Machine Learning-based (MLD) for Ligand-binding Site Prediction

Kowit Hengphasatporn
Center for Computational Sciences, University of Tsukuba

1. Project Purpose

This study aims to employ a machine learning (ML) model to identify and predict the binding site of siglec proteins. The first step involves generating a series of protein structures using molecular dynamics (MD) simulation. Specifically, the study focuses on singles-7, -9, and -15, which are cancer-signaling proteins known to bind to glycans. After that, molecular docking will be performed to predict the glycan binding site on each structure of the siglec protein. Then, ML model will be used to characterize the binding site and propose the new potential binding region.

2. Results

A total of 300 structures of siglec-7, -9, and -15 for molecular docking have been done and prepared as a receptor for molecular docking using autodock Vina 1.2.1. We performed docking for 100 iterations for 1 structure in a total of 900,000 docked structures.

3. Roles of the MCRP and its significance

MCRP was used to perform MD simulation for siglec-7, -9, and -15 and conduct the molecular docking study. Also, we plan to dock these siglecs with related proteins.

4. Future plan

Using the ML model, we plan to work on the docking data that contains about 100,000 structures.

5. Publications and conference presentations

(1) Journal papers

Kongkaew, N.; **Hengphasatporn, K.**; Injongkol, Y.; Mee-udorn, P.; Shi, L.; Mahalapbutr, P.; Maitarad, P.; Harada, R.; Shigeta, Y.; Rungrotmongkol, T.; Vangnai, A. S., Design of electron-donating group substituted 2-PAM analogs as antidotes for organophosphate insecticide poisoning. *RSC Advances* **2023,** *13* (46), 32266-32275.

(2) Presentations

-

(3) Others

-

Supercomputer	Use	Allocated resources*	
		Initial resources	Additional resources
Cygnus	Yes	12000	-
Wisteria/BDEC-01	No	-	-
*in units of node-hour product			