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

Kowit Hengphasatporn
Center for Computational Sciences, University of Tsukuba

1. Project Purpose

The objective of this study is 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 siglec-7, -9, and -15, which are cancer signaling proteins known to bind to glycans. After that, the molecular docking will be performed to predict the glycan binding site on each structure of siglec protein. Then, ML model will be used to characterize the binding site and prepose the new potential binding region.

2. Results

A total of 300 structures of siglec-7, -9, and -15 for molecular docking have 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 total is 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.

4. Future plan

We plan to work on the docking data that contain about 100,000 structures using ML model.

5. Publications and conference presentations

(1) Journal papers
(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			