How can we compare differenct drug target classes using 3d-similarity?

Sang-Hyeok Lee[1]

[1]*Shipping Big Data Analysis Center, Korea Maritime Istitute*

June 17, 2021

This presentation describes that applied mathematics can help to discover the new drug. We solve this problem in the chemoinformatics 'as we find unprecedented materials, which of proteins in our body does interact with the new ones?' using applied math including machine learning. The the summary of our study is as follows.

3D similarity is useful in predicting the profiles of unprecedented molecular frameworks that are 2D dissimilar to known compounds. When comparing pairs of compounds, 3D similarity of the pairs depends on conformational sampling, the alignment method, the chosen descriptors, and the similarity coefficients. In addition to these four factors, 3D chemocentric target prediction of an unknown compound requires compound–target associations, which replace compound-to-compound comparisons with compound-to-target comparisons. In this study, quantitative comparison of query compounds to target classes (one-to-group) was achieved via two types of 3D similarity distributions for the respective target class with parameter optimization for the fitting models: (1) maximum likelihood (ML) estimation of queries, and (2) the Gaussian mixture model (GMM) of target classes. While Jaccard–Tanimoto similarity of query-to-ligand pairs with 3D structures (sampled multi-conformers) can be transformed into query distribution using ML estimation, the ligand pair similarity within each target class can be transformed into a representative distribution of a target class through GMM, which is hyperparameterized via the expectation–maximization (EM) algorithm. To quantify the discriminativeness of a query ligand against target

classes, the Kullback–Leibler (K–L) divergence of each query was calculated and compared between targets. 3D similarity-based K–L divergence together with the probability and the feasibility index, (Fm), showed discriminative power with regard to some query–class associations. The K–L divergence of 3D similarity distributions can be an additional method for (1) the rank of the 3D similarity score or (2) the p-value of one 3D similarity distribution to predict the target of unprecedented drug scaffolds.