

Supplementary details

Grad-Cam Procedures

Implementation Steps: (1) Forward Pass: We input a cyclic peptide image-sequence pair into the model to obtain the permeability prediction. (2) Gradient Calculation: We compute the gradients of the prediction with respect to the feature maps of the final transform layer in the image encoder. These gradients indicate the importance of each feature map in contributing to the prediction. (3) Heatmap Generation: These gradients are averaged to produce weights, which are then combined with the feature maps to create a heatmap. This heatmap highlights the key regions that influence the prediction. (4) Visualization: The generated heatmap is upsampled to match the size of the original cyclic peptide image and overlaid onto it. This overlay aligns the regions (atoms or substructures) that are most influential in the model's permeability prediction.

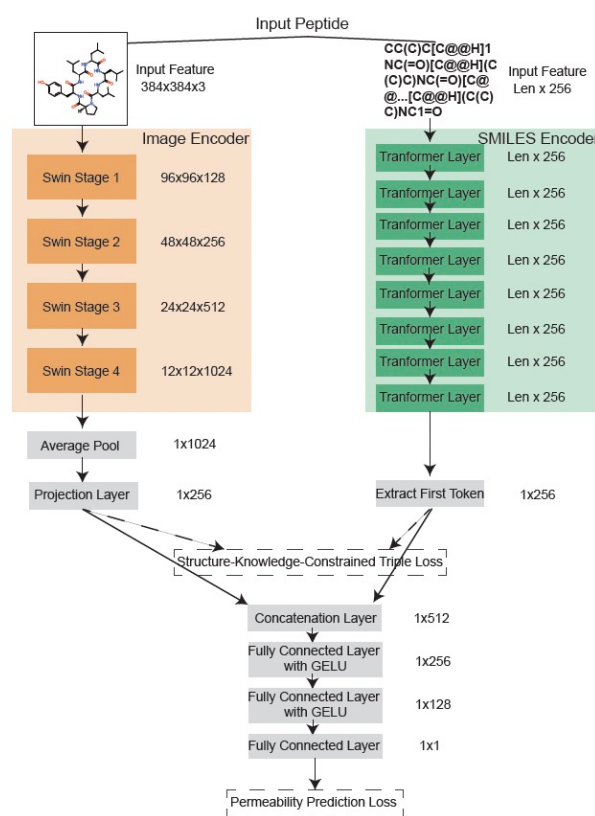


Figure S1. Diagram of Feature Changes Across Layers.

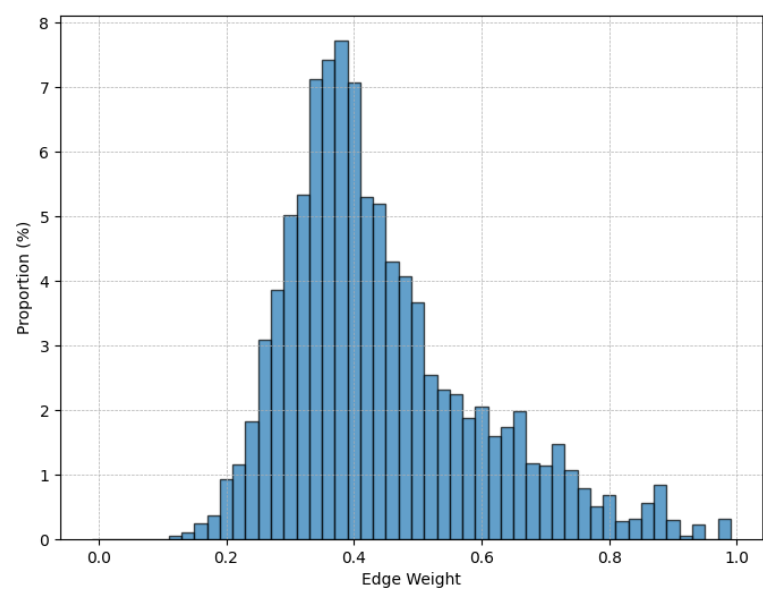


Figure S2. The distribution plot of edge weights for Structure Constrained Feature Fusion. The sampling method involves randomly selecting one edge for each cyclic peptide in the graph. A histogram of the edge weights is then created, with the y-axis representing the proportion in the histogram.

Table S1 Grid search results for margin determination.

	α	MSE ($\downarrow$)	PCC ($\uparrow$)
ID Setting	0.4	0.18	0.83
	0.2	0.17	0.83
	0.1	0.16	0.85
	0.05	0.17	0.85
OD Setting	0.4	0.46	0.47
	0.2	0.45	0.47
	0.1	0.44	0.48
	0.05	0.44	0.48
Cliff Setting	0.4	0.21	0.85
	0.2	0.18	0.87
	0.1	0.18	0.88
	0.05	0.17	0.87

Table S2 The results of several models in the ID setting.

	MSE	MAE	R ²	PCC	SCC
KNeighborsRegressor	0.35	0.42	0.40	0.66	0.58
ElasticNet	0.53	0.73	0.55	0.09	0.32
DecisionTreeRegressor	0.37	0.42	0.35	0.65	0.61
GradientBoostingRegressor	0.37	0.45	0.36	0.60	0.56
GaussianProcessRegressor	0.50	0.53	0.14	0.38	0.38
GIN	0.44	0.48	0.24	0.51	0.48

Table S3 The results of several models in the OD setting.

	MSE	MAE	R ²	PCC	SCC
KNeighborsRegressor	0.73	0.62	-0.32	0.12	0.16
ElasticNet	0.64	0.56	-0.16	0.07	0.12
DecisionTreeRegressor	0.70	0.61	-0.28	0.09	0.13
GradientBoostingRegressor	0.65	0.58	-0.17	0.17	0.19
GaussianProcessRegressor	0.68	0.64	-0.24	0.11	0.15
GIN	0.62	0.56	0.03	0.21	0.26

Table S4 The results of several models in the Cliff setting.

	MSE	MAE	R ²	PCC	SCC
KNeighborsRegressor	0.30	0.39	0.45	0.71	0.67
ElasticNet	0.47	0.52	0.16	0.46	0.45
DecisionTreeRegressor	0.33	0.39	0.40	0.68	0.65
GradientBoostingRegressor	0.29	0.40	0.47	0.69	0.64
GaussianProcessRegressor	0.29	0.39	0.48	0.69	0.66
GIN	0.55	0.56	0.24	0.50	0.50

Table S5 The ablation study results for pre-training in the ID setting.

	MSE	PCC
Image_Encoder (w/o pre-train)	0.59	0.11
Sequence_Encoder (w/o pre-train)	0.23	0.77
Image_Encoder (pre-train)	0.23	0.79
Sequence_Encoder (pre-train)	0.21	0.80

Table S6 The ablation study results for pre-training in the OD setting.

	MSE	PCC
Image_Encoder (w/o pre-train)	0.56	0.16
Sequence_Encoder (w/o pre-train)	0.55	0.30
Image_Encoder (pre-train)	0.49	0.39
Sequence_Encoder (pre-train)	0.52	0.39

Table S7 The ablation study results for pre-training in the Cliff setting.

	MSE	PCC
Image_Encoder (w/o pre-train)	0.76	0.11
Sequence_Encoder (w/o pre-train)	0.29	0.76
Image_Encoder (pre-train)	0.23	0.83
Sequence_Encoder (pre-train)	0.27	0.80