I Bangov, J Comput Sci Syst Biol 2014, 7:5
http://dx.doi.org/10.4172/0974-7230.S1.009

OMICS Group Conferences
Accelerating Scientific Discovery

2nd International Summit on
Integrative Biology
August 04-05, 2014 Hilton-Chicago/Northbrook, Chicago, USA

QSAR/QSPR Investigations and cluster analysis based on descriptor fingerprints

I Bangov
Konstantin Preslavski Shumen University, Bulgaria

Structural fingerprints have been widely used in structural similarity search procedures both in QSAR and QSPR investigations. Recently, it was proposed a new type of fingerprints, which is called descriptor fingerprints. A descriptor fingerprint is created by determining an interval and the resolution of its possible values for each descriptor, e.g. a initial value *initVal* and ending value *endVal* with a resolution step *resVal*. Hence, for each descriptor its interval is divided bitwise into *N=(initVal-endVal)/resVal* subintervals which are further concatenated, thus the different descriptors form the fingerprint itself. Then, the program determines in which bit the current descriptor value falls and put a bit 1 in this element, the other elements remaining zeros. Furthermore, in the case of small molecules, this approach allows the creation of mixed fingerprints consisting of both descriptor and structural fingerprints. Hence, both physical and structural properties are jointly coded and exploited in the process of similarity search. Database of either chemical or any material fingerprints (descriptor only or descriptor + structural) is created and the similarity of each query object is compared with the objects from this database by using Tanimoto criterion T: T=*NC/(NA+NB-NC)*

Where N_A is the number of 1s in the fingerprint of the first object, NB is the number of 1s of the second object fingerprint and NC is the number of 1s being in the same positions in both fingerprint arrays. The fingerprints are compared pairwise. The Tanimoto index has values in the range from 0.0 to 1.0. The higher is the value the more similar are the objects of a studied pair of fingerprints. It is expected, on the one hand, the most similar objects (with highest Tanimoto indices) to an object of the database to have some properties (biological activity, composition, etc.) similar to this object. On the other hand, the descriptor fingerprint similarity analysis allows the implementation of cluster analysis based on the Butina approach. This approach has been applied to QSAR (Quantitative Structure Activity relationship) investigations of food allergens, biologically active chemicals, as well as to QSPR (Quantitative Structure/Properties Relationship) studies of chromatographic and diesel fuel data. The results obtained are discussed in this work.

ivan.bangov@gmail.com