

The screening of chemical libraries is an important step in the drug discovery process. The existing chemical libraries contain up to millions of compounds. As the screening at such scale is expensive, the virtual screening is often utilized. There exist several variants of virtual screening and ligand-based virtual screening is one of them. It utilizes the similarity of screened chemical compounds to known compounds. Besides the employed similarity measure, another aspect greatly influencing the performance of ligand-based virtual screening is the chosen chemical compound representation.

In this thesis, we introduce a fragment-based representation of chemical compounds. Our representation utilizes fragments to represent a compound. Each fragment is represented by its physicochemical descriptors. The representation is highly parameterizable, especially in the area of physicochemical descriptors selection and application. In order to test the performance of our method, we utilized an existing framework for virtual screening benchmarking. The results show that our method is comparable to the best existing approaches and on some datasets it outperforms them.