[L6] Toward universal maps of chemical space

Alexandre Varnek

University of Strasbourg, 4 rue Blaise Pascal, 67000, Strasbourg, France

Explosive raise of chemical data causes a need in the development of new chemoinformatics tools able to analyse, to visualize and to models these data. Generative Topographic Mapping (GTM) is a promising method of Big Data handling. It allows one not only to visualize chemical structures as data points on 2-dimensional space but also to models a data probability distribution function. The latter can efficiently be used in QSAR modeling, in chemical databases comparison and help to extract privilege structural motifs corresponding to particular kind of biological activity. The ability of GTM to build regression and classification models opens a way to construct the “universal” maps complying with the neighbourhood principle for hundreds different activities. Generally, several maps are needed to describe the variety of activities recorded in databases. In this presentation, we present 8 universal maps describing >1.5 M compounds from the ChEMBL database. These maps correctly classify more than 600 biological activities, thus representing the polypharmacologically competent view of drug-like space. Extraction of privilege motifs for GPCR ligands as well as for protease and kinase inhibitors from these maps is discussed.

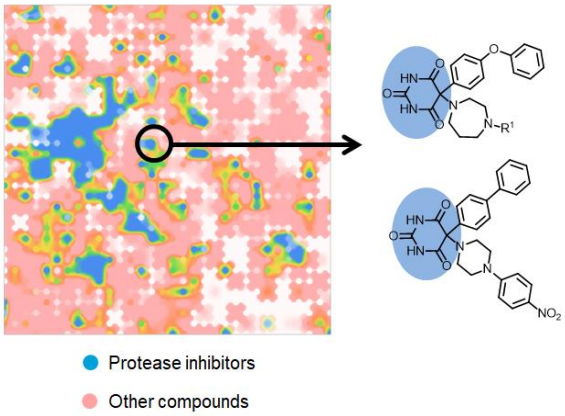

Figure 1. ChEMBL universal map displaying the zones populated by protease inhibitors (*left*) and privilege structural motifs extracted from the selected zone.

Bibliography:

1. P. Sidorov, H. A. Gaspar, Helena; A. Varnek, G. Marcou, D. Horvath J Comput Aided Mol Des. 2015, 29(12):1087-1108
2. K. Klimenko, G. Marcou, D. Horvath, A. Varnek J. Chem. Inf. Model., 2016, 56, 1438−1454
3. S. Kayastha, D. Horvath, E. Gilberg, M. Gütschow, J. Bajorath, A. Varnek J. Chem. Inf. Model., 2017, 56 (5), 1218-1232
4. A. Lin, D. Horvath, V. Afonina, G. Marcou, J.L. Reymond and A. Varnek, ChemMedChem., 2018; 13(6):540-554.