ChemicalToolBoX
a Galaxy for Cheminformatics

Björn Grüning
Pharmaceutical Bioinformatics
bjoern.gruening@pharmazie.uni-freiburg.de
Paint your structure and add it to Galaxy!

Editor:

Export Molfile to Galaxy  Export Smiles to Galaxy

Mol. Weight: Formula:
Integrated Tools

- filters
- physico-chemical properties
- Drug-likeliness prediction
- Conformer
- 2D/3D coordinate
- converters + filetypes
- similarity and substructure search
- Clustering
- spectrophore and pharmacophore search
- Remove duplicates
- molecule modification
- remove ions

- plotting of compounds
- Compound-DB integration
- fragmentation of compounds
ChemicalBoX: a workflow inside the ChemicalToolBoX

1. Collect data from sources (mixed formats)
2. Merge input databases (SMI format)
3. Select unique molecules (~40M)
4. Transform isotopes into main species
5. Remove molecules containing $M^{n+}$ or weird elements
6. Remove molecules with no C atoms
7. Remove small molecules (< 6 atoms)
8. Remove counterions and fragments
9. Select unique molecules
10. Select molecules with 100<MW<850 (~38.2M)
11. Create fingerprints (FP2)
12. Cluster @ 90% similarity (~1.3M)
13. Cluster @ 60% similarity (~124K)
14. Cluster @ 80% similarity (~600K)
ChemicalBoX

- PubChem ~ 32.5M*
- ZINC ~ 12M
- ChEMBL ~ 1.1M
- BindingDB ~ 300K*
- NCI ~ 260K
- ChEBI ~ 22K
- PDB ligand ~ 14K
- DrugBank ~ 6K*
Current features of the BoXes

**ChemicalBoX**
- Largest freely available chemical library
- Workflow FULLY automatized using tools from the ChemicalToolBoX
- Weird compounds filtered out
- Monthly updated
- Pre-clustered sets at 90, 80 and 60% similarity thresholds

**PurchasableBoX**
- ~ 40 companies are being fetched
- Largest filtered purchasable space known (~37M)
ftp://pharmaceutical-bioinformatics.org/chemicaltoolbox/

ftp://pharmaceutical-bioinformatics.org/chemicalbox/

$ git clone http://chemicalbox.pharmaceutical-bioinformatics.org/chemicalbox.git
Thanks

Pharmaceutical Bioinformatics
Institute of Pharmaceutical Sciences
Albert-Ludwigs-University Freiburg