PLCAI: Systematic Pipeline for Protein-Ligand Binding Prediction via Multi-Scale Convolution Neural Network With Applications in Interpretable Drug Design Process

Poysungnoen, Koravit (School: Triam Udom Suksa)
Manoret, Pongpak (School: Triam Udom Suksa)

Protein-ligand binding is at the heart of interaction between human or pathogen cells and ligand molecules. Among ligands, drug molecules usually bind specifically to their target proteins as the mechanism of action. In silico prescreening of ligand molecules can help accelerate the progress of drug discovery and reduce the cost of development significantly. This is particularly useful for both emerging infectious diseases (e.g. COVID-19) and non-communicable diseases (e.g. cancer). Herein, we proposed a new systematic pipeline to improve the current drug prescreening protocol consisting of 1) a deep learning model to predict protein-ligand interaction using amino acid sequences and ligand SMILES string with minimal preprocessing, followed by 2) postprocessing mutational scanning analysis for the result interpretation. To effectively encode the features of protein input, a stack of multi-scale convolution neural networks, each with different kernel sizes were designed to capture the local residue interaction patterns across the sequence. We also optimized the classification performance using a soft label technique. The model achieved a remarkable precision of 59.26%, recall of 88.18%, F-1 of 70.88%, and AUC-PR of 75.97% in the hold-out BindingDB testing set. Deep mutational scanning analysis described the importance of each residue to the binding between one protein and several of its ligand candidates. Furthermore, the web application for the binding prediction is available on Salesforce’s Heroku servers for? the general public. Finally, our prescreening pipeline can save the screening cost up to 25%, while losing only 10% of active compounds, and also offer the potential explanation behind each prediction for the future application in terms of new drug designs.

Awards Won: