GIST improves accuracy of predicting new drug candidates using AI technology utilizing AlphaFold3's 'internal representation information'

- *Professor Hyunju Lee's team from the Department of Artificial Intelligence Convergence develops 'Alpha DTA,' an AI model utilizing protein-drug 3D structures and internal AI representation information provided by AlphaFold3... Predicts new drug candidates with insufficient structural data with high accuracy*

- *Expected to reduce drug development time and costs, and facilitate drug repurposing*

- *Published in the international journal* ***Journal of Cheminformatics***

▲ ***(From left) Professor Hyunju Lee of the Department of Artificial Intelligence Convergence at GIST, master's student Minjae Chung, and Dr. Sejin Park of the Department of Electrical Engineering and Computer Science.***

To develop new drugs, researchers must first identify which substance among numerous candidate drugs binds best to the protein causing the disease. However, verifying all of this through experiments requires enormous time and cost; therefore, artificial intelligence (AI) technology, which uses computers to predict the likelihood of drug-protein binding in advance, is garnering attention as a key tool for new drug development.

To enhance the accuracy of these predictions, a GIST research team has developed a new AI model that utilizes internal expression information from 'AlphaFold3,' a world-class AI protein structure prediction model.

The Gwangju Institute of Science and Technology (GIST, President Kichul Lim) announced that a research team led by Professor Hyunju Lee of the Department of Artificial Intelligence Convergence has developed 'AlphaDTA,' an AI model capable of predicting the 'binding affinity'* of target proteins for new drug candidates without the need to verify the protein-drug binding structure through experiments.

'Alpha DTA' predicts binding affinity by utilizing the 3D structure of a protein-drug complex** predicted by Google DeepMind's AI protein structure prediction model AlphaFold3 and the internal expression information (embedding) generated during the structure prediction process.

> ** binding affinity: An indicator of how strongly and stably a drug binds to a target protein. The higher this value, the more effective the new drug candidate is considered.*
>
> *** complex: A general term for a structure formed by the combination of two or more molecules. In this paper, it specifically refers to a structure formed by the combination of a protein and a drug, and understanding its three-dimensional shape and binding method allows one to determine how well the two substances interlock.*

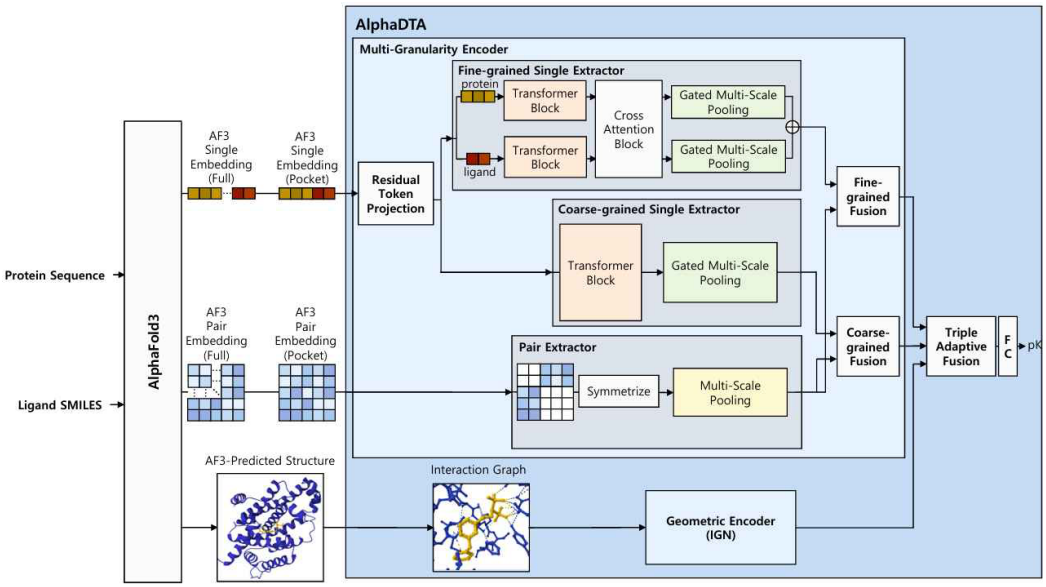
Until now, AI technology for predicting the binding affinity of drugs and proteins has largely evolved in two ways: 'sequence-based' and 'structure-based.'

The 'sequence-based' method, which utilizes only the sequence information of proteins and drugs, has a large amount of available data but may suffer from lower accuracy because it fails to reflect the actual binding form.

On the other hand, the 'structure-based' method, which utilizes the three-dimensional structure of protein-drug complexes identified through experiments, offers high accuracy but has the limitation of having very little experimental structural data.

To address these limitations, the research team focused on the 'embedding' generated by AlphaFold3 during the process of predicting the binding structure of proteins and drugs.

Embedding is numerical information designed to enable the AI to learn the chemical and structural characteristics of proteins and drugs. It consists of ▲ 'single embedding,' which captures the characteristics of the protein and drug individually, and ▲ 'pair embedding,' which captures the relationship and interaction between the two substances.

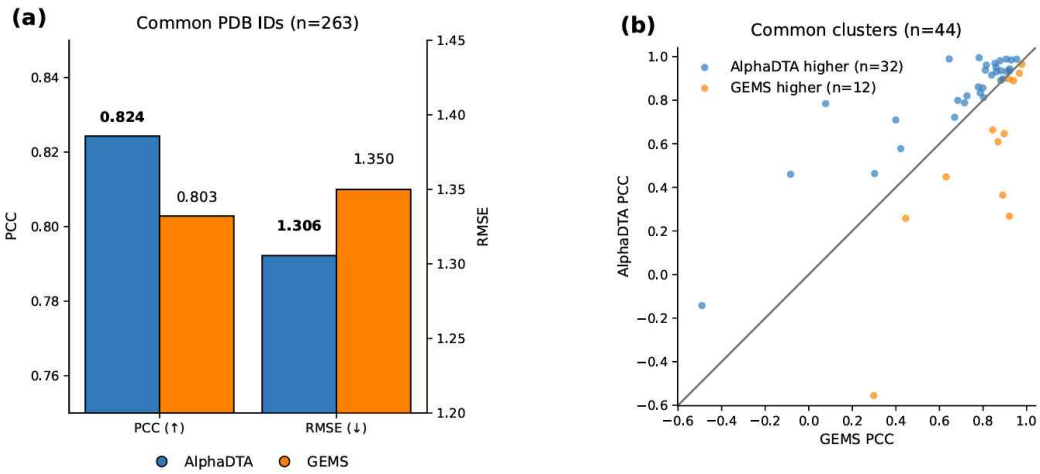


▲ ***Schematic diagram of Alpha DTA operation. When a protein sequence and the chemical structure of a drug are input, AlphaFold3 generates the 3D structure of the complex along with internal expression information (single expression vectors, paired expression vectors). AlphaDTA combines this structural information and expression information to predict binding affinity.***

The research team developed an AI model called "Alpha DTA" that combines a "multi-granularity encoder," which analyzes single- and paired expression information, with a "geometric encoder," which analyzes the three-dimensional structural information of protein-drug complexes.

"Alpha DTA" takes protein sequences and drug information as input and predicts binding affinity by analyzing the 3D structure generated by AlphaFold3 along with internal expression information.

Verifying its prediction performance with new protein-drug combinations not used in training, "Alpha DTA" demonstrated higher prediction performance than existing sequence-based methods and achieved a level of prediction performance comparable to existing structure-based methods even without experimentally obtained structural data.



▲ ***Comparison of prediction accuracy between Alpha DTA and the existing best-performing model (GEMS). Performance was compared based solely on the structures predicted by AlphaFold3, without any structures confirmed by experiments. (a) When comparing the accuracy (PCC, higher is better) and error (RMSE, lower is better) of binding affinity prediction, AlphaDTA (blue) showed higher accuracy and lower error than GEMS (orange). (b) When comparing the prediction accuracy of the two models for each of 44 different protein clusters, AlphaDTA (dot diagonally upward) showed superior performance to GEMS in 32 of the 44 clusters, indicating that it consistently outperformed a wide range of protein types without being limited to specific cases.***

This technology is expected to be widely utilized in drug repurposing research, as it can maintain high prediction performance even for new drug candidates for which experimental structural data is scarce, thereby increasing the efficiency of drug candidate screening and reducing the time and cost of new drug development.

> ** drug repurposing: A strategy of utilizing drugs already approved for the treatment of other diseases to treat new diseases different from their original purpose.*

Professor Hyunju Lee said, "This research is significant in that it demonstrates the ability to predict the binding affinity of new drug candidates with high accuracy, even without experimentally verifying the protein-drug binding structure. We expect this to contribute to efficiently selecting promising candidates for various target proteins lacking structural data, thereby increasing the speed and efficiency of new drug development."

This study, supervised by Professor Hyunju Lee of the GIST Department of Artificial Intelligence Convergence, led by master's student Minjae Chung as the first author, and featuring Dr. Sejin Park, was supported by the Ministry of Science and ICT and

the National Research Foundation of Korea's Mid-Career Researcher Support Program, the Ministry of Science and ICT and the Korea Institute of Information and Communication Technology Planning and Evaluation's Software Computing Industry Original Technology Development Program, and the Ministry of Science and ICT's GIST-InnoCORE program.

The research results — AlphaDTA: integrating AlphaFold3 embeddings and 3D complex structures for drug–target binding affinity prediction — were published online on July 21, 2026, in the international academic journal *Journal of Cheminformatics*.

Meanwhile, GIST stated that this research achievement takes into account both its academic significance and potential for industrial application, and that discussions regarding technology transfer can be conducted through the Technology Commercialization Center (hgmoon@gist.ac.kr).