

1 Supplementary Materials

Table 1: Summary of the reviewed machine learning studies for macrocycle permeability prediction. The table provides a comparative overview of datasets, input features, modeling approaches, and reported performance.

Study (Authors)	Year	Dataset & Size	Input Representation	Model Architecture	Primary Task(s)	Performance Metrics	Availability
Over et al. [?]	2016	214 non-peptidic macrocycles (Broad DOS)	2D/3D descriptors, experimental logD	PLS, SVM, Random Forest (RF)	Regression (Caco-2 log P_{app})	PLS r^2 : 0.83 (external), 0.93 (with logD)	Data only
Digiesi et al. [?]	2021	62 cyclic hexapeptides	2D (ALogP, TPSA), 3D (VolSurf), MSM/MD	QSPR (PLS), Physics-based (ΔG_I), MSM	Regression (Caco-2, PAMPA log P_{app})	PAMPA R^2 : 0.72 (mixed model); Caco-2 R^2 : 0.83 (3D-QSPR)	Not available
Cao et al. [Multi-CycGT] [?]	2024	6941 cyclic peptides (CycPeptMPDB)	1D SMILES, 2D Graph, Physico-chemical props.	Multimodal: GCN + Transformer + MLP	Binary Classification	ACC: 0.8206; AUC: 0.8650; F1: 0.8708	Not available
Li et al. [CycPeptMP] [?]	2024	6889 peptides (CycPeptM-PDB)	Atom-, Monomer-, Peptide-level features	Hybrid Fusion: Transformer + CNN + MLP	Regression	MAE: 0.355; R^2 : 0.772; R : 0.883	Available
Yu et al. [MuCoCP] [?]	2024	7451 peptides (CycPeptM-PDB)	Molecular Graph, HELM sequence	Multimodal Contrastive: Transformer + GNN	Classification & Regression	ACC: 0.87; R^2 : 0.503; MSE: 0.714	Not available
Wang et al. [CyclePermea] [?]	2024	CycPeptMPDB	1D SMILES, Morgan Fingerprints	Pre-trained BERT; Multi-loss Fusion	Regression	MSE: 0.217 (ID), 0.411 (OD); PCC: 0.801 (ID), 0.518 (OD)	Not available
Geylan et al. [?]	2024	6876 cyclic peptides (CycPeptMPDB)	2048-bit Morgan Fingerprints	SVM, RF, LightGBM, XGBoost	Uncertainty-aware Classification	Validity: >0.80 (recalibrated); Efficiency: variable by domain	Available
Jiang et al. [CPMP] [?]	2025	6701 PAMPA, 1310 Caco-2 (CycPeptM-PDB)	SMILES, 3D Inter-atomic distances	Molecular Attention Transformer (MAT)	Regression	R^2 : 0.67 (PAMPA), 0.75 (Caco-2); MAE: 0.308	Available
Cabezón et al. [CYCLOPS] [?]	2025	CycPeptMPDB: 7236 (AllPep) and 4114 (L6/7)	Sequential (SP) & Peptide (PP) properties	Hybrid: LSTM + MLP; CyclicConv layers	Classification & Regression	ACC: 0.824; MAE: 0.477; F1: 0.87; MCC: 0.59	Available
Wang et al. [MultiCycPermea] [?]	2025	CycPeptMPDB (ID, OD, Cliff splits)	2D Images, 1D SMILES	Swin Transformer + BERT; Triple Loss	Regression	MSE: 0.16 (ID), 0.44 (OD), 0.18 (Cliff); PCC: 0.85 (ID)	Available
Feller & Wilke [Peptide-CLM] [?]	2025	10M Synthetic, SmProt, CycPeptMPDB	SMILES tokens	BERT-style (RoFormer) with RoPE	Classification & Regression	ROC-AUC: 0.820; PR-AUC: 0.948	Not available

References

- [1] Björn Over, Pär Matsson, Christian Tyrchan, Per Artursson, Bradley C Doak, Michael A Foley, Constanze Hilgendorf, Stephen E Johnston, Maurice D Lee IV, Richard J Lewis, et al. Structural and conformational determinants of macrocycle cell permeability. *Nature Chemical Biology*, 12(12):1065–1074, 2016.

- [2] Vito Digiesi, Víctor de la Oliva Roque, Maura Vallaro, Giulia Caron, and Giuseppe Ermondi. Permeability prediction in the beyond-rule-of 5 chemical space: Focus on cyclic hexapeptides. *European Journal of Pharmaceutics and Biopharmaceutics*, 165:259–270, 2021.
- [3] Lujing Cao, Zhenyu Xu, Tianfeng Shang, Chengyun Zhang, Xinyi Wu, Yejian Wu, Silong Zhai, Liefeng Ma, and Hongliang Duan. Multi_cycgt: a dl-based multimodal model for membrane permeability prediction of cyclic peptides. *bioRxiv*, pages 2023–06, 2023.
- [4] Jianan Li, Keisuke Yanagisawa, and Yutaka Akiyama. Cycpeptmp: Enhancing membrane permeability prediction of cyclic peptides with multi-level molecular features and data augmentation. *Briefings in Bioinformatics*, 25(5):bbae417, 2024.
- [5] Yunxiang Yu, Mengyun Gu, Hai Guo, Yabo Deng, Danna Chen, Jianwei Wang, Caixia Wang, Xia Liu, Wenjin Yan, and Jinqi Huang. Mucocp: a priori chemical knowledge-based multimodal contrastive learning pre-trained neural network for the prediction of cyclic peptide membrane penetration ability. *Bioinformatics*, 40(8):btae473, 2024.
- [6] Zixu Wang, Yangyang Chen, Xiucui Ye, and Tetsuya Sakurai. Cyclepermea: Membrane permeability prediction of cyclic peptides with a multi-loss fusion network. In *2024 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2024.
- [7] Gökçe Geylan, Leonardo De Maria, Ola Engkvist, Florian David, and Ulf Norinder. A methodology to correctly assess the applicability domain of cell membrane permeability predictors for cyclic peptides. *Digital Discovery*, 3(9):1761–1775, 2024.
- [8] Dawei Jiang, Zixi Chen, and Hongli Du. Cyclic peptide membrane permeability prediction using deep learning model based on molecular attention transformer. *Frontiers in Bioinformatics*, 5:1566174, 2025.
- [9] Fergal J Duffy, Mélanie Verniere, Marc Devocelle, Elise Bernard, Denis C Shields, and Anthony J Chubb. Cyclops: generating virtual libraries of cyclized and constrained peptides including nonnatural amino acids. *Journal of chemical information and modeling*, 51(4):829–836, 2011.
- [10] Zixu Wang, Yangyang Chen, Yifan Shang, Xiulong Yang, Wenqiong Pan, Xiucui Ye, Tetsuya Sakurai, and Xiangxiang Zeng. Multicycpermea: accurate and interpretable prediction of cyclic peptide permeability using a multimodal image-sequence model. *BMC biology*, 23(1):63, 2025.
- [11] Aaron L Feller and Claus O Wilke. Peptide-aware chemical language model successfully predicts membrane diffusion of cyclic peptides. *Journal of Chemical Information and Modeling*, 2025.