論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Development of Prediction Models for Membrane Permeability and Plasma Protein Binding of Cyclic Peptides with Multi-Level Molecular Features by Deep Learning
著者(和文)	李佳男
Author(English)	Jianan Li
出典(和文)	学位:博士(工学), 学位授与機関:東京科学大学, 報告番号:甲第20号, 授与年月日:2024年12月31日, 学位の種別:課程博士, 審査員:秋山 泰,石田 貴士,岡﨑 直観,村田 剛志,関嶋 政和
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Institute of Science Tokyo, Report number:甲第20号, Conferred date:2024/12/31, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

系・コース： Department of, Graduate major in	情報工学 知能情報	系 コース	申請学位（専攻分野）： Academic Degree Requested	博士 Doctor of	（ 工学 ）
学生氏名： Student's Name	李 佳男		審査員主査： Chief Examiner	秋山 泰 教授	

要旨（英文 800 語程度）
Thesis Summary（approx.800 English Words ）

Recently, cyclic peptides have been considered breakthrough therapeutic agents that boast high binding affinity, minimal toxicity, and the potential to engage "undruggable" targets such as intracellular protein-protein interactions. Recent advances in novel synthesis and display technologies have led to breakthroughs in cyclic peptide drug discovery and simplified the discovery of functional cyclic peptides against specific targets. However, the pharmaceutical utility of cyclic peptides is limited by their low membrane permeability and plasma protein binding (PPB) rate related to plasma stability, which are essential indicators of oral bioavailability and intracellular targeting. Current machine learning (ML)-based prediction methods of cyclic peptide permeability and PPB rate show variable performance owing to the limitations of experimental data. Furthermore, these methods use features derived from the whole molecule that have traditionally been used to predict small molecules and ignore the unique structural properties of cyclic peptides.

In this dissertation, to accelerate cyclic peptide drug discovery, we aim to overcome the two major challenges in establishing fast computational prediction methods for membrane permeability and PPB rate of cyclic peptides: the limited availability of experimental data and the unique structural characteristics of cyclic peptides. We proposed a multi-level molecular feature design method, which integrates peptide-, monomer-, and atom-level features to comprehensively capture the structural complexity of cyclic peptides. It was inspired by the physicochemical knowledge of membrane permeability and PPB of cyclic peptides, aimed to concurrently capture both the critical local substructures and the overall peptide structures. Additionally, to address the scarcity of cyclic peptide data, we introduced various data augmentation strategies tailored to cyclic peptides, including generating multiple 3D conformations, sequence arrangement, and SMILES enumeration, to improve model training efficiency. These techniques effectively increased the diversity of the training dataset, considering the conformational flexibility and circularity of peptides.

For the membrane permeability prediction of cyclic peptides, we first conducted an extensive search through published papers and pharmaceutical company patent documents, compiling over 7,000 cyclic peptide structures along with their experimentally measured membrane permeabilities from 47 distinct sources. Based on these data and various valuable information, such as unified sequence representations that are essential for the development of cyclic peptide drugs, we published the world's first cyclic peptide membrane permeability database, CycPeptMPDB. Then, we constructed a permeability prediction model, CycPeptMP, based on the proposed fusion model. CycPeptMP (MAE=0.355, R=0.883) outperformed several baseline models, including traditional ML-based cyclic peptide permeability prediction approaches (MAE=0.418 to 0.488, R=0.781 to 0.834), as well as state-of-the-art deep learning (DL)-based small molecule property prediction methods (MAE=0.443 to 0.591, R=0.660 to 0.825). Ablation studies demonstrated that all feature levels contributed and were relatively essential for prediction, consistent with physicochemical findings that both global and local structural changes in cyclic peptides significantly affect membrane permeability. Moreover, CycPeptMP successfully predicted peptide permeabilities, which are challenging for MD-based methods, while significantly lower computational costs.

For the PPB rate prediction, we collected 380 experimental data from collaborations with pharmaceutical companies and published literature. The fusion model performed best on the internal test set (MAE=2.44%, R=0.973), while the monomer model (CycPeptPPB) showed strong generalization on the external DrugBank dataset (MAE=4.40%, R=0.947), consistent with physicochemical findings that the substructure significantly affects the PPB rate because it forms a specific bond with plasma proteins. Furthermore, we also introduced the concept of saliency scores to analyze the contribution of each monomer to the PPB rate prediction of cyclic peptides. We confirmed that our CycPeptPPB model accurately identified key monomers that significantly influence PPB rates. The ability to identify such important monomers has significant implications for drug

development, as it enables researchers to pinpoint the structural features that enhance or reduce PPB rates. Additionally, we attempted to predict PPB rates using docking simulations, but results indicated that ligand-based methods were more effective for predicting cyclic peptide PPB rates.

The completion of this study marks the first time that large-scale cyclic peptide membrane permeability and PPB rate prediction, which cannot be evaluated by traditional screening methods, has been realized. This advancement enables the efficient design of candidate cyclic peptides and accelerates the development of cyclic peptide drugs, which are expected to revolutionize drug discovery. As a result, cyclic peptide-based medications will become available for treating more diseases, and with innovations such as oral administration, the market share of peptide drugs, which stood at only 5% in 2019, is expected to expand significantly. Furthermore, the multi-level feature design and data augmentation methods proposed in this study are not limited to cyclic peptides. For example, they can also be extended and applied to linear peptides, enabling a broader range of drug discovery and optimization efforts.