Abstract

There are several approaches to cluster chemical structures. Classical 2D descriptor (for example chemical fingerprints or ECFP) and structure-based methods are the most widely used. Considering 3D information such as conformers, 3D pharmacophore maps or molecular shapes can give more insight thus facilitates natural interpretability of the results.

ChemAxon’s alignment tool provides automatic 3D shape-based flexible alignment of small molecules. The shape similarity score calculated for the best fit can serve as a basis for similarity-based clustering.

Introduction

It is generally accepted that molecular shape properties play a central role in ligand binding. According to the growing number of publications in this field, several descriptors and methods have been applied for shape-based similarity [1]. The usage of these shape-based methods are quite competitive with other virtual screening techniques, such as ligand and structure-based methods [2, 3]. Therefore, it is also expected that considering 3D shape alignment based similarity in clustering may also bring new and novel aspects besides the information provided by traditional 2D clustering methods.

ChemAxon in 3D

3D structure generation/conformational analysis

Generation of 3D structures is an important part of the drug discovery process. It allows to extend knowledge regarding the relationship of structure and activity. Several methods are available to generate 3D structures starting from a chemical structure. These methods can be divided into two categories: the first type is based on force-field calculations (e.g., Flex3D [4]), and the second type is based on molecular mechanics or semi-empirical calculations (e.g., Marvin [5]).

3D flexible alignment

The 3D flexible alignment procedure (released in 2005 [6]) overlays two structures by maximizing the intersection of their van der Waals volumes. The volume is partitioned by the underlying atomic properties, like extended atom types (H-bond types) or pharmacophoric types. Both molecules can be treated consistently, that is, by exchanging their rotatable bonds, flexible rings and ring systems in a continuous manner during the alignment. Single 2D conformer for each aligned structure is used as input for the alignment procedure. Thus, this method provides valid 3D-similarity scores for 2D input structures by automatically calling Generate3D 3D.

3D similarity - ligand based virtual screening

Screen3D is a ligand based 3D similarity calculation tool released in 2010Q3. Screen3D can return the intersection of the colored shape and the 3D Tanimoto, a dimensionless measure of similarity between 3D and 2D. A matrix of weight factors describes how much a shape part assigned to atom type “j” of the first molecule can be matched by a shape part assigned to atom type “i” in the second molecule. The identity matrix is used in this work: the intersections of the shapes of the same atom types is considered only.

JKlustor

JKlustor Suite [9, 10] performs similarity and structure-based clustering of compound libraries and focused sets in both hierarchical and non-hierarchical fashion as well as it carries out diversity calculations and library comparisons based on molecular fingerprints and other descriptors. It is an essential tool in combinatorial chemistry, virtual library design or other areas where a large number of compounds need to be analyzed.

The currently presented approach introduces 3D flexible alignment-based similarity calculation to the JKlustor Suite. This allows the available similarity based algorithms to use structural data in the clustering processes.

Proof of concept implementation

Interface to 3D flexible alignment functionality has been implemented in JKlustor by providing transparent pairwise similarity calculation. Visualization tools were also provided in order to compare the results of the alignment based similarity calculation with other descriptor implementations.

Clustering results of a small 3D fragment library

For a selected cluster (first one in the list above) all pairwise dissimilarities between cluster elements and all pairwise dissimilarities between cluster and non-cluster elements.

References