ChemicalToolBoX in Action!

Grüning, B.A.[1]; Lucas, X.[1]; Patel, H.[2]; Flemming, S.[1]; Chbeib, M.[1]; Günther, S.[1]

bjoern.gruening@pharmazie.uni-freiburg.de

[1] *Pharmaceutical Bioinformatics, Institute of Pharmaceutical Sciences, University of Freiburg*
[2] *Pharmaceutical Biology and Biotechnology, Institute of Pharmaceutical Sciences, University of Freiburg*

ChemicalToolBox is a set of tools integrated into the Galaxy[1] workflow-management system to enable researchers easy-to-use, reproducible, and transparent access to cheminformatics library and drug discovery tools. It includes standard applications for similarity and substructure searches, clustering of compounds, prediction of properties and descriptors, filtering, and many other tools that range from druglikeness classification to fragmentation and fragment-merging.

ChemicalToolBox is based on open-source software, web-accessible, freely available, and easily expandable. It can be downloaded and easily deployed locally or on a large scale cluster. Here we present our chemical toolbox and some use-cases that we have developed in our lab to demonstrate its capabilities. By combinations of the various tools many more powerful applications can be designed.

Collect data

- Upload custom data
- Download data
- Paint structures
- Fragmentation of libraries

Manipulate data

- Format conversion
- Extract scaffold
- Generate 2D/3D coordinates
- Conformer generation
- Remove small molecules/ions
- Remove duplicates

Filter libraries

- Physico-chemical properties
- Predefined rules (RoF, Lead-like, ...)
- Estimation of drug likeness (QED)[2]

Search libraries

- Similarity Search
- Pharmacophore Search
- Spectrophore Search
- Substructure Search

Clustering

- Fingerprint
- All-against-all clustering
- Taylor-Butina clustering

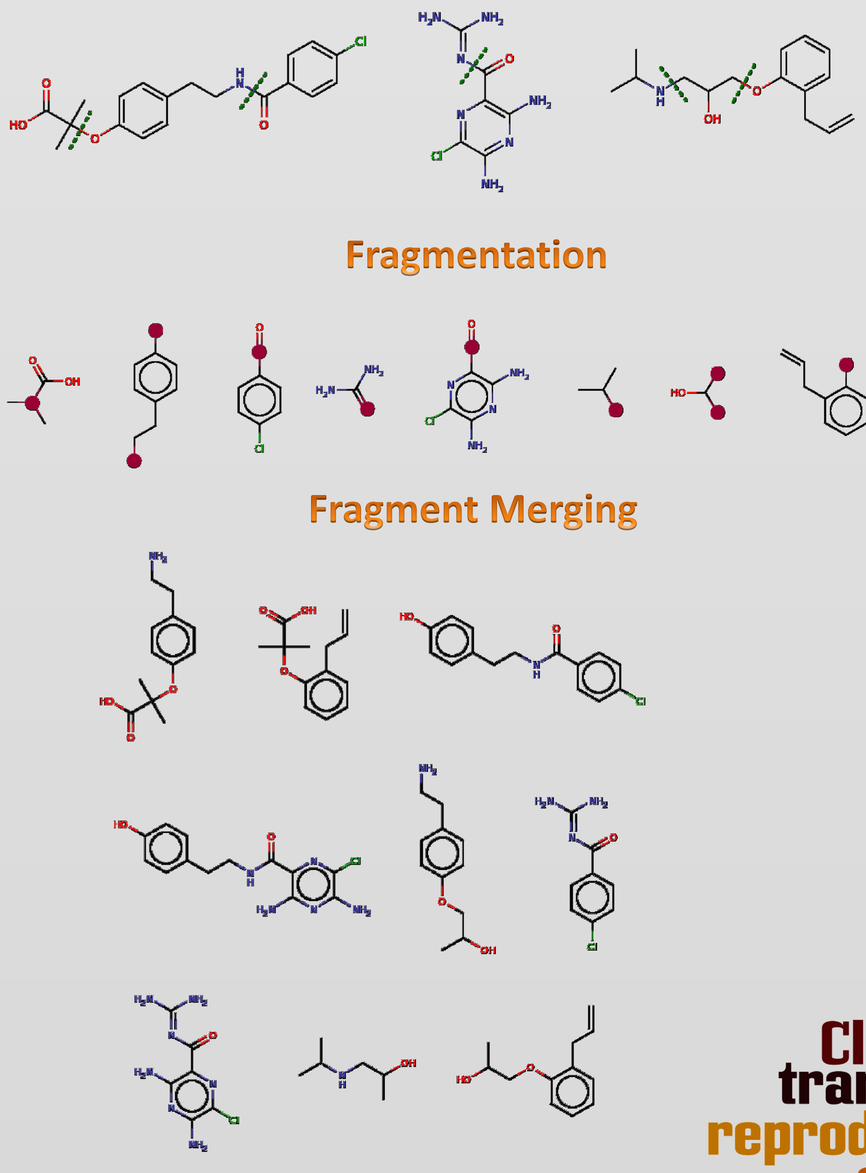
Visualisation

- Plotting of properties
- SVG representation
- 3D visualisation of whole libraries (work in progress)

ChemicalBox: a comprehensive library of about **39 million** unique small compounds collected from freely accessible repositories. The compounds included in ChemicalBoX have been filtered by physico-chemical properties and clustered by similarity to allow fast navigation throughout millions of compounds belonging to many different chemical classes.
ftp://ftp.pharmaceutical-bioinformatics.org/chemicalbox/

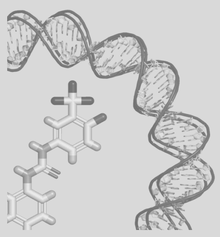
PurchasableBoX: a library consisting of **44.6 million** filtered, unique purchasable small compounds, of which **43.6 million** are drug-beauty (QED >= 0.2). Clustering with a similarity of 85 % revealed **1.2 million** cluster centers.

Both libraries are unique and, to the best of our knowledge, the **largest available libraries**. Furthermore, merging both libraries together will create **~70 million** unique, filtered compounds. That is more than the amount of compounds registered in CAS.



This illustration shows a workflow that is implemented in the ChemicalToolBoX. It fragmentizes an input library according to predefined rules (e.g. RECAP[3]) and labels the broken bonds (red dots). In the next step it merges the fragments according to predefined synthetic rules to form new promising compounds that enlarge the known chemical space.
The fragmentation of 1,425 approved drugs yields 1,895 unique fragments that build the foundation of **150,642 unique drug-like compounds**.

The majority is totally new and not included in **ChemicalBoX** and **PurchasableBoX**.



References

[1] Goecks, J, et al. “Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences.” Genome Biol. 2010
[2] G. Richard Bickerton et al. "Quantifying the chemical beauty of drugs"; Nature Chemistry 2012
[3] Lewell XQ et al. “RECAP--retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry.” J Chem Inf Comput Sci. 2012