Chemogenomics analysis of small molecule bioactivity data: Privileged scaffolds and conserved structural elements in proteins

L. Humbeck, O. Koch

Department of Chemistry and Chemical Biology, TU Dortmund, Dortmund, Germany

The term “privileged scaffolds” is often used for multiple molecules that show bioactivity on different targets but consist of the same scaffold. [1] Within proteins, conserved structural elements can also be found in different proteins, ranging from conserved motifs that interact with specific functional groups to similar spatial arrangements of secondary structure elements around the ligand binding site (the “ligand sensing core”) in proteins with different folding patterns that can bind similar scaffolds. [2] Information about similar ligand sensing cores can be useful for rational identification of new lead structures [3] or predicting polypharmacology. [2]

Chemical compound databases like DrugBank (http://www.drugbank.ca/) or ChEMBL (https://www.ebi.ac.uk/chembl/) contain a huge amount of data about molecules and their bioactivity on different protein targets. Therefore we decided to develop a python based tool for knowledge discovery to get new insights about the relationship of privileged scaffolds and conserved structural elements in proteins. The main idea of this data mining approach is the identification of scaffolds that bind to different and unrelated protein targets for analyzing potential conserved structural elements.

In a first step, a command line version of Scaffold Hunter [4] is used to reduce all molecules in a database to their containing scaffolds. The second step analyses the sequence similarity of protein targets of all molecules sharing a common scaffold. Only protein targets with identity below 40 % are regarded as unrelated. The last step visualizes the results for an in-depth analysis of the results.

We will present the overall workflow and the result of an exhaustive chemogenomics analysis of the DrugBank. Around 1500 scaffolds were identified that are active against different protein targets. An analysis of one example already ended up in a new ligand sensing core that is shared between four different protein targets and can help to identify new lead structures for the respective targets.

[1] M. E. Welsch, S. A.Snyder, B. R. Stockwell, *Curr Opin Chem Biol*, **2010**, *14(3)*, 347-361.
[2] O. Koch, *Fut Med Chem*, **2011**, *3(6)*, 699-708.
[3] D. Willmann, S. Lim, S. Wetzel, E. Metzger, A. Jandausch, W. Wilk, M. Jung, I. Forne, A. Imhof, A. Janzer, J. Kirfel, H. Waldmann, R. Schüle, R. Buettner, *Int J Cancer*, **2012**, *131(11)*, 2704-9.
[4] K. Klein, O. Koch, N. Kriege, P. Mutzel, T. Schäfer, *Mol Inf*, **2013**, *32*, 964-975.