Capturing SAR-trends from chemogenomical spaces

Bernd Wendt
Tripos International, Martin-Kollar-Str. 17, Munich, Germany

Recent years have seen huge efforts in creating databases that store chemical structures with their biological activities for the public domain. With its new release Chembl [1] stores more than 520,000 individual compound records with 2.4 million records of their effects on biological systems making it one of the richest chemogenomical spaces currently available. In order to be useful for rational design of potent and selective lead structures it is necessary to extract structure-activity-relationships from this wealth of information. Mining databases the size of Chembl requires fast, robust and effective workflows.

We have developed a procedure around a new method: quantitative series enrichment analysis (QSEA)[2] that enables fully automated 3D-QSAR model creation and prediction. SAR-tables, lists of structures with associated activities that build input for the procedure, are constructed from shape similarity searches[3] using a collection of 255 marketed drugs[4] as queries. The visual output of the procedure reveals crucial information about the applicability domain of each series within a SAR-table together with insights where and how structural changes in the series might affect the biological activity. Results from mining of Chembl, PubChem and ChemBank databases will be reported.

References:
[1] http://www.ebi.ac.uk/chembl/db/