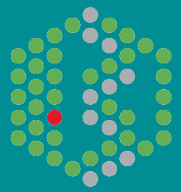
Clippy: A cross-platform wxPython app

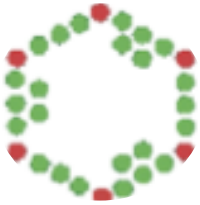
George Papadatos, PhD
Senior Technical Officer
ChEMBL group
georgep@ebi.ac.uk

EMBL-EBI



Outline

- myChEMBL
- Clippy



myChEMBL

myChEMBL

- Ubuntu 13.04 VM with:
  - Fresh build of RDKit from GitHub
  - ChEMBL 17 db in postgresql
  - RDKit cartridge
  - Custom ChEMBL web app and web services
  - ipython (qtconsole & notebook), numpy, scipy, pandas, etc.
  - Connectable via http, ssh, postgres client, KNIME, etc.
  - Available for free at:
    - ftp://ftp.ebi.ac.uk/pub/databases/chembl/VM/myChEMBL/current
    - Portable hard disk

Manuscript submitted

myChEMBL installation

- Copy/download folder to your disk
- VirtualBox → import appliance…
- Choose .ovf file
- Next → Import
- Wait for 5-10 minutes
- Username: `chembl`
- Password: `chemblvm`
- `sudo pip install scikit-learn`
- `ipython notebook --ip='*'`

myChEMBL: Interface

EMBL-EBI

RDKit | Ora2Pg | ChEMBL-og

myChEMBL

Home | Structure Search | Web Services | Property Calculation | Tutorial | Acknowledgements

Welcome to myChEMBL

The aim of the myChEMBL Project is to make the process installing and running a local copy of the ChEMBL database as easy as possible. Additional goals of the myChEMBL Project included:

- The database we provide should be 'Chemically Aware', allowing users to run chemical substructure and similarity searches.
- The removal of any cost restrictions imposed to software licensing, so only freely available open-source software should be used in the project.

In order to achieve each of these goals we have created a Virtual Machine, which contains the following:

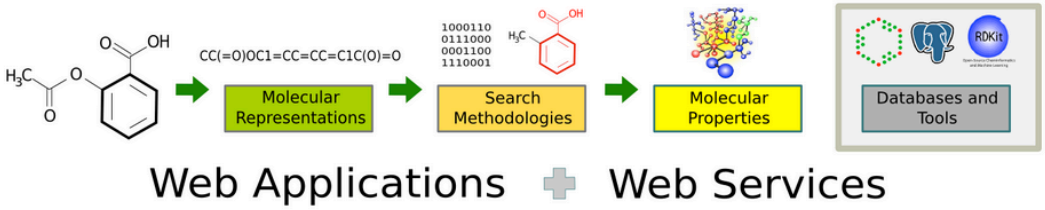
- PostgreSQL version of the ChEMBL database (Release: ChEMBL_17).
- The RDKit Chemical Cartridge, which is an open source software library and adds 'Chemical-Awareness' to a PostgreSQL database.
- A web application (this web application), which demonstrates some of the functionality of the RDKit Chemical Cartridge, such as chemical structure searching and chemical structure property calculations.*

*Note you do not need to use the web application and can connect directly to the database using PostgreSQL client.

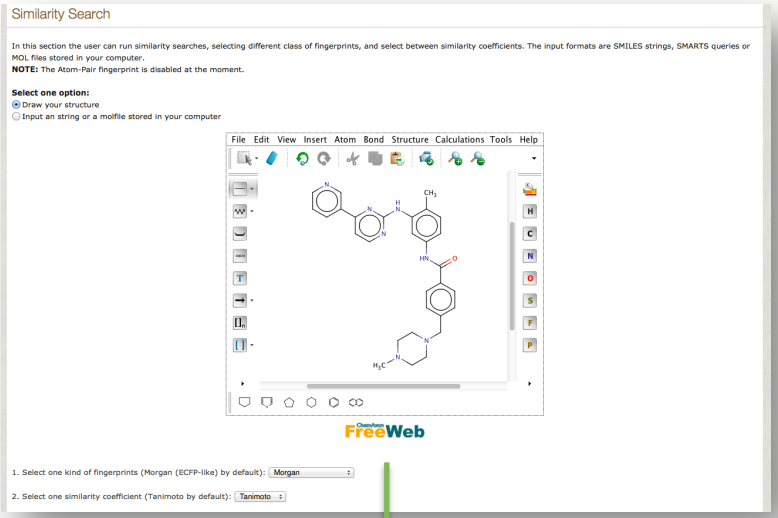
Your feedback will help us determine how successful this project has been and also help drive the projects future direction. So please get in get in touch and share with us your myChEMBL comments and questions.

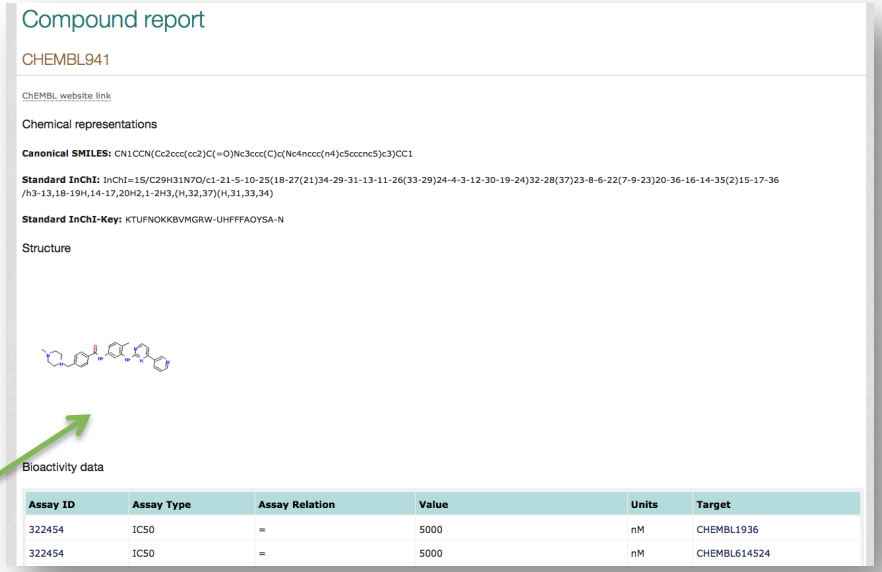
myChEMBL Web Application Overview

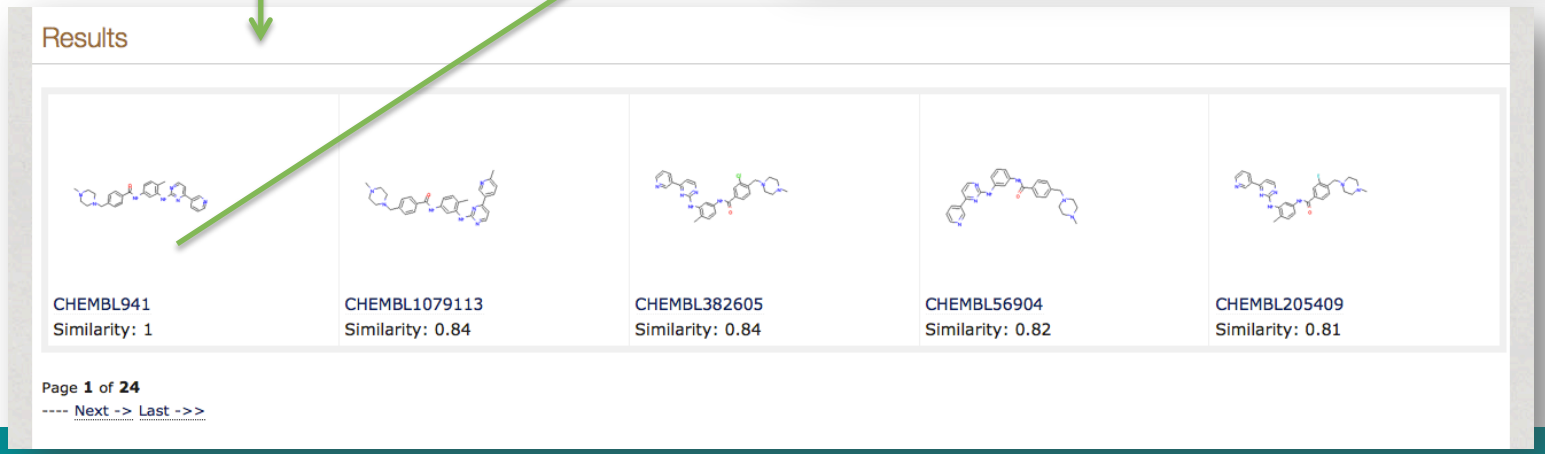
The myChEMBL web application is used to expose some of the functionality provided by the RDKit Chemical Cartridge. It allows a user to draw or upload a chemical structures, which can then be used in chemical structure searches against the ChEMBL database or in chemical property calculations. Programmatic access to each of these services is also provided through a RESTful Web Service



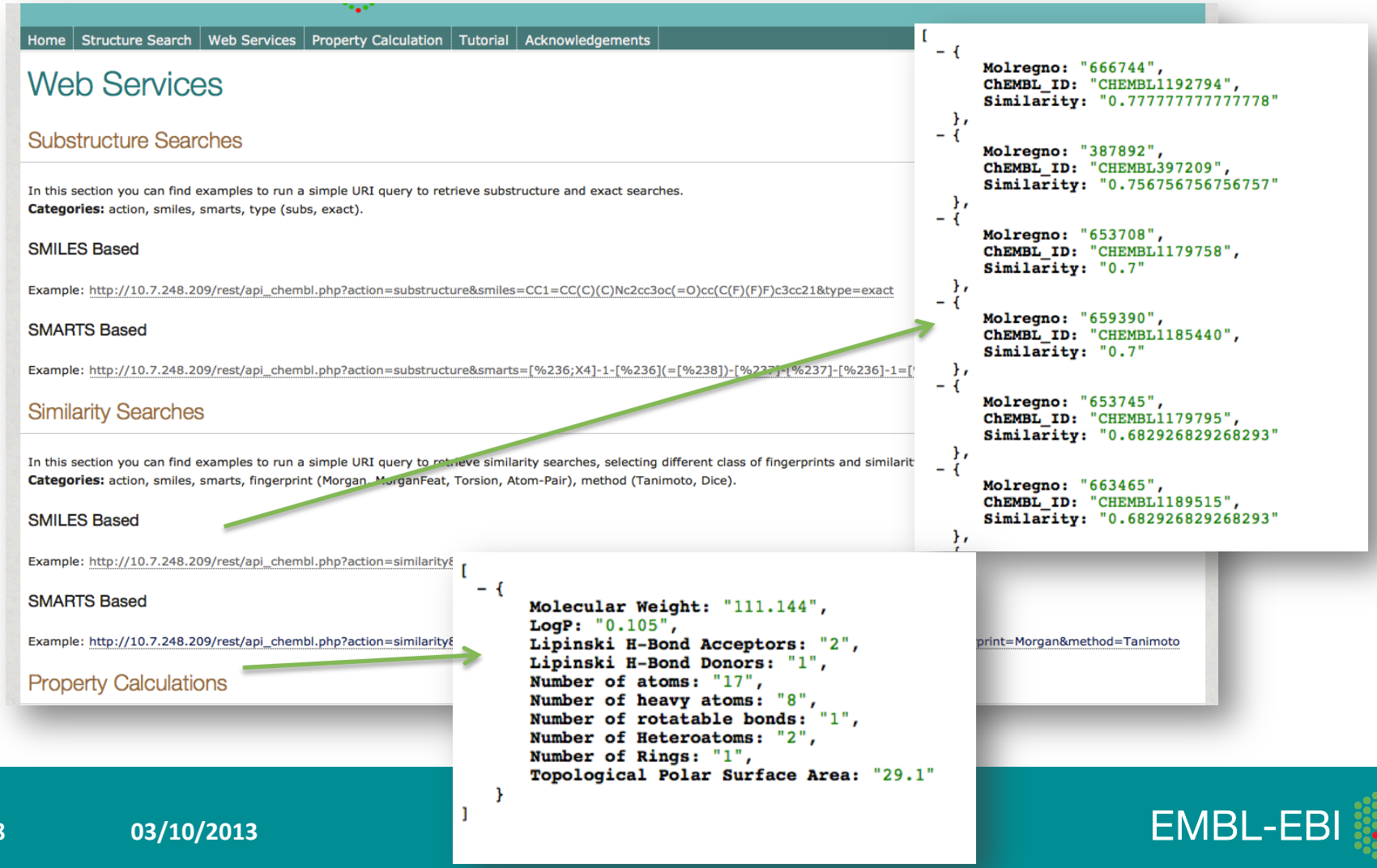
myChEMBL: Interface



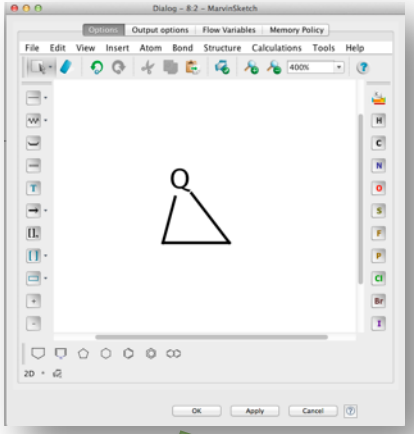




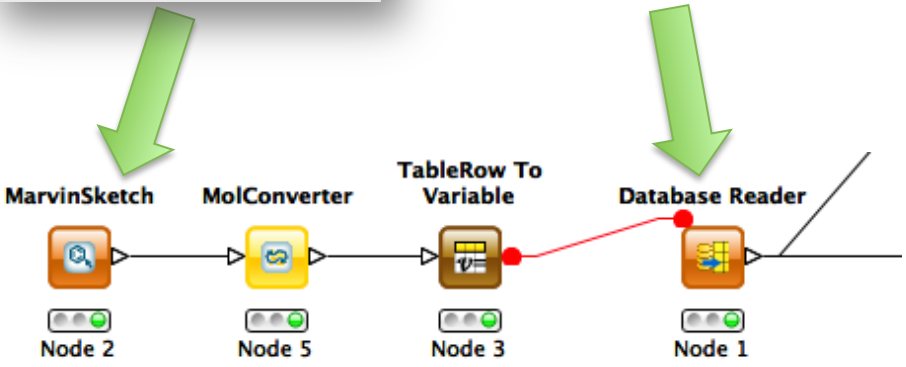
myChEMBL: Web Services

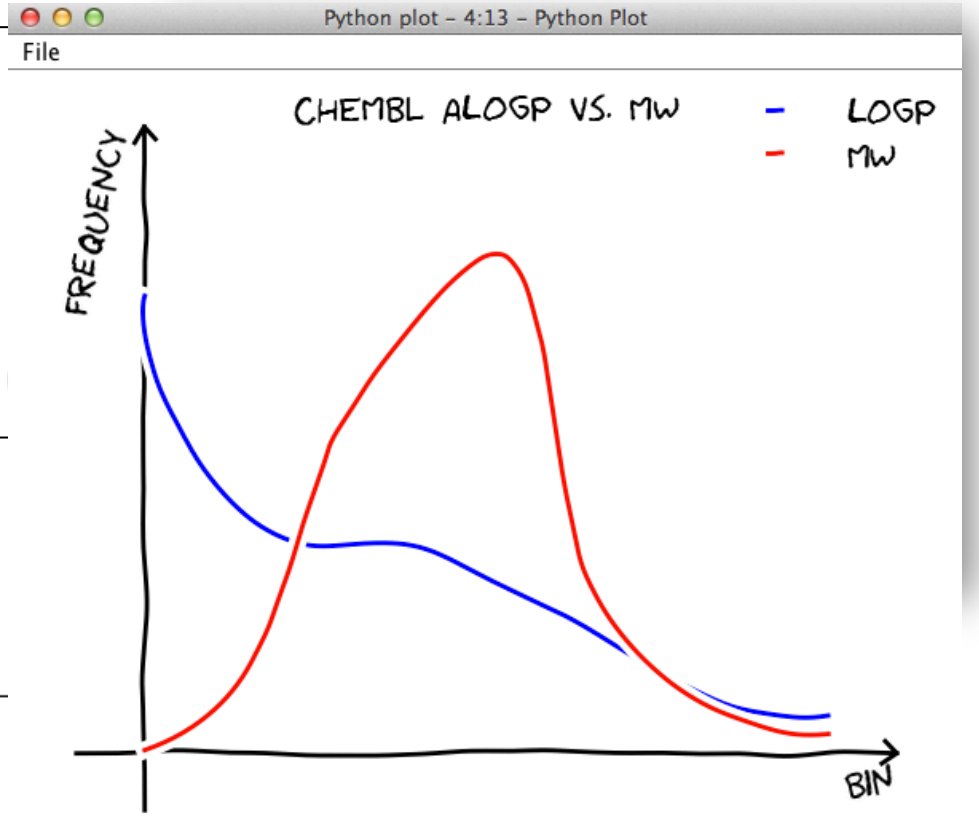


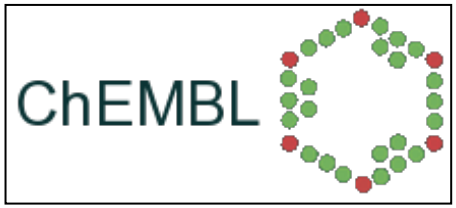
Using KNIME to connect to the VM



```
SELECT mr.*, md.chembl_id,
cp.full_mwt, cp.alogp
from mols_rdkit mr,
molecule_dictionary md,
compound_properties cp
where
mr.m @> '$${SMolecule}$$':
and
mr.molregno = md.molregno
and
md.molregno = cp.molregno;
```

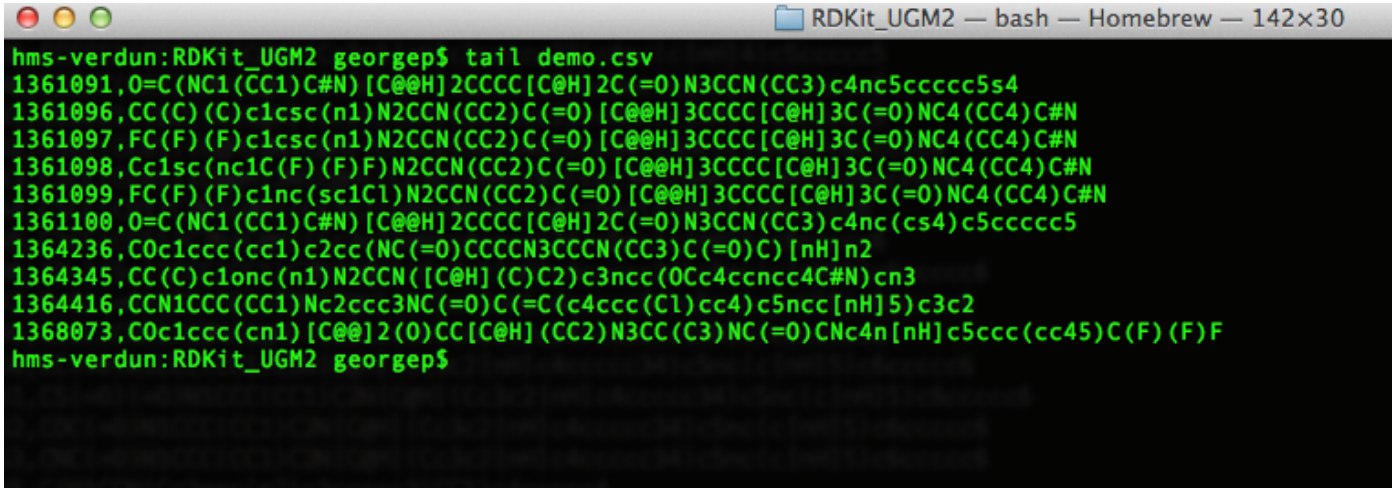






Clippy

Rationale 1


RDKit_UGM2 — bash — Homebrew — 142×30
hms-verdun:RDKit_UGM2 georgep$ tail demo.csv
1361091,O=C(NC1(CC1)C#N)[C@@H]2CCCC[C@H]2C(=O)N3CCN(CC3)c4nc5ccccc5s4
1361096,CC(C)(C)c1csc(n1)N2CCN(CC2)C(=O)[C@@H]3CCCC[C@H]3C(=O)NC4(CC4)C#N
1361097,FC(F)(F)c1csc(n1)N2CCN(CC2)C(=O)[C@@H]3CCCC[C@H]3C(=O)NC4(CC4)C#N
1361098,Cc1sc(nc1C(F)(F)F)N2CCN(CC2)C(=O)[C@@H]3CCCC[C@H]3C(=O)NC4(CC4)C#N
1361099,FC(F)(F)c1nc(sc1Cl)N2CCN(CC2)C(=O)[C@@H]3CCCC[C@H]3C(=O)NC4(CC4)C#N
1361100,O=C(NC1(CC1)C#N)[C@@H]2CCCC[C@H]2C(=O)N3CCN(CC3)c4nc(cs4)c5ccccc5
1364236,COc1ccc(cc1)c2cc(NC(=O)CCCCN3CCCN(CC3)C(=O)C)[nH]n2
1364345,CC(C)c1onc(n1)N2CCN([C@H](C)C2)c3ncc(OCc4ccncc4C#N)cn3
1364416,CCN1CCC(CC1)Nc2ccc3NC(=O)C(=C(c4ccc(Cl)cc4)c5ncc[nH]5)c3c2
1368073,COc1ccc(cn1)[C@@]2(O)CC[C@H](CC2)N3CC(C3)NC(=O)CNc4n[nH]c5ccc(cc45)C(F)(F)F
hms-verdun:RDKit_UGM2 georgep$

Rationale 2

new compounds — Trash

2 Attachments, 66 KB Save Quick Look

From: "The Chemist" <thechemist@ebi.ac.uk>
Subject: Re: New actives
Date: 2 October 2013 12:55:34 BST
To: George Papadatos <georgep@ebi.ac.uk>

Hi George,

These are the latest screening hits:

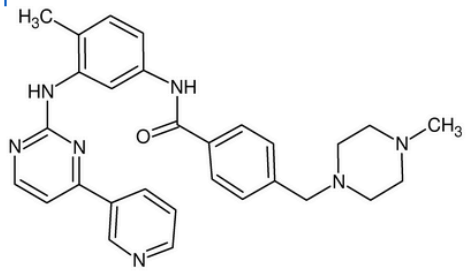
chembl123
chembl111

and these
CCCCC1(CCCC)N[C@H](Cc2c1[nH]c3ccccc23)c4nc(c[nH]4)c5ccccc5
C1CCC(CC1)C2N[C@H](Cc3c2[nH]c4ccccc34)c5nc(c[nH]5)c6ccccc6

And I also found this analogue in a vendor catalogue:

InChI=1S/C14H18N4O3/c1-19-10-5-8(6-11(20-2)12(10)21-3)4-9-7-17-14(16)18-13(9)15/h5-7H,4H2,1-3H3,(H4,15,16,17,18)

and this one:



and

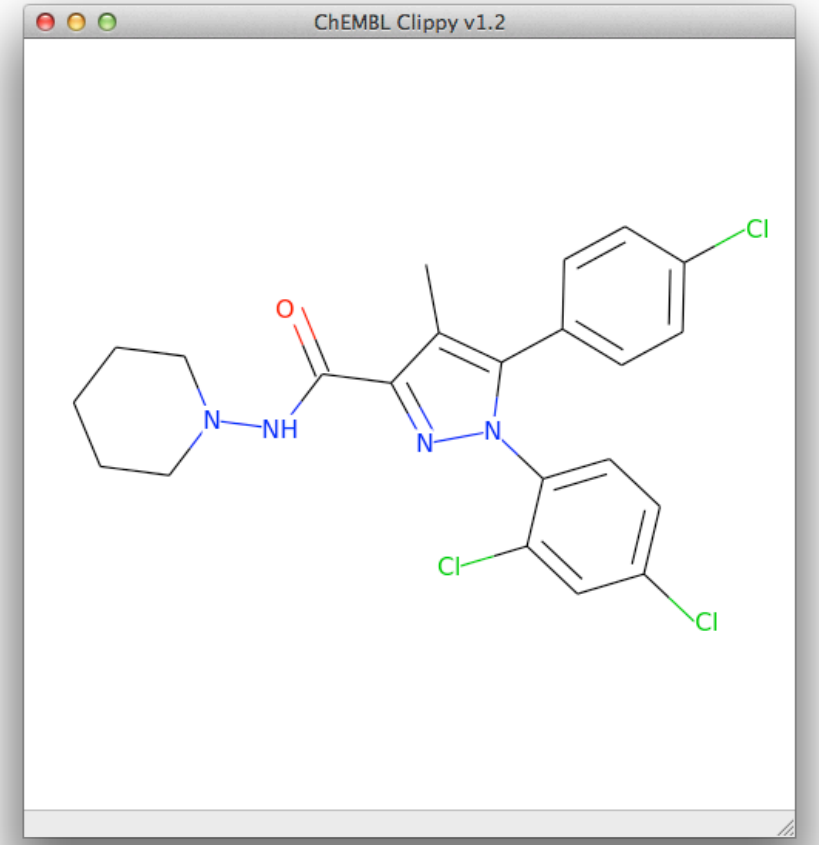
The answer: Clippy

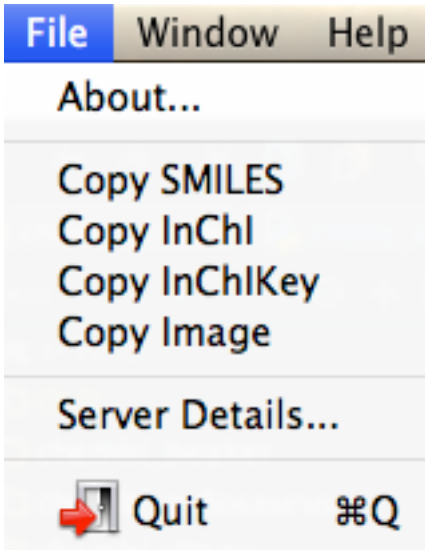
- A simple wxPython GUI app
- Cross-platform
  - Win, MacOS and Ubuntu
- Reads data from clipboard
  - ChEMBL IDs, SMILES, InChIs, images
- Visualises structure
- Exports structure in different formats
  - SMILES, InChI, InChIKey, image
- Uses Beaker web services
- Can be packaged and deployed anywhere

MarvinView vs. Clippy

- Requires more clicks
- Closed source
- Not customisable
- Java

Live Demo


ChEMBL Clippy v1.2
O
N
NH
N
N
Cl
Cl
Cl


File
Window
Help
About...
Copy SMILES
Copy InChI
Copy InChIKey
Copy Image
Server Details...
Quit
⌘Q

Technologies

clippy

wxPython

smiles2ctab, ctab2smiles

ctab2inchi, inchi2inchikey

image2ctab, ctab2image

chembl2smiles

beaker

RDKit

OSRA

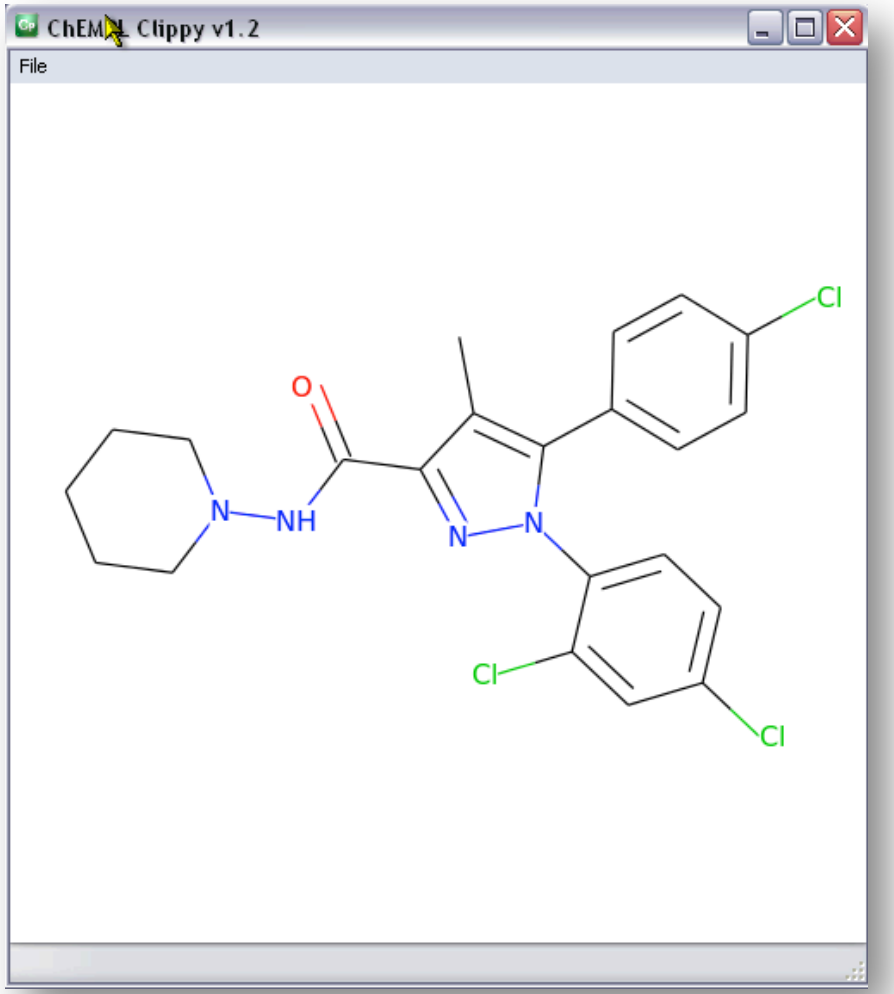
ChEMBL WS

Packaging


```
pyinstaller -wFs --clean clippy_all.py
```

Live Demo

- Cross-platform
  - Same code runs on MacOS, Windows and Ubuntu
- Beaker WS & packaging remove all dependencies
  - RDKit, Python, etc.
- Can be deployed to any desktop



Summary

- myChEMBL VM
- Clippy app

Acknowledgements

- ChEMBL group
  - John Overington
  - Anne Hersey
  - **Francis Atkinson**
  - Louisa Bellis
  - Jon Chambers
  - **Mark Davies**
  - **Michal Nowotka**
  - Nathan Dedman
  - Anna Gaulton
  - **Rodrigo Ochoa**
- RDKit community
  - Greg
- All of you for listening

wellcome trust

Clippy: A cross-platform wxPython app

George Papadatos, PhD
Senior Technical Officer
ChEMBL group
georgep@ebi.ac.uk

EMBL-EBI