

Virtual screening is a part of computer-aided drug design, which aims to identify biologically active molecules. The ligand-based virtual screening employs known biologically active molecules and similarity search. A common approach to computation of molecular similarity is to utilize molecular fingerprints. Hashed structural molecular fingerprints hash fragments (subgraphs) of molecular graphs into a bit string reducing the problem of molecular similarity to the bit string similarity. Due to the hashing two distinct fragments may collide, which causes information loss. For this reason collisions are considered unwanted and they are generally believed to decrease a performance. Our goal was, contrary to the general believe, test whether collisions can have positive impact on the performance. For this purpose we designed several similarity models based on fragments. In order to make testing and evaluation easy we implemented testing environment. Results of our experiments prove that some collisions can outperform commonly used methods. Moreover some collisions in a specific model can lead to a performance of AUC over 0.99.