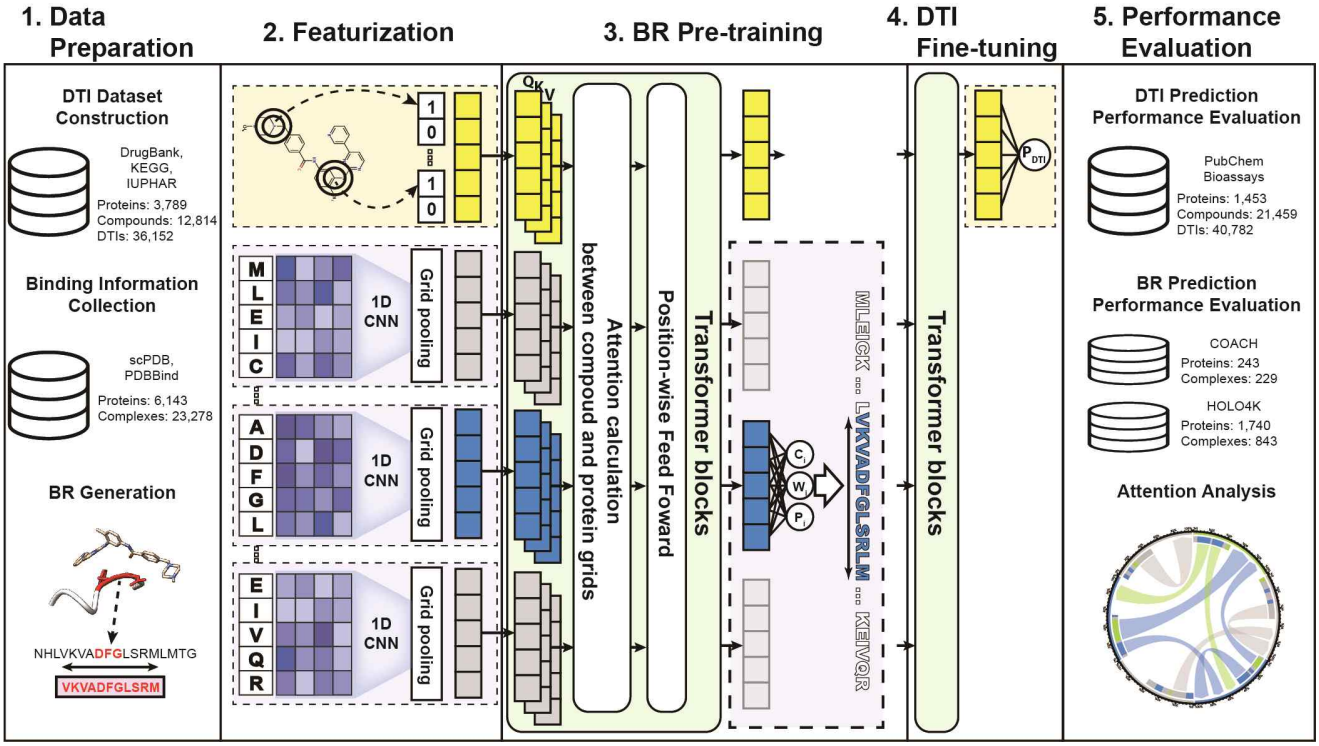
Development of artificial intelligence model to search for new drug candidates

— Development of artificial intelligence technology capable of developing new drugs without information on the three-dimensional structure of proteins

▲ From left: Professor Hojung Nam and integrated degree student Ingoo Lee

The new drug development industry, which requires astronomical time and cost, is attracting attention as an industry that can lead revolutionary change using artificial intelligence technology. By using artificial intelligence to shorten the search time for new drug candidates, the time and cost required for new drug development can be dramatically reduced.

GIST (Gwangju Institute of Science and Technology, President Kiseon Kim) School of Electrical Engineering and Computer Science Professor Hojung Nam's research team developed an artificial intelligence technology that predicts (Highlights on Target Sequence, HoTS) the binding region and interaction between drugs and target proteins based on protein sequences.



▲ HoTS model overview that comprehensively shows the training dataset, model structure, evaluation and analysis methods of the HoTS model.

Candidate discovery stage, which is the initial stage of new drug development, is the stage of finding compounds that are active on target proteins. It is a laborious process in which it is necessary to find a compound that exhibits activity on a target protein from tens and hundreds of thousands of compounds.

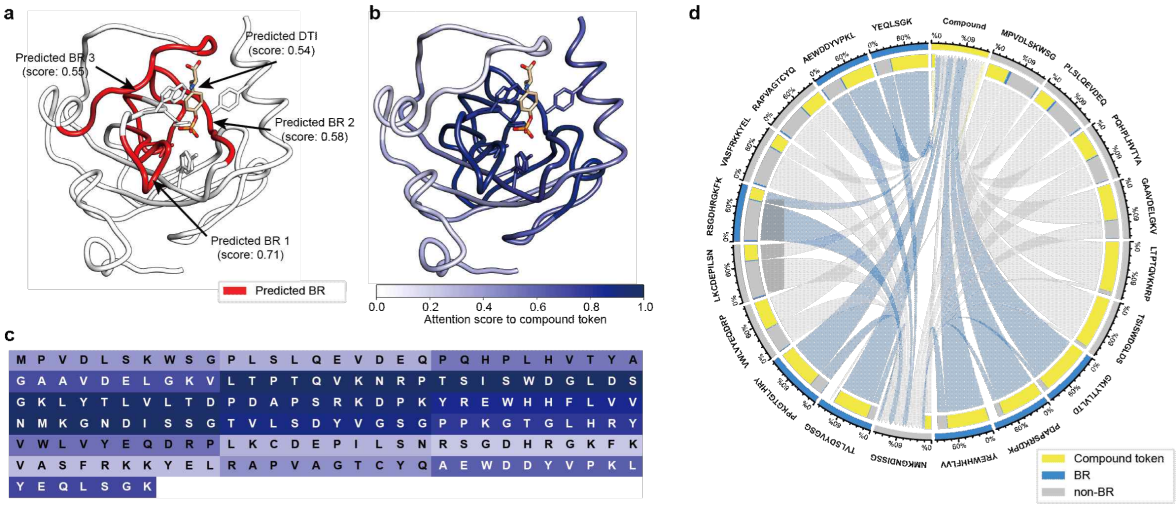
To solve this problem, various artificial intelligence models for predicting drug-target protein interaction have been developed, but, despite good predictive performance, they were reluctant to actively introduce them in actual drug development because of the lack of explanatory power for the predicted results.

However, the HoTS model developed by the research team predicts the drug-target protein binding area after pre-learning, thereby providing high prediction accuracy and evidence for drug-target protein interaction prediction as well as developing new drugs. It has become possible to present researchers with more reliable effective compound prediction results.

In this study, by extracting the binding region with the compound from a large-scale protein 3D structure database, a deep learning model based on convolutional neural network (CNN) and transformer was trained to predict the binding region on the protein sequence.

After learning the binding region, based on that learning, drug-target protein interactions can be predicted through more layers of transformers. As a result, a deep learning model can predict drug-target interactions along with binding regions.

As a result, the HoTS model showed higher predictive power than other deep learning models, and the binding region prediction was also confirmed to have a similar level of performance to other three-dimensional structure-based prediction models despite using only protein sequence information.



▲ HoTS binding region prediction and transformer attention distribution show that the HoTS transformer focuses on the binding region of the protein.

Professor Hojung Nam said, "The result of this research is a technology that greatly improves the efficiency of discovering effective compounds during the new drug development stage, and, above all, it is meaningful in that it opened up the possibility of developing new drugs for new target proteins without 3D structural information. In the future, this model will enable rapid and efficient discovery of effective compounds in the drug development stage."

This research conducted by GIST Professor Hojung Nam's team was supported by the 'Development of a system for predicting toxicity and side effects of drug candidates based on explainable artificial intelligence' (National Research Foundation's Senior Researcher Support Project), GIST-Chonnam National University Hospital Joint Research Project, and the GIST Research Institute and was published online on February 8 in the Journal of Cheminformatics.

GIST
Since 1993