Using Protein Ligand Interaction Fingerprints and Machine Learning tools for the prediction of novel dual active compounds

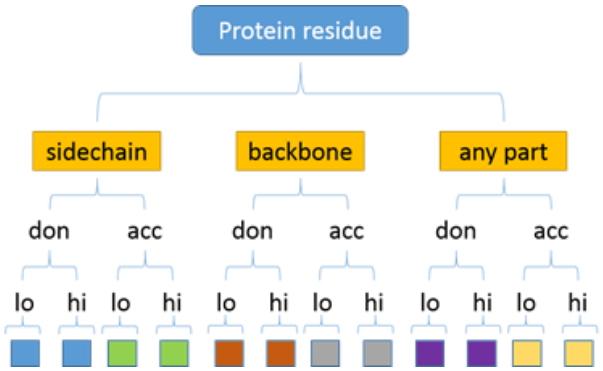
Lena Hefke[1]*, Ewgenij Proschak[1]

[1]*Goethe University Frankfurt, Frankfurt am Main, Hessen, 60438, Germany*

**kalinowsky@pharmchem.uni-frankfurt.de*

Structure-based drug design relies on accurate affinity prediction through analysis of protein-ligand interactions, which remains an only partially solved problem up to date. Furthermore, if affinities towards dual active compounds have to be approximated, the uncertainty of the prediction rises.

Here we introduce a combination of Protein Ligand Interaction Fingerprints (PLIF) and Machine Learning tools for the target-specific prediction of novel dual active compounds. The PLIF tool (Chemical Computing Group) uses a fingerprint representation of the interactions between ligands and proteins. Currently 10 types of interactions (hydrogen bonds, ionic, surface-, metal binding- and π interactions) are used to describe these interactions.[1]-[3] Using this fingerprints as a descriptor for specific protein-ligand interactions, different machine learning methods including support vector machines, self-organizing maps, neuronal networks, and random forest, were applied to classify active and inactive compounds.



Using the described workflow, we created a data set containing active co-crystallized ligands of the leukotriene A-4 hydrolase (LTA4H) and soluble epoxide hydrolase (sEH). After generating the PLIF, a random forest was trained to classify the compounds. Potential hits are manually inspected, selected compounds will be synthesized and characterized afterwards in an biochemical assay.

[1] P. Labute, *Journal of the Chemical Computing Group*, **2001**.
[2] A.M. Clark, P. Labute, M. Santavy, *J. Chem. Inf. Model.*, **2006**, *46*, 1107–1123.
[3] A.M. Clark, P. Labute, *J. Chem. Inf. Model.*, **2007**, *47*, 1933–1944.