Abstract

There is a huge number of chemical compounds that is collected as free access database for worldwide using. Several molecules show potential to use as pharmacological compounds. If we can screen and selected only some compound for experimental testing, the huge of cost and time for investigation could be reduced, especially when the time of epidemic. So several of drug screening techniques are of great interested and developed. The structure of chemical compounds can be represented with molecular graph which theories and algorithms of graph can be applied for structural searching or analyzing, e.g. substructure searching, structure similarity searching, common substructures analysis etc., which are important information for pharmacological drug screening when we have already some used drugs or chemical compounds. The 3D and 2D molecular structures can be reduced and represented by text in one dimension in the form of SMILES (simplified molecular-input line-entry system) format. It can be also reverse transformed back to 2D and 3D, respectively. So it is one format that widely used in molecular databases. In this research we develop the algorithm for extracting structure data from SMILES format. The meaning molecular parts or fragment are defined as "Meaning Fragments", and denoted by MFs for use in the molecular reduced graphs. A similarity score RTCS (Refinement Tanimoto Coefficient of Similarity) is proposed and is obtained based on types and number of MFs as appear in both and each of pair of compared molecules. Then the RTCS is used as the screening criteria in comparison with original TCS (Tanimoto Coefficient of Similarity) that is generally used widely. The screening of drug for SARS coronavirus proteinase is used for case study. The known five drug structures, Ribavirin, Lopinavir, Ritonavir, Calanolide A, Navirapine and 3C-LIKE proteinase are obtained in free accessing databases, e.g. PubChem and DrugBank. The results show that the obtained RTCS values are most higher than the corresponding TCS values, and also yields more hit molecules but not too much to use as input of molecular docking step.