

## 機械学習を用いた高精度なドッキングスコアの開発

(名大院理<sup>1</sup>・名大 ITbM<sup>2</sup>) ○南 翔太<sup>1</sup>・藤本 和宏<sup>1,2</sup>・柳井 毅<sup>1,2</sup>

*Development of a highly accurate docking scoring function using machine learning* (<sup>1</sup>Graduate School of Science, Nagoya University, <sup>2</sup>WPI-ITbM, Nagoya University) ○Shota Minami,<sup>1</sup> Kazuhiro J. Fujimoto,<sup>1,2</sup> Takeshi Yanai<sup>1,2</sup>

Structure based drug design (SBDD) is expected to greatly contribute to reducing the enormous cost of new drug development, and indeed many drugs have been developed using SBDD. Among several SBDD methods, docking simulation is a powerful tool for predicting protein-ligand binding conformations and ligand binding affinities based on computer simulations. While currently available methods are able to predict the ligand binding conformations with high accuracy, they remain incapable of accurately predicting the ligand binding affinity. In this study, we add fingerprints that describe the structure of the ligand and the surrounding protein information as new descriptors to the scoring function, and perform regression analysis using machine learning models such as Lasso and LightGBM. As a result, we succeeded in developing a scoring function to achieve highly accurate binding affinity prediction.

**Keywords :** Docking Simulation; Machine Learning; Fingerprint **【最大 5 words】**

標的タンパク質の立体構造情報に基づく化合物設計(SBDD : Structure based drug design) は新薬開発にかかる莫大なコスト削減に大いに貢献できるものとして期待されており、SBDD を用いた多くの薬剤が開発されている。SBDD にはいくつかの手法があるが、タンパク質-リガンド間の結合様式やリガンドの結合親和性を、コンピューターシミュレーションを用いて予測する手法はドッキングシミュレーションと呼ばれる。現在提案されている手法はリガンド結合様式を高い精度で予測可能となっているが、リガンドの結合親和性を精度よく予測できないという問題を抱えている。そこで、本研究では、PMFScore と呼ばれるスコア関数に新しい記述子としてリガンドの構造やリガンド周辺タンパク質の情報を記述するフィンガープリントを加え、Lasso 回帰や LightGBM といった機械学習の手法を用いて回帰分析を行うことで高精度な結合親和性予測を実現するためのスコア関数を開発した。

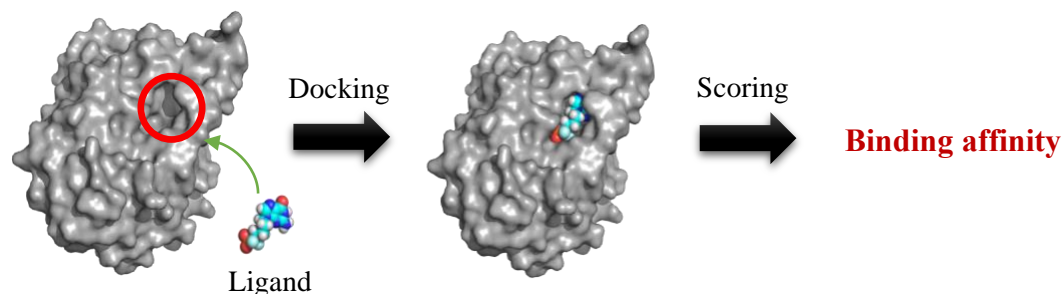


Fig 1 ドッキング計算の流れ