[BIOINFO 2024]

Scoring method of POM compound activity prediction for *Mycobacterium tuberculosis*

Jihyeon Yun[1,2], and Inhee Choi[1,2,*]

[1] *Advanced Drug Discovery & Development, University of Science & Technology*
[2]*AI Drug Discovery, Institut Pasteur Korea*
**Corresponding author: inhee.choi@ip-korea.org*

Tuberculosis (TB), which is caused by bacteria of the *Mycobacterium tuberculosis* complex, is one of the oldest diseases known to affect humans and a major cause of death worldwide. Thus, it is important to discover new TB inhibitors. In this study, the Principal Component Analysis (PCA) and Exploratory Data Analysis (EDA) were performed to analyze chemical structural and physicochemical features to understand TB inhibition activities. For this, anti-tubercular compounds with activities from the PubChem Bioassay were used for data analysis. Through the PCA and EDA analyses, 7 RDKit descriptors were selected to predict good TB inhibition activity. Those descriptor values were used as RDKitScore (R). For HyperScore (H), 3 parameters from HYPER LAB were used to score compounds. Finally, RH scoring method based on RDKitScore and HyperScore was defined for TB inhibition activity prediction. To predict, two groups of compounds were used. One is spirocyclic POM compounds which are targeting TB's MmpL3 (Mycobacterial membrane protein Large 3) and have experimental TB inhibition activity data. The other is HyperDesigned compounds which were generated by HYPERLAB for suggesting new POM derivatives. In this RH scoring method, group 2 (HyperScore≥4, 2.3≤RDKitScore≤3.6) for POM compounds and group3 (HyperScore≥4, 3.6<RDKitScore) for HyperDesigned compounds were the best groups. To predict how these compounds bind well with MmpL3, homology modeling of TB MmpL3 with bound ligands by SWISS-MODEL and Glide were used for docking. As the result, the group 2 has lowest binding score for POM compounds, it means that group 2 compounds have high possibilities to bind with MmpL3 stably. For HyperDesigned compounds, group 3 has lowest binding score. As the result, the best groups of RH scoring method and docking results showed correlation where best groups have lowest docking energy. Now newly designed POM compounds can be predicted for their TB inhibition activities based on RH scoring method combined with docking.