[P33] Using a new 3D pharmacophore presentation for ligand-based modeling

A. Kutlushina[1], T. Madzhidov[1], P. Polishchuk[2]

[1] Laboratory of Chemoinformatics, A.M. Butlerov Insitute of Chemistry, Kazan Federal University, Kremlevskaya Str. 10, Kazan, Russia.
[2] Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University and University Hospital in Olomouc, Hněvotinska, 1333/5, Olomouc, Czech Republic

Pharmacophore modeling is one of the major drug design approaches in the absence of structural data about the target receptor. There are many tools for 3D pharmacophore modeling (Topomer Search, Unity, LigandScout, Surflex-Sim, MOE, Galahad, GASP, DiscoTech and others), and they are all commercial [1].

We propose a new approach to 3D pharmacophore representation. In the framework of this approach a 3D pharmacophore is represented as a hash encoded its topology and stereoconfiguration. A pharmacophore is considered a complete graph with labeled vertices (H-acceptor, H-donor, center of positive or negative charge, hydrophobic or aromatic) and binned distances (with predefined step) as edges to enable fuzzy matching. The pharmacophore is split on all possible feature quadruplets. A canonical graph signature and stereoconfiguration are assigned to each quadruplet as its name. The number of identical quadruples is calculated and quadruplet names and counts are sorted in alphanumerical order followed by computation of md5 hash of this data structure. Distinct pharmacophores have different hashes and this property can be used for ligand-based pharmacophore modeling.

Ligand-based pharmacophore modeling is iterative and considers active and inactive compounds. On the first step counts of identical four-point pharmacophores are calculated for active and inactive compounds. Pharmacophores which match mainly active ones are selected for the next iteration. More complex five-point pharmacophores are generated based on the selected ones. The procedure is continued until the most complex pharmacophores matching mostly active compounds will be found.

Since we use active and inactive compounds for model generation two strategies of a training set selection were implemented. The first one supposes that all active compounds have identical or very similar binding modes, the second one uses clustering of compounds with 2D pharmacophore fingerprints to separate compounds with possibly different binding modes and models are generated for each cluster independently.

The proposed procedure was applied to develop pharmacophore models of acetylcholinesterase inhibitors. The best models had high precision and reasonable recall values on an external validation test set demonstrating applicability of the developed approach.

Acknowledgement:
This research was funded by the Ministry of education, youth and sport of Czech Republic, agreement number MSMT-5727/2018-2 and by the Ministry of Education and Science of the Russian Federation, agreement No14.587.21.0049 (unique identifier RFMEFI58718X0049).

Bibliography :
[1] Vuorinen A., Schuster D., Methods, 71, (2015), 113-134