[L15] Consensus queries in ligand-based virtual screening experiments

Francois Berenger[1], Yoshihiro Yamanishi[1]

[1]System Cohort Division, Medical Institute of Bioregulation, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan.

In Ligand-Based Virtual Screening (LBVS) experiments, a known active ligand is used in similarity searches to find putative active compounds for the same protein target. When there are several known active molecules, screening using all of them is more powerful than screening using a single ligand. A consensus query can be created by either screening serially with different ligands before merging the obtained similarity scores, or by combining the molecular descriptors (i.e. chemical fingerprints) of those ligands.

We report [1] on the discriminative power and speed of several consensus methods, on two datasets only made of experimentally verified molecules. Our consent software is released in open source [2]. The two datasets contain a total of 19 protein targets, 3776 known active and ~2*10^6 inactive molecules. Three chemical fingerprints are investigated: MACCS 166 bits, ECFP4 2048 bits and an unfolded version of MOLPRINT 2D. Four different consensus policies and five consensus sizes were benchmarked.

The best consensus method is to rank candidate molecules using the maximum score obtained by each candidate molecule versus all known actives. When the number of actives used is small, the same screening performance can be approached by a consensus fingerprint. However, if the computational exploration of the chemical space is limited by speed (i.e. throughput), a consensus fingerprint allows to outperform this consensus of scores.

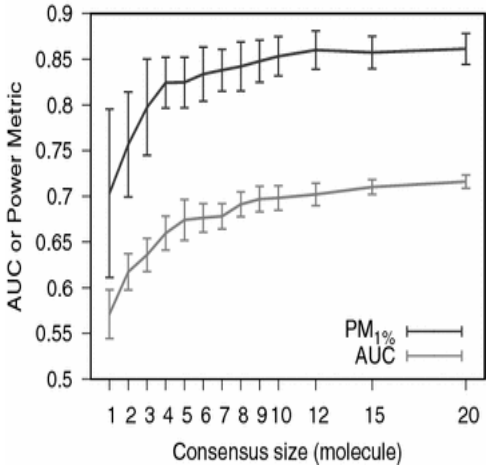

Figure 1: effect of the consensus size on the consensus query global classification performance (AUC) and early recovery capability (PM1%) [3]. Experiment: HTS dataset PubChem SAID 463087, ECFP4 fingerprint and optimist consensus (logical union of fingerprints). Values shown are medians over 100 experiments ± 1 median absolute deviation.

Bibliography:

[1] Berenger, Francois, Oanh Vu, and Jens Meiler. "Consensus queries in ligand-based virtual screening experiments." Journal of Cheminformatics 9.1 (2017): 60.
[2] https://github.com/UnixJunkie/consent
[3] Lopes, Julio Cesar Dias, et al. "The power metric: a new statistically robust enrichment-type metric for virtual screening applications with early recovery capability." Journal of cheminformatics 9.1 (2017): 7.