Utilization of Hybrid Fragmentation Fingerprint in SARMs Dataset

Rezi Riadhi, Nara Institute of Science and Technology

Virtual screening (VS) of compounds against a target macromolecule has proved its usefulness in many drug discovery projects by efficiently identifying active compounds at low cost. Given diverse compounds, machine learning (ML)-based VS is improving the efficiency of different drug discovery stages. However, in the case of using a set of analogous active compounds, such as a structure-activity relationship matrix (SARM) dataset, the usefulness of ML-based VS for diverse compound selection has not been apparent. SARM datasets can mimic a situation in the early phase of drug discovery when analogous active compounds are identified. One reasonable approach is to utilize both intermolecular interaction information between a ligand and the target macromolecule, and the structural information of the ligand: a hybrid of ligand- and structure-based methods. In this study, we focus on fingerprint (FP) as a representation for ML and propose a new FP, named FIFI (Fragmented Interaction FIngerprints) as a way to utilize the combined information. The fingerprint is based on interaction fingerprints and atom-environment fingerprints. Benchmark calculations using a set of SARMs data sets against four target macromolecules showed that FIFI had comparable performance to other approaches in VS including another state-of-the-art fragmentation fingerprint, PLEC.