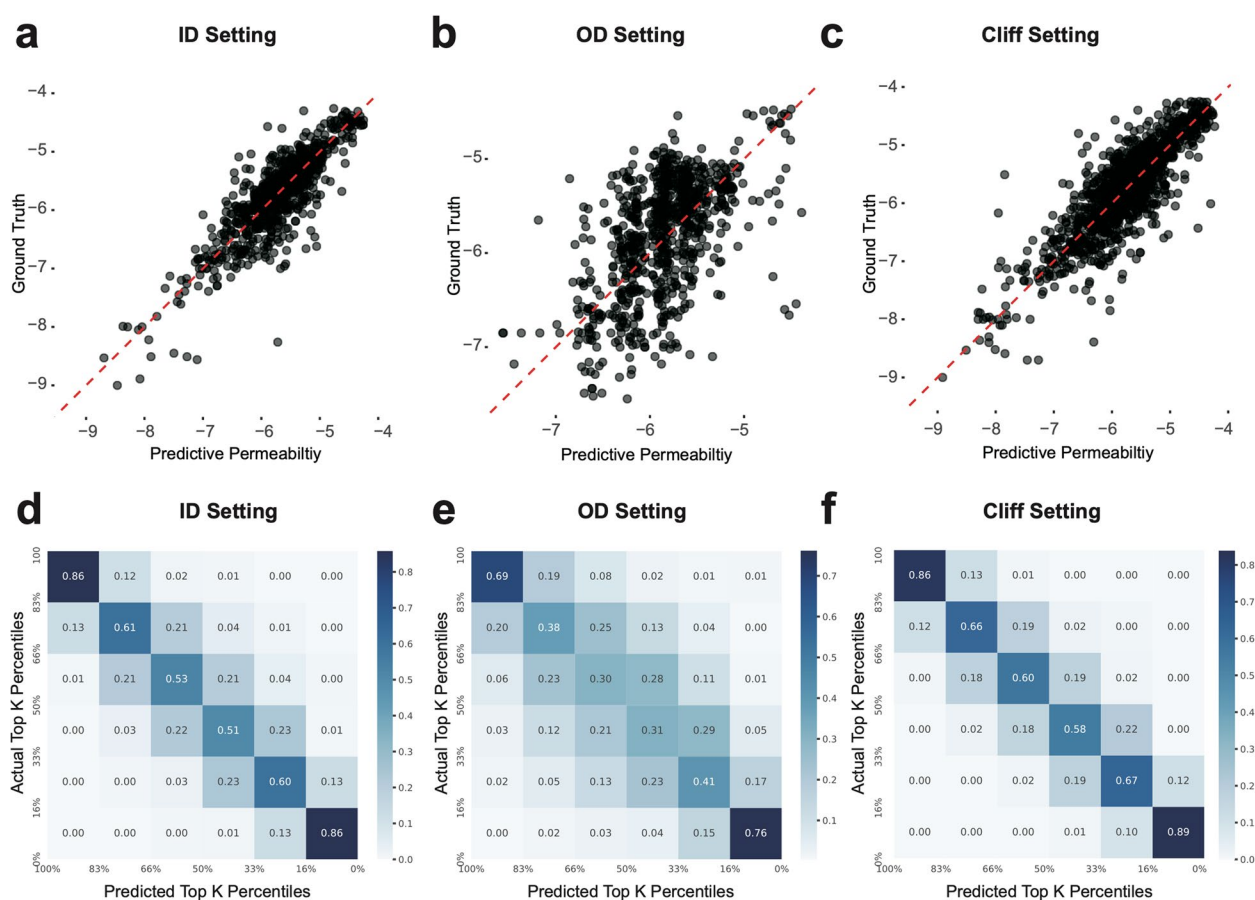


Fig. 2 The prediction results of MultiCycPermea. **a–c** The scatter plots of predicted permeability of cyclic peptides and ground truth under 3 different data setting. **d–f** Confusion matrix showing the predicted versus actual top k percentiles of cyclic peptide permeability. MultiCycPermea can ensure that the predicted rankings reflect the relative positions of the actual permeability values

mea. It captures the 2D structural features and 1D sequence features of cyclic peptides based on image information and sequence information. To the best of our knowledge, this is the first work to introduce images as a modality in the field of peptides.

- 2) We demonstrated that pretraining encoder models on small molecule data improves the performance of cyclic peptide permeability prediction. The model can enhance its encoding capabilities by pre-learning the structural knowledge of small molecules, highlighting the importance of pretrained models.
- 3) We comprehensively evaluated the performance of models, especially in the previously overlooked data partitioning methods (simulated out-of-distribution setting and permeability cliff setting). MultiCycPermea achieved state-of-the-art performance across all data partitioning scenarios on the CycPeptMPDB dataset. We evaluated the predictive capability of MultiCycPermea in a recent cyclic peptide design study. MultiCycPermea successfully filtered out

approximately 80% of cyclic peptides with low membrane permeability.

- 4) MultiCycPermea shows interpretability. Leveraging visual model analysis tools helps us understand how the model constructs the relationship between cyclic peptide substructures and membrane permeability.

Results

The performance for MultiCycPermea and other baselines

To validate the efficacy of the model, we apply three methods to split the data. The predictive capability of MultiCycPermea is shown in Fig. 2. The comparison results of MultiCycPermea with other baselines are shown in the following: in-distribution (ID) setting (Table 1 and Additional file 1: Table S2), simulated out-distribution (OD) setting (Table 2 and Additional file 1: Table S3), permeability values cliff setting (Table 3 and Additional file 1: Table S4).

ID setting: In this basic data partitioning evaluation, apart from models like TextCNN and BiLSTM (with R^2

Table 1 Permeability prediction results under ID setting

Model	MSE (↓)	MAE (↓)	R^2 (↑)	PCC (↑)	SCC (↑)
SVM	0.48	0.48	0.18	0.46	0.46
RandomForest	0.30	0.38	0.48	0.70	0.66
TextCNN	0.59	0.57	0.00	0.07	0.12
BiLSTM	0.59	0.57	0.00	−0.05	−0.05
GAT	0.47	0.50	0.19	0.45	0.46
GCN	0.49	0.52	0.17	0.41	0.43
ChemBERTa	0.26	0.38	0.61	0.76	0.69
Unimol	0.24	0.36	0.59	0.77	0.72
MolCLR	0.19	0.33	0.67	0.82	0.76
CyclePermea	0.22	0.34	0.63	0.81	0.74
Multi_CycGT	0.29	0.40	0.45	0.71	0.68
MultiCycPermea	0.16	0.28	0.73	0.85	0.81

Table 2 Permeability prediction results under OD setting

Model	MSE (↓)	MAE (↓)	R^2 (↑)	PCC (↑)	SCC (↑)
SVM	0.62	0.55	−0.12	−0.05	−0.01
RandomForest	0.51	0.51	0.07	0.32	0.37
TextCNN	0.60	0.57	−0.04	−0.09	−0.04
BiLSTM	0.61	0.58	−0.08	−0.09	−0.02
GAT	0.58	0.54	0.02	0.17	0.22
GCN	0.56	0.52	0.02	0.22	0.27
ChemBERTa	0.50	0.53	0.07	0.30	0.36
Unimol	0.48	0.52	0.12	0.38	0.48
MolCLR	0.48	0.49	0.12	0.44	0.48
CyclePermea	0.47	0.49	0.11	0.43	0.47
Multi_CycGT	0.53	0.54	0.03	0.41	0.45
MultiCycPermea	0.44	0.47	0.19	0.48	0.51

Table 3 Permeability prediction results under Cliff setting

Model	MSE (↓)	MAE (↓)	R^2 (↑)	PCC (↑)	SCC (↑)
SVM	0.47	0.58	0.16	0.64	0.63
RandomForest	0.30	0.40	0.57	0.76	0.72
TextCNN	0.75	0.66	0.00	0.10	0.11
BiLSTM	0.75	0.66	−0.02	0.00	−0.04
GAT	0.56	0.56	0.23	0.48	0.50
GCN	0.62	0.58	0.15	0.41	0.45
ChemBERTa	0.29	0.39	0.60	0.79	0.73
Unimol	0.32	0.41	0.58	0.69	0.64
MolCLR	0.23	0.35	0.71	0.84	0.80
CyclePermea	0.27	0.38	0.63	0.80	0.72
Multi_CycGT	0.31	0.41	0.56	0.73	0.65
MultiCycPermea	0.18	0.29	0.76	0.88	0.85

approximately equal to 0), the remaining models can predict the permeability of cyclic peptides to a certain extent. From the perspective of the MSE metric, the better-performing methods were RandomForest (0.30), ChemBERTa (0.26), Unimol (0.24), MolCLR (0.19), Multi_CycGT (0.29), and MultiCycPermea (0.16). Notably, models pretrained on small molecules (regardless of the pretraining strategy) all showed higher predictive performance than the latest cyclic peptide permeability prediction model Multi_CycGT (ChemBERTa: −0.03, Unimol: −0.05, MolCLR: −0.10, and MultiCycPermea: −0.13). This highlights the importance of pre-trained models, even when trained on cross-domain data (the molecular mass of the small molecule dataset is mostly below 400 Da, while cyclic peptide mass is often above 800 Da). Additionally, in correlation metrics, models showed similar tendencies. For example, with the PCC (Pearson correlation coefficient) metric, Multi_CycGT scores 0.71, while other pretrained models show improved performance (ChemBERTa: +0.05, Unimol: +0.06, MolCLR: +0.11, and MultiCycPermea: +0.14). MultiCycPermea exceeded other baselines across various evaluation metrics, demonstrating the effectiveness of the MultiCycPermea architecture for the task of predicting cyclic peptide permeability.

OD setting: After reducing the structural similarity between the training and testing data, all models showed a significant performance decline. Considering both MSE and PCC metrics, pretrained models still have some advantages over not pretrained Multi_CycGT (MSE: 0.53, PCC: 0.41). Specifically, all pretrained models perform better than Multi_CycGT in terms of MSE, while MolCLR (+0.03) and MultiCycPermea (+0.07) show better performance in terms of PCC. Under this relatively rigorous evaluation, MultiCycPermea outperformed other baselines.

Cliff setting (Fig. 3): In the real-world cyclic peptide design work, researchers often start from known cyclic peptide structures and gradually modify certain parts to improve their properties [32, 33]. Therefore, it is crucial for the model to detect significant changes in membrane permeability resulting from different modification methods. The models that performed well in the previous two tasks still had a significant advantage here. Among them, MultiCycPermea performed well, achieving an MSE of 0.18 and a PCC of 0.88.

The effectiveness of each component in MultiCycPermea

We conducted a series of ablation studies to analyze the importance of each module in MultiCycPermea. Figure 4 illustrates the impact of each component across three data settings (represented by MSE and PCC metrics).

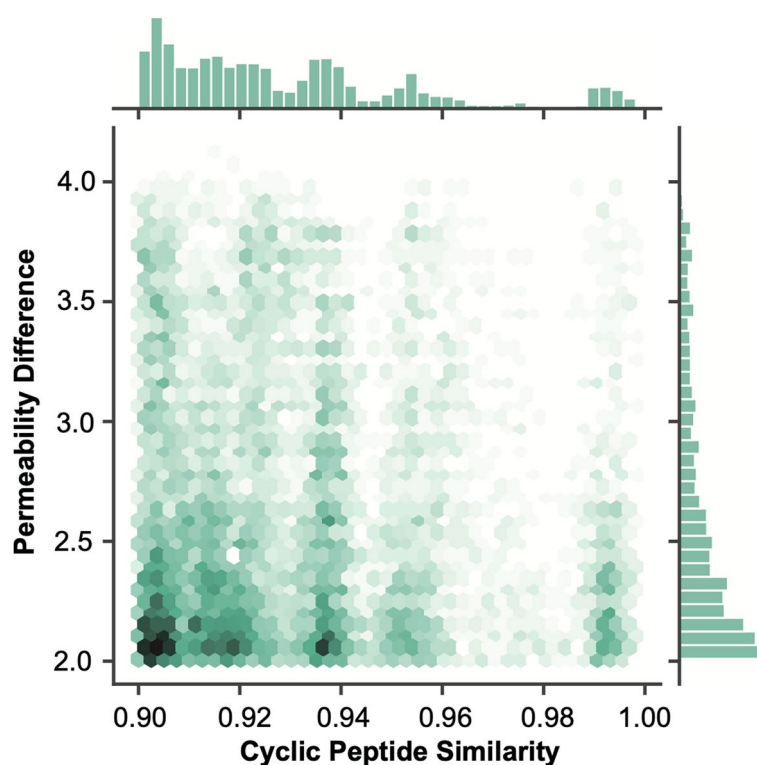


Fig. 3 Distribution of permeability cliff data pairs. Although many known cyclic peptide pairs have high structural similarity (greater than 0.9), their corresponding membrane permeability often differs a lot (permeability difference greater than 2, note that the permeability values shown here are log-transformed)

We summarized several findings from the ablation experiments.

- (1) Images can serve as a data modality for peptides to help predict membrane permeability and can be integrated with sequence information to improve model performance. Although images have never been used as a data modality in previous peptide-related work, we have demonstrated their potential application value in this study. Under all three data partitioning methods, the predictive performance of the ImageEncoder (only) and the SequenceEncoder (only) are comparable. With the help of image information, MultiCycPermea (w/o triple loss) outperformed the SequenceEncoder (only). For example, in the OD setting, the fusion of information results in an improvement of 0.05 and 0.06 points in MSE and PCC, respectively, compared to the sequence-only performance.
- (2) Pretraining on small molecules can help the encoders improve their encoding capabilities and significantly enhance model performance on cyclic peptide membrane permeability prediction. For encoders without pretraining, their performance

showed a decline across all data settings (Additional file 1: Tables S5–S7). Specifically, due to the inherent sparsity of molecular images, the ImageEncoder required pretraining guidance. For example, in the ID setting, the ImageEncoder (pretrained) outperformed the ImageEncoder (w/o pretrain) by 0.36 in MSE and 0.68 in PCC. The SequenceEncoder, in comparison, does not rely as heavily on pretraining guidance, but pretraining still improved the performance to a certain extent. Although the CycPept-MPDB dataset represents a qualitative improvement over previous datasets, it is still insufficient for training “data-hungry” deep learning models. We chose to pretrain on a small molecule dataset instead of a peptide dataset because most peptide datasets consist of standard amino acids. Designing membrane-permeable cyclic peptides often requires various modifications, making these datasets primarily composed of non-standard amino acids.

- (3) Feature aligning after incorporating structural knowledge can improve the model’s predictive capability. MultiCycPermea demonstrated performance improvements when incorporating triple

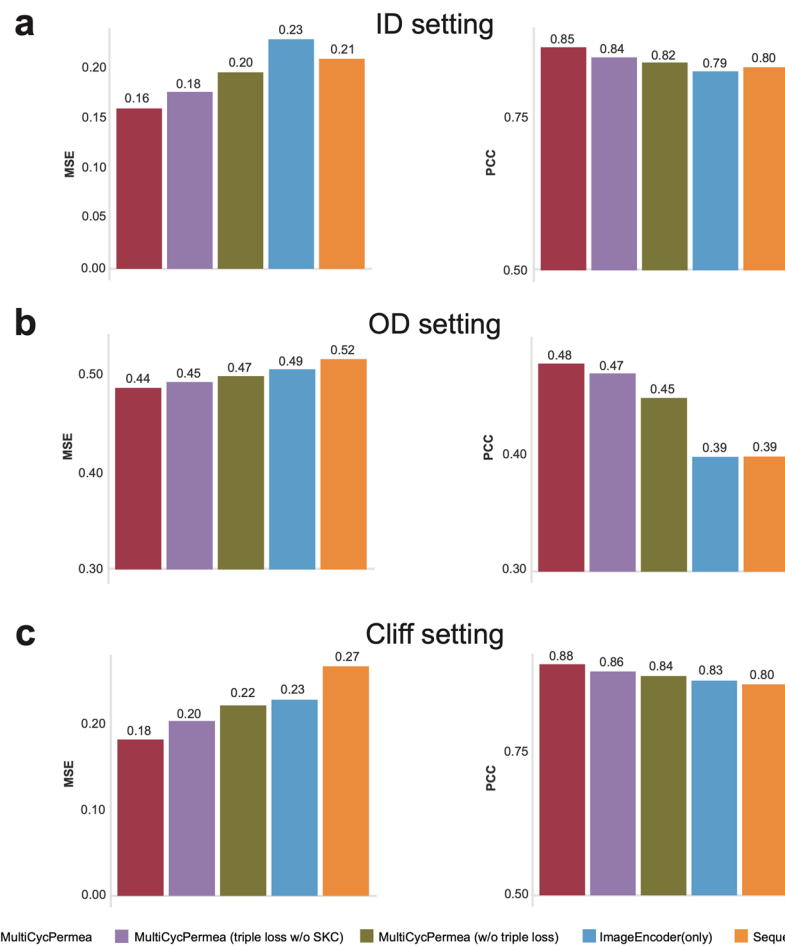


Fig. 4 The ablation study of MultiCycPermea. **a–c** The results of the ablation experiments under ID setting, OD setting, and Cliff setting, respectively. The MultiCycPermea (w/o triple loss) version directly concatenated the features from the two modalities to predict membrane permeability. SKC stands for substructure-knowledge-constrained

loss and triple loss with SKC. For example, in the Cliff setting, the original model improved MSE from 0.22 to 0.18, while the PCC increased from 0.84 to 0.88. We revealed the source of the performance improvement by depicting the relationship of Tanimoto similarity between cyclic peptide pairs and the feature similarity proposed by the model. Using the model trained in the ID setting as an example, we randomly sample peptides from the test set, where each cyclic peptide is paired with another randomly selected cyclic peptide. We then calculated the Tanimoto similarity and feature similarity between the peptide pairs. The results demonstrated that the model trained with edge weight constraints better integrated substructure similarity into feature extraction, evidenced by an increase in PCC from 0.55 (Fig. 5b, before constraints) to 0.68 (Fig. 5c, after constraints). Incorporating structural knowledge into the feature alignment process

that the features extracted by the model were more closely aligned with the inherent structural characteristics of the cyclic peptides.

Leveraging the visual model analysis tool to uncover the structural knowledge of cyclic peptide membrane permeability

One major advantage of using images as a modality is the ability to leverage analysis tools from the field of computer vision to visually understand how the model makes decisions. We employed Grad-CAM (Gradient-weighted Class Activation Mapping) [34] to elucidate how MultiCycPermea determines the permeability cliff pairs of cyclic peptides. Grad-CAM is a visualization technique that highlights the image regions the network emphasizes. Figure 6 presents three cyclic peptide pairs in the permeability cliff setting, with an importance heatmap

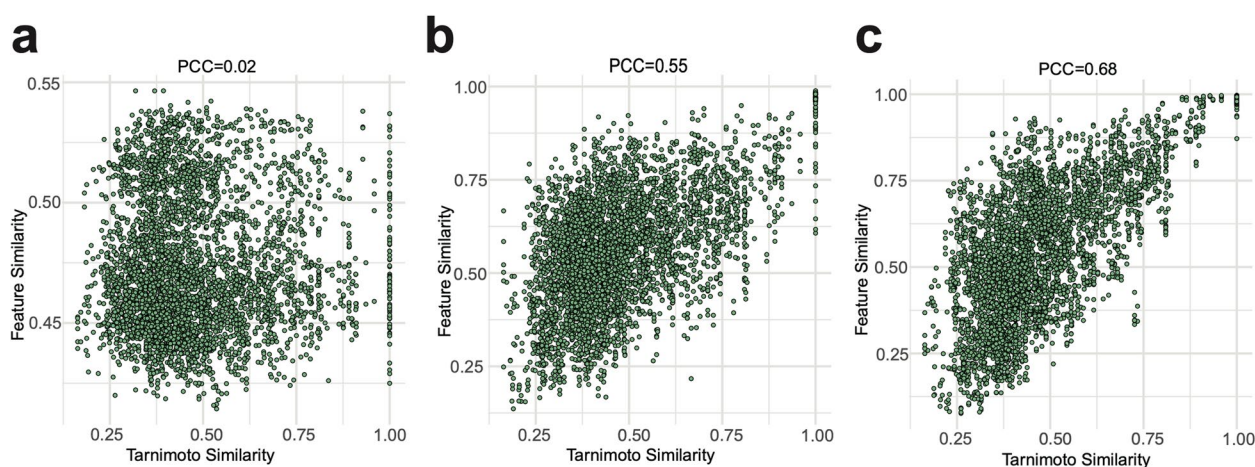


Fig. 5 The scatter plot distribution of feature similarity and Tanimoto similarity for test data in the ID setting. **a** The feature component is extracted from MultiCycPermea (w/o triple loss). **b** The feature component is extracted from MultiCycPermea (triple loss w/o SKC). **c** The feature component is extracted from MultiCycPermea

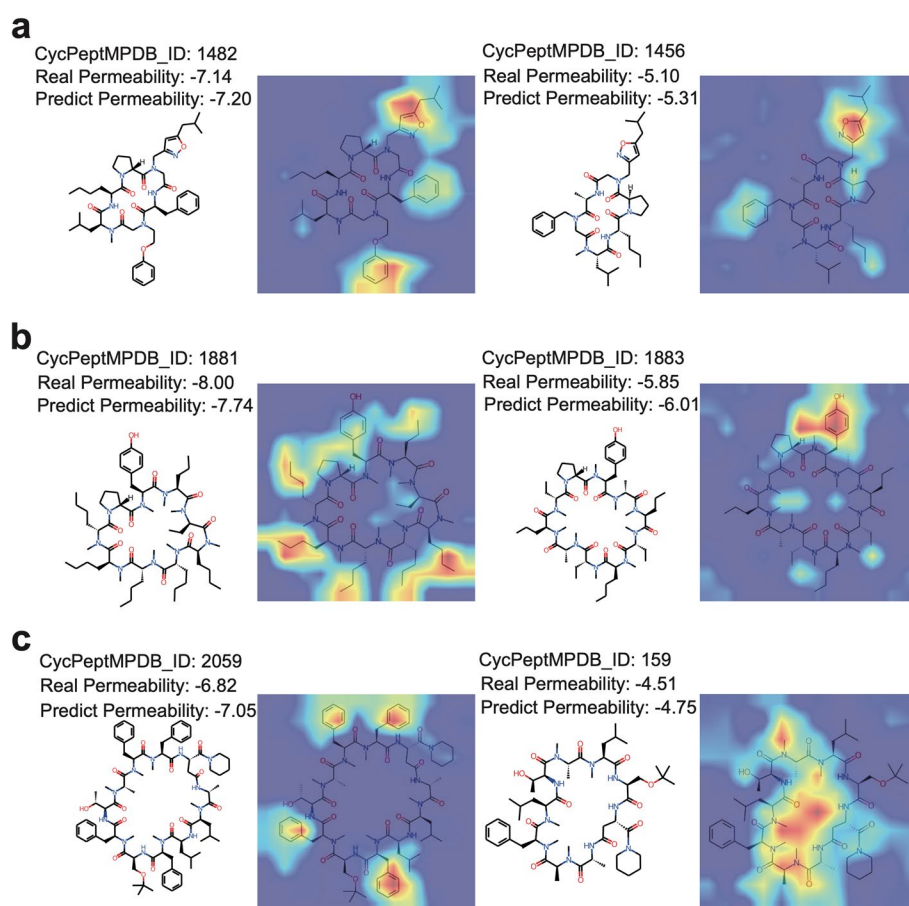


Fig. 6 Cases of permeability cliff pairs of cyclic peptides and the regions of interest identified by MultiCycPermea using Grad-CAM analysis. In each cliff pair, the cyclic peptide on the left is difficult to permeate (< -6), while the corresponding peptide on the right is easy to permeate (> -6). Redder colors indicate higher attention, while bluer colors indicate lower attention. **a** The two cyclic peptides have a Levenshtein similarity of 0.92. **b** The two cyclic peptides have a Tanimoto similarity of 0.91. **c** The two cyclic peptides have a Tanimoto similarity of 0.92

overlaid on the cyclic peptide images. For the cyclic peptides in examples **a** and **c** that are difficult to permeate (ID:1482 and 2059), the model paid close attention to the number of aromatic rings, which is the structural difference compared to the corresponding easily permeable cyclic peptides (ID:1456 and 159). The structural differences between the two cyclic peptides in example **b** are mainly on whether the side chains of the peptide backbone are heavily alkylated. MultiCycPermea successfully reveals this distinction. Notably, for the more permeable cyclic peptide (ID:1883), MultiCycPermea not only focused on the phenol structure (polar functional groups that reduce membrane permeability) but also gave high attention to areas around all N-methylation sites (a very popular modification strategy to enhance cyclic peptide membrane permeability [10]). These examples illustrate that MultiCycPermea has captured certain relationships between the substructures of cyclic peptides and their membrane permeability. By using Grad-CAM to reveal the decision-making process of MultiCycPermea, it can provide insights for medicinal chemists to improve the membrane permeability of cyclic peptides.

MultiCycPermea can help screen for membrane-permeable cyclic peptides

In practical applications, it is crucial to determine whether the designed cyclic peptides are easily permeable. We applied MultiCycPermea to a recent cyclic peptide design work [35], where researchers designed millions of potential cyclic peptides and tested the membrane permeability of some representative ones. The final readable test data includes 24 cyclic peptides, of which 9 are difficult to permeate. We used the MultiCycPermea model trained in the OD setting for evaluation. The comparison between the predicted results and the permeability values obtained from PAMPA experiments is shown in Fig. 7a. MultiCycPermea achieved an MSE of 0.45 and a PCC of 0.55, indicating that MultiCycPermea can perceive changes in membrane permeability due to structural variations. Notably, in practical scenarios, we generally consider a permeability value of -6 as an important evaluation point. Cyclic peptides with a permeability value higher than -6 are considered easily permeable, while those below -6 are considered difficult to permeate. In this dataset, among the 9 cyclic peptides that are difficult to permeate, MultiCycPermea identified 7, with 1 being uncertain (predicted value close to the boundary at greater than -6.2 but less than -5.8), successfully eliminating approximately 80% of the difficult-to-permeate cyclic peptides. For the 15 cyclic peptides that are easy to permeate, MultiCycPermea only misjudged the permeability of 2 peptides. The ability to distinguish between easily permeable and

difficult-to-permeate cyclic peptides with high accuracy suggests that MultiCycPermea can assist in initial screening for membrane permeability in cyclic peptide synthesis experiments.

Discussion

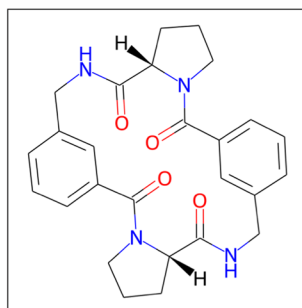
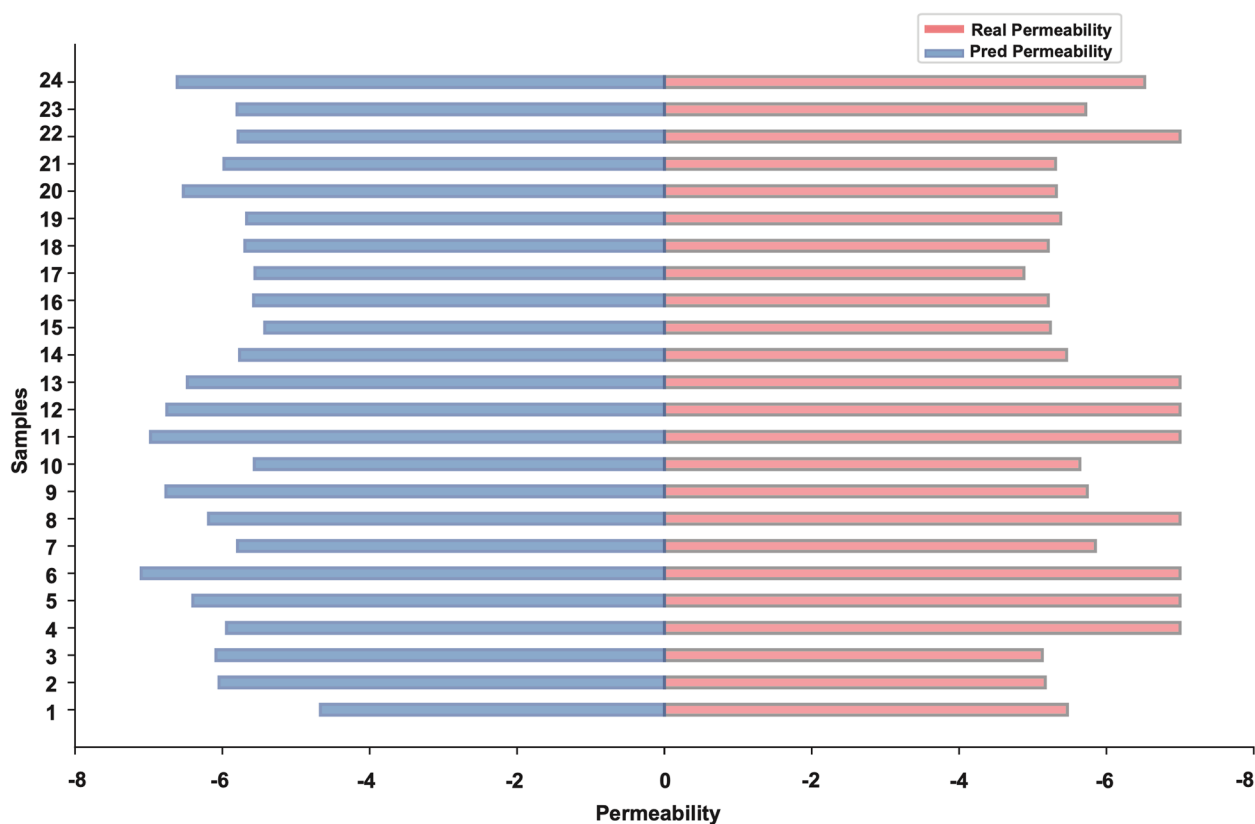
Our image-based approach has certain limitations: (1) The representation of cyclic peptides in this work relies on Lewis structures, which cannot capture stereochemical information, thus limiting the model's ability to consider three-dimensional structural features that may affect permeability. (2) The resolution of images generated by RDKit is generally adequate for small to medium-sized peptides (up to a molecular weight of 1777 Da, the largest in CycPeptMPDB). However, for larger cyclic peptides, the existing image resolution becomes insufficient for accurate representation, potentially affecting the model's performance on such molecules.

Conclusions

In the early stages of cyclic peptide drug design, determining membrane permeability is crucial. In this study, we introduced MultiCycPermea, a multimodal deep learning model for cyclic peptide permeability prediction. We tested MultiCycPermea on the most comprehensive cyclic peptide dataset (CycPeptMPDB) using three data partitioning methods, and it outperformed the state-of-the-art models. We addressed a challenge in cyclic peptide permeability prediction, the limited availability of training data, through transfer learning, where the model first learns fundamental chemical structure representations from extensive small molecule datasets before fine-tuning on cyclic peptide prediction. We also demonstrated that images, as a highly intuitive modality, cannot only be used for cyclic peptide property prediction but also reveal decision-making principles through visual interpretability tools. For future work, we plan to enhance MultiCycPermea in two directions: implementing online learning mechanisms to continuously incorporate newly available permeability data and integrating stereochemical information through video-based representations to capture dynamic conformational changes of cyclic peptides. We believe these explorations can further improve the model's applicability and accuracy.

Methods

The overall architecture of MultiCycPermea for predicting the permeability of cyclic peptides is illustrated in Fig. 1a. MultiCycPermea uses two input modalities of cyclic peptide data (image and sequence) and integrates them through a structure-knowledge-constrained triple loss fusion (Fig. 1c). In this section, we first introduced the architecture and the pretraining of the image encoder

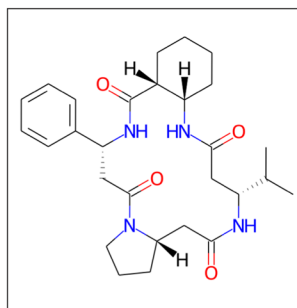


Sample_ID: 12

O=C1NCc2cccc(c2)C(=O)N2CCC[C@H]2C(=O)NCc2cccc(c2)C(=O)N2CC[C@H]12

Real Permeability: -7.0

Predicted Permeability: -6.75

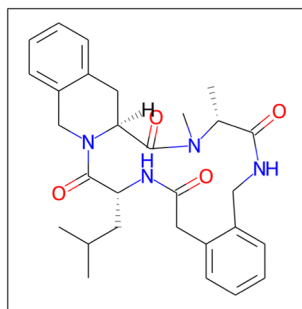


Sample_ID: 13

CC(C)[C@H]1CC(=O)N[C@H]2CCCC[C@H]2C(=O)N[C@H](c2cccc2)C(=O)N2CCC[C@H]2CC(=O)N1

Real Permeability: -7.0

Predicted Permeability: -6.47

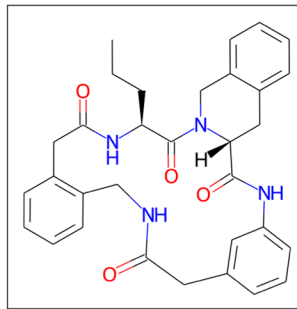


Sample_ID: 15

O=C([C@H](C)N(C)C([C@H](CC(C=CC=C1)=C1C2)N2C([C@H](CC(C)N3)=O)=O)NCC4=CC=CC=C4CC3=O

Real Permeability: -5.24

Predicted Permeability: -5.43



Sample_ID: 23

O=C([C@H](CCC)NC(CC1=C(CNC(C2)=O)C=CC=C1)=O)N3[C@H](C(NC4=CC=CC2=C4)=O)CC5=CC=CC=C5C3

Real Permeability: -5.72

Predicted Permeability: -5.80

Fig. 7 Using MultiCycPermea to assist in screening easily permeable cyclic peptides. **a** Actual membrane permeability and predicted values by MultiCycPermea for the 24 cyclic peptides in the external evaluation set. **b** Examples of cyclic peptides that are difficult to permeate. **c** Examples of cyclic peptides that are easy to permeate

model. Second, we introduced the architecture of the peptide sequence encoder and the pretraining model. Third, we described the feature fusion model and the training target.

Cyclic peptide image encoder

Architecture of the image encoder: MultiCycPermea employed the Swin Transformer (Swin-B) architecture [30], which consists of 4 stages and 12 transformer layers. Each stage uses a 4×4 patch size with embedding dimensions of 128, 256, 512, and 1024, and has 4, 8, 16, and 32 attention heads, respectively. The model features a window size of 7 and an MLP ratio of 4. Swin-B has a shifted window mechanism where self-attention [32] is applied within windows that shift between layers, enhancing the capture of both local and global features, reducing computational complexity, and maintaining the ability to model long-range dependencies. In MultiCycPermea, $\mathbf{X}_i \in \mathbb{R}^{384 \times 384 \times 3}$ represent the input cyclic peptide image and is divided into non-overlapping small windows as shown in Fig. 1b. The self-attention mechanism within a window can be described by:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{SoftMax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (1)$$

where \mathbf{Q} , \mathbf{K} , and \mathbf{V} are the query, key, and value matrices. After four stages of dimensionality reduction with the Swin-B, the feature map $\mathbf{H}_{X_i} = \text{ImageEncoder}(\mathbf{X}_i)$ of the cyclic peptide molecular image becomes $12 \times 12 \times 1024$.

Pretraining of the image encoder: Inspired by the excellent performance of deep learning-based molecular OCR recognition tools, represented by MolScribe [36], we first loaded the encoders pretrained on the OCR task, and then transferred it to peptide prediction task. Specifically, during the pretraining process, a decoder is used to decode the feature map of the molecular image into the atoms and bonds:

$$\mathbf{X}_i^A, \mathbf{X}_i^B = \text{ImageDecoder}(\mathbf{H}_{X_i}), \quad (2)$$

where \mathbf{X}_i^A represents the type of each atom in the image while \mathbf{X}_i^B represents the type of covalent bonds between the atoms in the image. The main part of the decoder consists of 6 transformer layers with 8 attention heads. A fully connected linear network was used to predict the types of covalent bonds. After pretraining, the decoder part was completely discarded, and only the encoder part was retained for MultiCycPermea.

Cyclic peptide sequence encoder

Architecture of the sequence encoder: First, the SMILES representation of the cyclic peptides S_i is tokenized into

individual characters $[t_0, t_1, \dots, t_L]$ up to 250 tokens using regular expressions. A start token and an end token are added to the beginning and end of the sequence, respectively. These tokens are then embedded into a high-dimensional space: $[\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_L] = \text{Embedding}([t_0, t_1, \dots, t_L])$, where $\mathbf{w}_\bullet \in \mathbb{R}^{256}$ and L is the length of the sequence. Second, the sequence of vectors is processed through 8 transformer layers with 8 heads [37], $\mathbf{H}_{S_i} = \text{SequenceEncoder}([\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_L])$. Each layer consists of a self-attention mechanism followed by a position-wise feedforward network with an inner layer size of 1024 and an output size of 256. Layer normalization is applied before each sub-layer, and a dropout rate of 0.1 is used after both the attention and feedforward networks. Residual connections surround each sub-layer to facilitate gradient flow. The mechanism of the self-attention layer is identical to Eq. 1. The difference lies in that one captures features within image windows, while the other captures features in molecular sequences.

Pretraining of the sequence encoder: For the pretraining of the sequence encoder, we adopt a method similar to the masked language modeling of BERT [38]. In this process, random tokens in the SMILES sequences are masked, and the model is trained to predict these masked tokens. This allows the model to learn contextual dependencies and relationships within the molecular sequence effectively: $\mathcal{L}_{\text{Mask}} = -\log P(t_i | t_1, \dots, t_{i-1}, [\text{MASK}], t_{i+1}, \dots, t_L)$. In practice, each token has a 15% chance of being masked. Finally, we discarded the parameters of the output layer used for token prediction and retained the remaining parameters for MultiCycPermea.

Structure constrained feature fusion and permeability prediction

We proposed a substructure-knowledge-constrained triple loss for fine-grained alignment of features from two modalities. Specifically, we created a knowledge graph based on the substructure similarity between peptides (Fig. 1c). The knowledge network is defined as $G = (V, E)$, where V represents the set of nodes (cyclic peptides without distinguishing between different modalities) and E represents the set of edges. The weights on the edges are calculated by the similarity of substructures between two cyclic peptides. For example, the edge weight $e_{i,j}$ between node i and j in the knowledge is calculated as:

$$e_{i,j} = 1 - \text{TanimotoSimilarity}(S_i, S_j). \quad (3)$$

This way of constructing the knowledge graph ensures that the weights between nodes of cyclic peptides with similar substructures are smaller, while cyclic peptides with different substructures are larger.

MultiCycPeptide aligns features based on this structural knowledge graph constraint. First, average pooling was applied to the 1024-dimensional feature vectors of peptide images, which was then mapped to a 256-dimensional space via a linear layer to align with the sequence features:

$$\mathbf{h}_X = \mathbf{W}_{\text{proj}} \cdot \text{AveragePooling}(\mathbf{H}_X), \quad (4)$$

where $\mathbf{W}_{\text{proj}} \in \mathbb{R}^{256 \times 1024}$ is a projection matrix. Second, we selected the start token feature $\mathbf{h}_S = \mathbf{H}_S[0, :]$ from the encoded sequence features and normalize features from two modalities:

$$\bar{\mathbf{h}}_X = \frac{\mathbf{h}_X}{\|\mathbf{h}_X\|}, \bar{\mathbf{h}}_S = \frac{\mathbf{h}_S}{\|\mathbf{h}_S\|}. \quad (5)$$

Third, to align the features from the two modalities, we used the structure-knowledge-constrained triple loss as follows:

$$\mathcal{L}_{\text{kc_triple}} = e_{ij} \cdot \max\left(0, d(\bar{\mathbf{h}}_{X_i}, \bar{\mathbf{h}}_{S_i})^+ - d(\bar{\mathbf{h}}_{X_i}, \bar{\mathbf{h}}_{S_j})^- + \alpha\right), \quad (6)$$

where e_{ij} is the edge weight between the cyclic peptide i and j , $d(\bullet, \bullet)$ is the Euclidean distance between features, and α is the margin of triple loss [39] which forced the distance within the positive feature pairs $d(\bullet, \bullet)^+$ is at least closer than their negative counterparts $d(\bullet, \bullet)^-$. A positive pair is constructed by combining features from different modalities of a cyclic peptide. In contrast, a negative pair consists of one modality (either image or sequence) from a positive pair and another modality sourced from a different cyclic peptide within the same training batch. One merit of introducing pre-knowledge about the substructure of cyclic peptides is that the triple loss can dynamically adjust the degree of the distance punishment.

In the final stage of MultiCycPeptide, we concatenated the features from the two modalities together and predict the permeability of the cyclic peptides through three linear layers with activation function (ReLU [40]), reducing the dimensions from 512 to 256, 256 to 128, and finally to 1:

$$\mathbf{h}_{C_i} = \text{concat}(\bar{\mathbf{h}}_{X_i}, \bar{\mathbf{h}}_{S_i}), \quad (7)$$

$$\mathcal{L}_{\text{Permea}}(y_i, X_i, S_i) = (y_i - \text{LinearLayers}(\mathbf{h}_{C_i}))^2, \quad (8)$$

where y_i is the experimentally determined membrane permeability of the cyclic peptide. The loss function for training the MultiCycPermea is:

$$\mathcal{L}_{\text{MultiCycPermea}} = \mathcal{L}_{\text{kc_triple}} + \mathcal{L}_{\text{Permea}}. \quad (9)$$

Experimental setup

The tools used in the experiment included RDKit [41] (for all kinds of fingerprint calculation and generating cyclic peptide images) and the PyTorch platform [42]. The baseline models were all sourced from their respective GitHub open-source repositories. For encoders of MultiCycPermea, we loaded the encoders trained for 30 epochs on the pretraining tasks and then fine-tuning on cyclic peptides. The AdamW optimizer [43] was used with a learning rate of $1e-4$, and a batch size of 32 was selected for training. The margin of the triplet loss α was set to 0.1 in all experiments. Data augmentation was adopted to enhance the representation of cyclic peptides during the training stage. For the image modality, cyclic peptide images were randomly augmented using the following methods (implemented via the Albumentations Python library [44]): random rotation, cropping and padding, scaling, blurring, Gaussian noise injection, and salt-and-pepper noise injection. For the sequence modality, standard cyclic peptide SMILES were randomly converted into other non-standard SMILES representations. Each augmentation method was applied with a 50% probability during each data loading. The structural knowledge network used for triple loss is computed prior to training and queried using the peptide IDs in the dataset. All models in this work were trained for 200 epochs. The best parameters on the validation set were retained for the final evaluation on the test set. All reported metrics were the averages obtained from three independent experiments.

Data and evaluation metrics

We implemented MultiCycPermea on the current largest cyclic peptide membrane permeability dataset, CycPept-MPDB [25]. The sequence information of cyclic peptides and their logarithmic permeability values were extracted from this dataset. The cyclic peptide images were generated using the RDKit tool. We removed duplicate entries and those whose permeability was not measured through experiments. We validated the model performance under three data settings (Fig. 8, Table 4):

- (1) In-distribution (ID) setting: This setting conforms to the traditional practice of random partitioning, which entails uniformly allocating the dataset into training, validation, and test sets.
- (2) Simulated out-distribution (OD) setting: The purpose of this setting is to simulate the model's performance when predicting structures with low similarity to the known training set. We referred to the data processing methods of previous works [46] to ensure that the similarity between the training set and the test set exceeds a threshold. Specifically, this setting begins with the extraction of the Murcko scaffold from each peptide

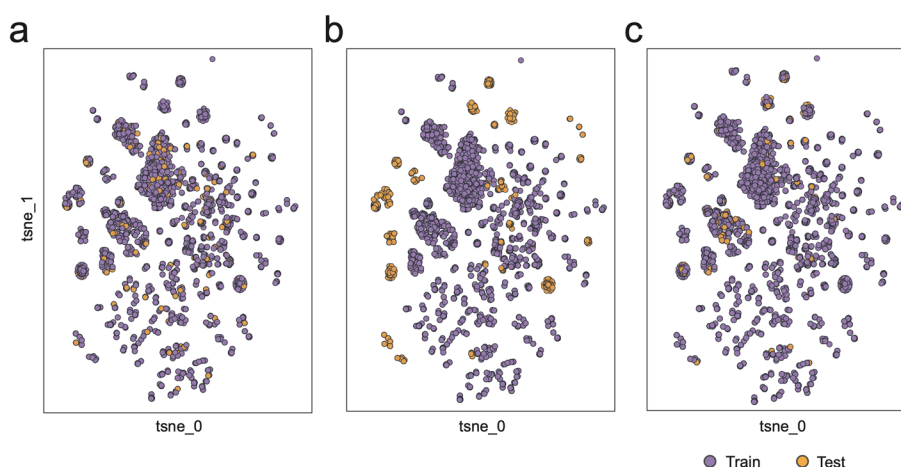


Fig. 8 The results after dimensionality reduction using t-SNE [45] based on the molecular fingerprints of cyclic peptides. **a** In-distribution setting. The training and testing data have similar distributions. **b** Simulated out-of-distribution setting. The training and testing sets are different in their distributions. **c** Permeability cliff setting. In this scenario, the distribution of the test set is not as uniform as in the ID setting. It primarily targets molecules with highly similar structures but drastic changes in membrane permeability

Table 4 The number of cyclic peptide molecules in different sets after the three data partitioning methods. It is worth noting that the numbers in parentheses for the permeability cliff setting indicate the number of known cliff data points

	Training	Validation	Test
In-distribution setting	5619	730	743
Simulated out-of-distribution setting	5554	816	722
Permeability cliff setting	5650 (2282)	720 (284)	722 (290)

$M_i = \text{MurckoScaffold}(S_i)$, followed by the computation of the Jaccard distance matrix:

$$\text{Distance}(M_i, M_j) = 1 - \text{Jaccard}(\text{MF}(M_i), \text{MF}(M_j)), \quad (10)$$

where $\text{MF}(\bullet)$ represents the Morgan fingerprints and $\text{Jaccard}(\bullet, \bullet)$ represents the Jaccard similarity. Subsequently, hierarchical clustering based on the Jaccard distance matrix was employed to categorize the data. We divided the data according to their cluster number as the training, validation, and test sets, ensuring that the minimum distance between samples from any two clusters is greater than 0.3.

[3] Permeability values cliff (Cliff) setting: This setting aims to evaluate the performance of models in distinguishing between cyclic peptides with very similar structures but significantly different permeability values. We referred to the data processing methods of previous works [47, 48] to identify pairs of cyclic peptides where a small structural change leads to a large difference in permeability (a difference of 2 in log-transformed values). We evaluated the structural

similarity between peptide pairs using three different metrics: substructure similarity, scaffold similarity, and sequence similarity. Substructure similarity is assessed by calculating the Tanimoto coefficient of two peptides. Scaffold similarity is measured by computing the Tanimoto coefficient for the scaffolds of two peptides. Sequence similarity is determined by calculating the Levenshtein distance between the SMILES representations of two peptides. We then collected pairs with any similarity greater than 0.9 and a permeability value difference exceeding 2. In total, 2856 cyclic peptide pairs with cliff relationships were identified and subsequently evenly distributed with other non-cliff peptides into the training, validation, and test sets. Note that during the evaluation phase, the metrics we presented are specifically targeted at the peptides containing cliff pairs.

For the pretraining data of the cyclic peptide image encoder and the cyclic peptide sequence encoder, we followed the datasets used in previous works. The cyclic peptide image encoder was pretrained using the PubChem database [49] and patent data collected in the MolScribe study [36]. The cyclic peptide sequence encoder was pretrained using small molecule data from ChemBERTa [50], but we randomly selected 1 million molecules from it due to the dataset's large size.

The following metrics are used to evaluate the performance of models:

- (1) MSE (mean squared error): Measures the average squared difference between the estimated values and the actual values.

- (2) MAE (mean absolute error): Measures the average magnitude of errors in a set of predictions without considering their direction.
- (3) R^2 : Indicates goodness of fit by measuring how well unseen samples are likely to be predicted by the model through the proportion of variance explained by the model.
- (4) PCC (Pearson correlation coefficient): Measures the linear correlation between two variables.
- (5) SCC (Spearman's correlation coefficient): A nonparametric measure of rank correlation that assesses how well the relationship between two variables can be described using a monotonic function.

Baselines

The baseline models can be summarized into the following three categories:

- (1) Classical machine/deep learning models for regression task:

SVM (support vector machine), RandomForest, KNeighbors, ElasticNet, DecisionTree, GradientBoosting, and GaussianProcess [51]: these methods utilize molecular fingerprints as input and then apply different forms of machine learning methods to extract features for predicting permeability of cyclic peptide.

BiLSTM and TextCNN [52, 53]: these models directly process SMILES sequences. TextCNN employs convolutional networks to identify patterns in SMILES, whereas BiLSTM processes these sequences bidirectionally through LSTM.

GAT (graph attention networks), GCN (graph convolutional networks), and GIN (graph isomorphism network) [54–56]: these models convert SMILES strings into graph, representing atoms and bonds as nodes and edges, respectively. GAT incorporates an attention mechanism to assess node importance, GCN applies graph convolutions for feature learning, and GIN is designed to capture the isomorphism properties of graphs.

- (2) Three deep learning models pretrained on small molecule datasets:

ChemBERTa [50]: It is a transformer-based model pretrained on molecular data, utilizing random masking and prediction to capture relationships within molecular sequences.

CyclePermea [57]: It is a transformer-based model pretrained on molecular data and utilizes molecular fingerprints to assist prediction.

MolCLR [58]: It is a self-supervised learning framework using large amounts of data, trained by molecular contrastive learning of representations with graph neural networks.

UniMol [59]: It is a universal 3D framework, containing a pretrained model with the SE [3] transformer architecture: a molecular model pretrained by molecular conformations.

- (3) A recently published deep learning model for permeability prediction task:

Multi_CycGT [26]: it is a multimodal permeability prediction model, based on the graph, sequence, and various molecular properties of cyclic peptides.

Abbreviations

FDA	The Food and Drug Administration
PAMPA	Parallel artificial membrane permeability assay
QSPR	Quantitative structure–property relationship
ID setting	In-distribution setting
OD setting	Simulated out-distribution setting
Cliff setting	Permeability values cliff setting
SVM	Support vector machine
GAT	Graph attention networks
GCN	Graph convolutional networks
GIN	Graph isomorphism network
MSE	Mean squared error
MAE	Mean absolute error
PCC	Pearson correlation coefficient
SCC	Spearman's correlation coefficient

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-025-02166-2>.

Additional file 1. Figures S1–S2, Tables S1–S7, and supplementary details. Fig. S1 Diagram of feature changes across layers. Fig. S2 The distribution plot of edge weights for structure constrained feature fusion. Table S1 Grid search results for margin determination. Table S2 The results of several models in the ID setting. Table S3 The results of several models in the OD setting. Table S4 The results of several models in the Cliff setting. Table S5 The ablation study results for pretraining in the ID setting. Table S6 The ablation study results for pretraining in the OD setting. Table S7 The ablation study results for pretraining in the Cliff setting. Supplementary Details: Grad-CAM procedures.

Acknowledgements

We are very much indebted to the anonymous reviewers, whose constructive comments are very helpful for this paper.

Authors' contributions

Zixu Wang and Yangyang Chen design experiments as well as writing manuscripts. Yifan Shang, Xiulong Yang and Wenqiong Pan draw figures and analyse the results. Xiucui Ye, Tetsuya Sakurai revise the manuscript. Xiangxiang Zeng provides financial help with experiments as well as revising papers. All authors read and approved the final manuscript.

Funding

X.Z. acknowledges support from the National Natural Science Foundation of China (Grant Nos. 62425204, 62450002, 62122025, U22A2037, 62432011) and the Beijing Natural Science Foundation (Grant No. L248013). X.Y.

acknowledges support from the Japan Society for the Promotion of Science (Grant No. JP22K12144).

Data availability

All data generated or analysed during this study are included in this published article, its supplementary information files and publicly available repositories. The source code for this work is available in the Github (<https://github.com/viko-3/MultiCycPermea>) and Zenodo (<https://zenodo.org/records/14795688>). The dataset used in this study is available at figshare (https://figshare.com/articles/dataset/Dataset_used_for_permeability_training/28351319?file=52,148,750).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 13 June 2024 Accepted: 17 February 2025

Published online: 27 February 2025

References

- Verdine GL, Walensky LD. The challenge of drugging undruggable targets in cancer: lessons learned from targeting BCL-2 family members. *Clin Cancer Res*. 2007;13(24):7264–70.
- Vinogradov AA, Yin Y, Suga H. Macrocyclic peptides as drug candidates: recent progress and remaining challenges. *J Am Chem Soc*. 2019;141(10):4167–81.
- Li J, Yanagisawa K, Yoshikawa Y, Ohue M, Akiyama Y. Plasma protein binding prediction focusing on residue-level features and circularity of cyclic peptides by deep learning. *Bioinformatics*. 2022;38(4):1110–7.
- Zhang H, Chen S. Cyclic peptide drugs approved in the last two decades (2001–2021). *RSC Chem Biol*. 2022;3(1):18–31.
- Valdez B, Brammer J, Li Y, Murray D, Liu Y, Hosing C, et al. Romidepsin targets multiple survival signaling pathways in malignant T cells. *Blood Cancer J*. 2015;5(10):e357–e.
- Whitty A, Zhong M, Viarengo L, Beglov D, Hall DR, Vajda S. Quantifying the chameleonic properties of macrocycles and other high-molecular-weight drugs. *Drug Discov Today*. 2016;21(5):712–7.
- Danielius E, Poongavanam V, Peintner S, Wieske LH, Erdélyi M, Kihlberg J. Solution conformations explain the chameleonic behaviour of macrocyclic drugs. *Chem—A Eur J*. 2020;26(23):5231–44.
- Lee D, Lee S, Choi J, Song Y-K, Kim MJ, Shin D-S, et al. Interplay among conformation, intramolecular hydrogen bonds, and chameleonicity in the membrane permeability and cyclophilin A binding of macrocyclic peptide cyclosporin O derivatives. *J Med Chem*. 2021;64(12):8272–86.
- Biron E, Chatterjee J, Ovadia O, Langenegger D, Bruegggen J, Hoyer D, et al. Improving oral bioavailability of peptides by multiple N-methylation: somatostatin analogues. *Angew Chem Int Ed*. 2008;47(14):2595–9.
- Bockus AT, Schwachert JA, Pye CR, Townsend CE, Sok V, Bednarek MA, Lokey RS. Going out on a limb: delineating the effects of β -branching, N-methylation, and side chain size on the passive permeability, solubility, and flexibility of sanguinamide A analogues. *J Med Chem*. 2015;58(18):7409–18.
- Hosono Y, Uchida S, Shinkai M, Townsend CE, Kelly CN, Naylor MR, et al. Amide-to-ester substitution as a stable alternative to N-methylation for increasing membrane permeability in cyclic peptides. *Nat Commun*. 2023;14(1):1416.
- Huang J, Xu Y, Xue Y, et al. Identification of potent antimicrobial peptides via a machine-learning pipeline that mines the entire space of peptide sequences. *Nat Biomed Eng*. 2023;7:797–810.
- Ji X, Nielsen AL, Heinis C. Cyclic peptides for drug development. *Angew Chem Int Ed*. 2024;63(3):e202308251.
- Ottaviani G, Martel S, Carrupt P-A. Parallel artificial membrane permeability assay: a new membrane for the fast prediction of passive human skin permeability. *J Med Chem*. 2006;49(13):3948–54.
- Witek J, Wang S, Schroeder B, Lingwood R, Dounas A, Roth H Jr, et al. Rationalization of the membrane permeability differences in a series of analogue cyclic decapeptides. *J Chem Inform Model*. 2018;59(1):294–308.
- Ono S, Naylor MR, Townsend CE, Okumura C, Okada O, Lokey RS. Conformation and permeability: cyclic hexapeptide diastereomers. *J Chem Inf Model*. 2019;59(6):2952–63.
- Cipcigan F, Smith P, Crain J, Hogner A, De Maria L, Llinas A, Ratkova E. Membrane permeability in cyclic peptides is modulated by core conformations. *J Chem Inf Model*. 2020;61(1):263–9.
- Katritzky AR, Lobanov VS, Karelson M. QSPR: the correlation and quantitative prediction of chemical and physical properties from structure. *Chem Soc Rev*. 1995;24(4):279–87.
- Liu T, Qiao H, Wang Z, Yang X, Pan X, Yang Y, et al. CodLncScape provides a self-enriching framework for the systematic collection and exploration of coding LncRNAs. *Adv Sci*. 2024;11:2400009.
- Ru X, Ye X, Sakurai T, Zou Q. NerLTr-DTA: drug–target binding affinity prediction based on neighbor relationship and learning to rank. *Bioinformatics*. 2022;38(7):1964–71.
- Gu Z-F, Hao Y-D, Wang T-Y, Cai P-L, Zhang Y, Deng K-J, et al. Prediction of blood–brain barrier penetrating peptides based on data augmentation with Augur. *BMC Biol*. 2024;22(1):86.
- Pang Y, Liu B. DisoFLAG: accurate prediction of protein intrinsic disorder and its functions using graph-based interaction protein language model. *BMC Biol*. 2024;22(1):3.
- Poongavanam V, Atilaw Y, Ye S, Wieske LH, Erdélyi M, Ermondi G, et al. Predicting the permeability of macrocycles from conformational sampling—limitations of molecular flexibility. *J Pharm Sci*. 2021;110(1):301–13.
- Digiesi V, de la Oliva RV, Vallaro M, Caron G, Ermondi G. Permeability prediction in the beyond-Rule-of 5 chemical space: focus on cyclic hexapeptides. *Eur J Pharm Biopharm*. 2021;165:259–70.
- Li J, Yanagisawa K, Sugita M, Fujie T, Ohue M, Akiyama Y. CycPeptMPDB: a comprehensive database of membrane permeability of cyclic peptides. *J Chem Inf Model*. 2023;63(7):2240–50.
- Cao L, Xu Z, Shang T, Zhang C, Wu X, Wu Y, et al. Multi_CycGT: a deep learning-based multimodal model for predicting the membrane permeability of cyclic peptides. *J Med Chem*. 2024;67(3):1888–99.
- Li J, Yanagisawa K, Akiyama Y. CycPeptMP: enhancing membrane permeability prediction of cyclic peptides with multi-level molecular features and data augmentation. *Brief Bioinform*. 2024;25(5):bbae417.
- Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training. 2018. https://cdn.openai.com/research-covers/languageunsupervised/language_understanding_paper.pdf.
- Natarajan P, Wu S, Vitaladevuni S, Zhuang X, Sakalidis S, Park U, et al. Multimodal feature fusion for robust event detection in web videos. In Proceedings of the IEEE international conference on computer vision; 2012. p. 1298–305.
- Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin transformer: hierarchical vision transformer using shifted windows. In Proceedings of the IEEE international conference on computer vision; 2021. p. 10012–22.
- Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inform Comput Sci*. 1988;28(1):31–6.
- Lee LL, Buckton LK, McAlpine SR. Converting polar cyclic peptides into membrane permeable molecules using N-methylation. *Pept Sci*. 2018;110(3):e24063.
- Frost JR, Scully CC, Yudin AK. Oxadiazole grafts in peptide macrocycles. *Nat Chem*. 2016;8(12):1105–11.
- Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE international conference on computer vision; 2017. p. 618–26.
- Salveson PJ, Moyer AP, Said MY, Gökçe G, Li X, Kang A, et al. Expansive discovery of chemically diverse structured macrocyclic oligoamides. *Science*. 2024;384(6694):420–8.
- Qian Y, Guo J, Tu Z, Li Z, Coley CW, Barzilay R. MolScribe: robust molecular structure recognition with image-to-graph generation. *J Chem Inf Model*. 2023;63(7):1925–34.

37. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Adv Neural Inf Process Syst*. 2017;30:5998–6008.
38. Devlin J, Chang M W, Lee K, et al. Bert: pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*; 2019. p. 4171–86.
39. Schroff F, Kalenichenko D, Philbin J. Facenet: a unified embedding for face recognition and clustering. *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2015. p. 815–23.
40. Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. *Proceedings of the fourteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings*; 2011. p. 315–23.
41. Landrum G. RDKit: a software suite for cheminformatics, computational chemistry, and predictive modeling. *Greg Landrum*. 2013;8(31.10):5281.
42. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: an imperative style, high-performance deep learning library. *Advances in neural information processing systems*. 2019;32.
43. Loshchilov I, Hutter F. Decoupled weight decay regularization. *7th International Conference on Learning Representations*; 2019. p. 1–8.
44. Buslaev A, Iglovikov VI, Khvedchenya E, Parinov A, Druzhinin M, Kalinin AA. Albumentations: fast and flexible image augmentations. *Information*. 2020;11(2):125.
45. Van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res*. 2008;9(86):2579–065.
46. Li S, Wan F, Shu H, Jiang T, Zhao D, Zeng J. MONN: a multi-objective neural network for predicting compound-protein interactions and affinities. *Cell Syst*. 2020;10(4):308–22 e11.
47. van Tilborg D, Alenicheva A, Grisoni F. Exposing the limitations of molecular machine learning with activity cliffs. *J Chem Inf Model*. 2022;62(23):5938–51.
48. Wu K, Yang X, Wang Z, Li N, Zhang J, Liu L. Data-balanced transformer for accelerated ionizable lipid nanoparticles screening in mRNA delivery. *Brief Bioinform*. 2024;25(3):bbae186.
49. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res*. 2018;47(D1):D1102–9.
50. Chithrananda S, Grand G, Ramsundar B. ChemBERTa: large-scale self-supervised pretraining for molecular property prediction. *arXiv preprint arXiv:201009885*. 2020.
51. Hao J, Ho TK. Machine learning made easy: a review of scikit-learn package in Python programming language. *J Educ Behav Stat*. 2019;44(3):348–61.
52. DiPietro R, Hager GD. Deep learning: RNNs and LSTM. *Handbook of medical image computing and computer assisted intervention*. Academic Press; 2020. p. 503–19.
53. Guo B, Zhang C, Liu J, Ma X. Improving text classification with weighted word embeddings via a multi-channel TextCNN model. *Neurocomputing*. 2019;363:366–74.
54. Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. *arXiv preprint arXiv:171010903*. 2017.
55. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:160902907*. 2016.
56. Xu K, Hu W, Leskovec J, Jegelka S. How powerful are graph neural networks? *arXiv preprint arXiv:181000826*. 2018.
57. Wang Z, Chen Y, Ye X, Sakurai T. CyclePermea: membrane permeability prediction of cyclic peptides with a multi-loss fusion network. *International joint conference on neural networks (IJCNN)*; 2024. p. 1–8.
58. Wang Y, Wang J, Cao Z, Barati FA. Molecular contrastive learning of representations via graph neural networks. *Nat Mach Intell*. 2022;4(3):279–87.
59. Rynefors K. UNIMOL: a program for Monte Carlo simulation of RRKM unimolecular decomposition in molecular beam experiments. *Comput Phys Commun*. 1982;27(2):201–12.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.