[P50] Development of an universal workflow for the preparation of molecular databases for Virtual Screening

José-Manuel Gally, Nicolas Bosc, Emilie Pihan, Rohit Arora, Stéphane Bourg, Baptiste Canault, Pascal Bonnet

Structural Bioinformatics and Chemoinformatics, Université d'Orléans, CNRS, ICOA, UMR 7311, 45067 Orléans, France

The development of new bioactive molecules is very challenging since it is constrained by multiple factors such as time and cost of drug development as well as intellectual property, which make the field of drug discovery extremely competitive. Therefore, novel tools have been developed in early discovery stage to identify and prioritize IP free scaffolds with suitable physicochemical properties. Virtual Screening is a well integrated tool in drug discovery processes to identify potential bioactive molecules from docking, to generate novel molecules from chemical library enumeration or to select molecules from databases of providers which can then be optimized using drug-like properties. The success of this technique relies, not only on the main *in silico* tools but also on how the ligand database has been prepared. A general protocol was therefore developed using Knime platform to ensure that the input ligands were adequate for docking by combining the benefits of several toolkits, such as RDKIT, Indigo and Chemaxon. First, the systematic false positives and reactive compounds are filtered by using the definition of Baell et all [1, 2] and invalid compounds (wrong valence, isotopes) are discarded. For mixtures, the largest molecule for each entry is retained. Since molecular databases can be provided by many external sources, similar molecules might have different tautomeric forms which would cause extra work load. To address this issue, all molecules are transformed in a unique tautomeric form and duplicate entries were removed. As a compromise between exhaustiveness and computational time, only rational tautomers are enumerated and the molecules are protonated at physiological pH. Furthermore, for unspecified stereochemistry, the stereoisomers are fully enumerated depending on the number of chiral centers in each molecule. Finally, for 1D or 2D entries, three dimensional coordinates are generated for the molecules and optionally conformational states could be enumerated [3]. Due to the nature of the packages used in this workflow, it will be freely available for academics. We welcome suggestions from the chemoinformatics community to improve this tool and to make it broadly available for virtual screening projects

[1] Baell J.B., Holloway, G.A. J. Med. Chem. 53, 2719-2740 (2010).
[2] http://www.chemaxon.com/marvin/help/calculations/tautomers.html
[3] Ebejer, J.-P., Morris, G.M., Deane, C.M. Journal of Chemical Information and Modeling 52, 1146-1158 (2012).

.