September 30, 2020

Interprotein Corporation (Interprotein) to identify inhibitors for an undisclosed protein-protein interaction (PPI) target in collaboration with Janssen Biotech Inc.

One of the major problems for pharmaceutical companies is low productivity of drug research and development (R&D), and the purpose of application of AI to drug R&D is to improve the productivity. The major cause of the low productivity is mainly failures in phase II studies, which is often called phase II attrition **[1]** , and the biggest reason for the failure is lack of efficacy **[2]** . These facts suggest that the most important factors in drug R&D are "validity of concept (including selection of drug target protein and binding site)" and "selection of the compound that binds to proposed binding site with appropriate binding pose". Almost all biotech companies conducting AI-based drug discovery are comprehensively supporting selection of new drug candidates including toxicological and pharmacokinetic aspects. On the other hand, Interprotein is focusing on binding pose, binding affinity and activity prediction of human proteins based on the above-mentioned point.

On a basis of this concept, Interprotein established a new AI-implemented activity prediction system and named it AI-guided INTerprotein's Engine for New Drug Design (INTENDD). AI-guided INTENDD was constructed based on the accumulation of the many experiences and successful results obtained from the examination with INTENDD as a basal technology, which is a proprietary binding mechanism-based *in silico* screening system with unique algorithm (but not universal algorithm) or each target pocket discriminated from conventional binding energy-based docking methodologies **[3] .**

Table 1 shows the differences between main current AI-based drug discovery technologies and AI-guided INTENDD from the aspect of activity prediction of small molecules. Approach of AI-guided INTENDD is purely structure-based, which enables to newly identify active compounds without pre-reported active ligand information. Unique training data represent 3D-

and atomic-level (but not partial structure-level) interactions between target proteins and compounds with known bioactivities. These training data are produced by proprietary data-preprocessing method. Such pre-preprocessing approach is contributing to short-term machine-learning and activity prediction without supercomputer. Predicted activity can be divided into 8 classes based on the order of active concentration. Furthermore, it has been clarified that challenging drug targets such as ubiquitin-proteasome system-related proteins **[4]** are applicable as well as PPIs **[5]** .

Table 1. Differences between main current AI-based drug discovery technologies and AI-guided INTENDD from the aspect of activity prediction of small molecules.

Compared point	Main current technologies	AI-guided INTENDD
Main aim	Comprehensive support of selection of new drug candidates including toxicological and pharmacokinetic aspects	Focused on selection of highly active compounds for lead optimization
Approach for activity prediction	Ligand-based or mixture of ligand- and structure-based	Purely structure-based
Training data for deep learning	Mainly partial structure information of ligands at 2D- or 3D-level	Atomic information at 3D-level (including unique hyper parameters produced by proprietary data preprocessing method)
Balance of enthalpy and entropy	Not clearly reported	Precise prediction of compound with a good of enthalpy and entropy
Predicted activity classification	2 Classes	8 Classes
Main target	Non-PPIs	PPIs

In such circumstances, Interprotein and Janssen Biotech Inc. (Janssen), one of the Janssen Pharmaceutical Companies of Johnson & Johnson, have decided to enter a research collaboration to identify inhibitors for an undisclosed challenging target. The collaboration will focus on drug discovery research for a specific protein-protein interaction (PPI) target (autoimmune disease area). The collaboration was facilitated by the Johnson & Johnson Asia Pacific Innovation Center.

Under the agreement, Interprotein and Janssen will collaborate to identify active compounds with new scaffolds and high potencies by combining Janssen's outstanding drug discovery expertise and Interprotein's platform technologies, INTENDD®/AI-guided INTENDD® (see News Release at the Interprotein's Web site). The structure of the agreement has not been disclosed.

Interprotein is currently conducting *in silico* screening from around 10 million compounds that we can readily purchase in Japan, and based on our experience, the potency of hit compounds seems to be limited to 100 nmol/L order in the best case. For promotion of the process for lead optimization, it is preferable to obtain more potent compounds as early as possible, but such compounds are rarely included in the commercially available real compound libraries. To solve this problem, we are currently trying to establish an AI system which predict activities of virtual compounds. As the sources, patent compounds, medicinal chemist-designed compounds and commercially available virtual compounds. Reymond reported that the number of virtual compounds with selected existing partial structures was estimated to be 166.4 billion when the number of molecular-composing atoms was increased up to 17 types **[6]**. Interprotein's proprietary data-preprocessing method and resultant unique features are expected to contribute to short-term activity prediction of huge number of compounds without supercomputer and improvement of productivity of drug R&D.

[1] Patel, D. D., Antoni, C., Freedmen, S. J., Antoni, C., Freedman, S. J., Levesque, M. C., and Sundy, J. S. (2017) *J. Allergy Clin. Immunol.*, **140**, 685-687.
[2] Harrison, R. K. (2016) *Nat. Rev. Drug Discov.*, **15**, 817-818.
[3] Komatsu, H., Tanaka, T., and Matsuzaki, T. (2019) Medical Science Digest, 45, 225-231.
[4] 内藤幹彦, 小松弘嗣 (2020) 日薬理誌, in press.
[5] Komatsu, H., Ikeda, K., Tanaka, T., and Matsuzaki, T. (2019) *J. Biol. Macromol.*, **19**, 5-19.
[6] Klebe, G. (2015) *Nat. Rev. Drug Discov.*, **14**, 95-110.

Hirotsugu KOMATSU, Ken IKEDA, Takeshi TANAKA and Masato HOSODA
Interprotein Corporation