

Flexophore, a new versatile 3D pharmacophore descriptor

M. Korff,¹

T. Sander², J. Freyss³

^{1,2,3} Actelion Ltd., Allschwil, Switzerland

A new molecular descriptor encoding three-dimensional pharmacophore information is described and evaluated. The encoding of a molecule starts by generating a reduced graph. Its nodes, which are called pharmacophore points, are classified by determining their enhanced atom types from a predefined list that was derived from an analysis of the protein data bank. This analysis also yielded a similarity matrix between these atom types characterizing the similarity of any two atom types concerning their coordination behavior to different protein atoms. The pharmacophore descriptor consists of the complete graph of these pharmacophore points. Any of the graph's edges is represented by a histogram of the distances through space between the corresponding nodes considering a representative set of conformers. These conformers are generated by a self-organization based algorithm for conformation sampling.

The pharmacophore similarity of two molecules is then determined by a sub-graph matching procedure considering node and edge similarities. Node similarities are taken from the enhanced atom type similarity matrix described above and edge similarities are calculated from the overlapping areas of the distance histograms.

To evaluate our descriptor's capability to model similarities of protein binding affinities we compiled a data set from the free available DUD dataset.¹ Because the DUD dataset contains proteins and ligands as well as decoys it is not only possible to compare ligand based and structure based screening, this dataset also enabled us to compare our results with other groups using the same dataset. The DUD database was designed to evaluate docking programs and contains 2,950 active compounds against 40 target proteins. Additionally the database contains 36 decoys for each ligand with similar physicochemical properties. We extracted the ligands from the target proteins to extend the applicability of the dataset to ligand based virtual screening. From the 40 target proteins 37 contained ligands which we used as query molecules for virtual screening evaluation. The query molecules were used to screen the test datasets consisting of ligands and decoys. In a large virtual screening experiment the Flexophore descriptor was challenged with five other descriptors and with our in-house docking tool. Four descriptors were chemical fingerprints, all encoding the molecular structure in a different way; the fifth descriptor was a topological pharmacophore histogram. Our experiments showed that the Flexophore descriptor outperformed the chemical descriptors as well as the topological pharmacophore descriptors considering the ability to detect structurally different actives while still being competitive concerning enrichment rates. Thus, it is well suited to find new chemical entities via "scaffold hopping". The Flexophore descriptor can be explored with a Java applet at <http://www.cheminformatics.ch/flexophore>. Its usage is free of charge and doesn't need any registration.

1. Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking Sets for Molecular Docking. *J. Med. Chem.* **2006**, 49, 6789-6801.