

Clippy: A cross-platform wxPython app

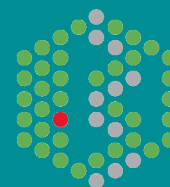
George Papadatos, PhD

Senior Technical Officer

ChEMBL group

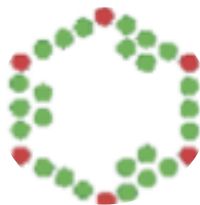
georgep@ebi.ac.uk

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](#).

CC(=O)OC1=CC=CC=C1C(=O)O

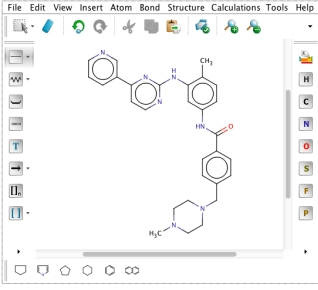
Web Applications + Web Services

myChEMBL: Interface

Similarity Search

In this section the user can run similarity searches, selecting different class of fingerprints, and select between similarity coefficients. The input formats are SMILES strings, SMARTS queries or MOL files stored in your computer.
NOTE: The Atom-Pair fingerprint is disabled at the moment.

Select one option:
☒ Draw your structure
☐ Input an string or a molfile stored in your computer



1. Select one kind of fingerprints (Morgan (ECFP-like) by default):

2. Select one similarity coefficient (Tanimoto by default):

Compound report

CHEMBL941

[ChEMBL website link](#)

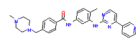
Chemical representations

Canonical SMILES: CN1CCN(Cc2ccc(cc2)C(=O)Nc3ccc(C)c(Nc4cccc(n4)c5ccccc5)c3)CC1

Standard InChI: InChI=1S/C29H31N7O/c1-21-5-10-25(18-27(21)34-29-31-13-11-26(33-29)24-4-3-12-30-19-24)32-28(37)23-8-6-22(7-9-23)20-36-14-35(2)15-17-36/n3-13,18-19H,14-17,20H2,1-2H3,(H,32,37)(H,31,33,34)

Standard InChI-Key: KTUFNOKKBVMGRW-UHFFFAOYSA-N

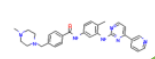
Structure



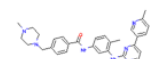
Bioactivity data

Assay ID	Assay Type	Assay Relation	Value	Units	Target
322454	IC50	=	5000	nM	CHEMBL1936
322454	IC50	=	5000	nM	CHEMBL614524

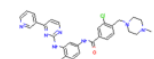
Results



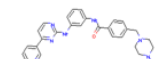
CHEMBL941
Similarity: 1



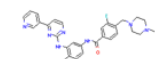
CHEMBL1079113
Similarity: 0.84



CHEMBL382605
Similarity: 0.84



CHEMBL56904
Similarity: 0.82



CHEMBL205409
Similarity: 0.81

Page 1 of 24
---- Next -> Last ->>



myChEMBL: Web Services

Home

Structure Search

Web Services

Property Calculation

Tutorial

Acknowledgements

Web Services

Substructure Searches

In this section you can find examples to run a simple URI query to retrieve substructure and exact searches.
Categories: action, smiles, smarts, type (subs, exact).

SMILES Based

Example: [http://10.7.248.209/rest/api_chembl.php?action=substructure&smiles=CC1=CC\(C\)\(C\)Nc2cc3oc\(=O\)cc\(C\(F\)\(F\)F\)c3cc21&type=exact](http://10.7.248.209/rest/api_chembl.php?action=substructure&smiles=CC1=CC(C)(C)Nc2cc3oc(=O)cc(C(F)(F)F)c3cc21&type=exact)

SMARTS Based

Example: [http://10.7.248.209/rest/api_chembl.php?action=substructure&smarts=\[%236;X4\]-1-\[%236\]\(=\[%238\]\)-\[%222\]-\[%237\]-\[%236\]-1=\[%236\]](http://10.7.248.209/rest/api_chembl.php?action=substructure&smarts=[%236;X4]-1-[%236](=[%238])-[%222]-[%237]-[%236]-1=[%236])

Similarity Searches

In this section you can find examples to run a simple URI query to retrieve similarity searches, selecting different class of fingerprints and similarity methods.
Categories: action, smiles, smarts, fingerprint (Morgan, MorganFeat, Torsion, Atom-Pair), method (Tanimoto, Dice).

SMILES Based

Example: [http://10.7.248.209/rest/api_chembl.php?action=similarity&smiles=CC1=CC\(C\)\(C\)Nc2cc3oc\(=O\)cc\(C\(F\)\(F\)F\)c3cc21&method=Tanimoto](http://10.7.248.209/rest/api_chembl.php?action=similarity&smiles=CC1=CC(C)(C)Nc2cc3oc(=O)cc(C(F)(F)F)c3cc21&method=Tanimoto)

SMARTS Based

Example: [http://10.7.248.209/rest/api_chembl.php?action=similarity&smarts=\[%236;X4\]-1-\[%236\]\(=\[%238\]\)-\[%222\]-\[%237\]-\[%236\]-1=\[%236\]&method=Tanimoto](http://10.7.248.209/rest/api_chembl.php?action=similarity&smarts=[%236;X4]-1-[%236](=[%238])-[%222]-[%237]-[%236]-1=[%236]&method=Tanimoto)

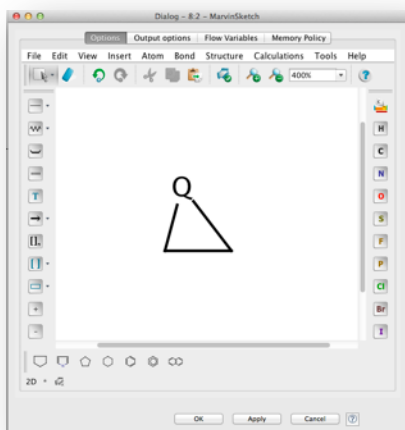
Property Calculations

```
[
- {
  Molregno: "666744",
  ChEMBL_ID: "CHEMBL1192794",
  Similarity: "0.7777777777777778"
},
- {
  Molregno: "387892",
  ChEMBL_ID: "CHEMBL397209",
  Similarity: "0.756756756756757"
},
- {
  Molregno: "653708",
  ChEMBL_ID: "CHEMBL1179758",
  Similarity: "0.7"
},
- {
  Molregno: "659390",
  ChEMBL_ID: "CHEMBL1185440",
  Similarity: "0.7"
},
- {
  Molregno: "653745",
  ChEMBL_ID: "CHEMBL1179795",
  Similarity: "0.682926829268293"
},
- {
  Molregno: "663465",
  ChEMBL_ID: "CHEMBL1189515",
  Similarity: "0.682926829268293"
},
]
```

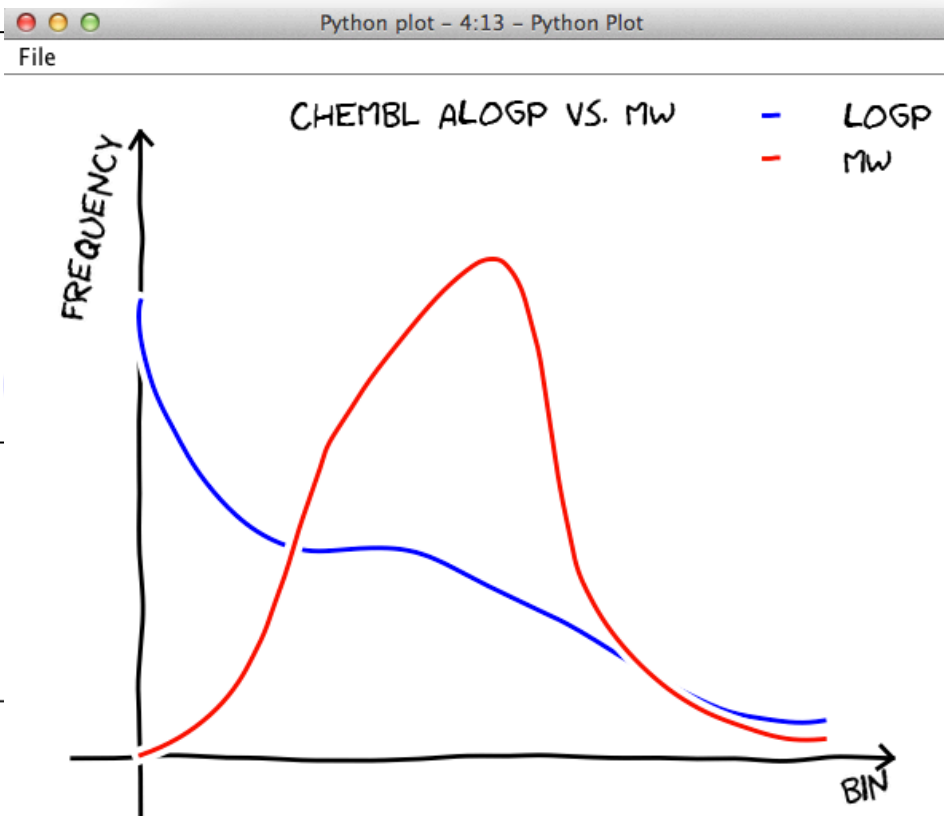
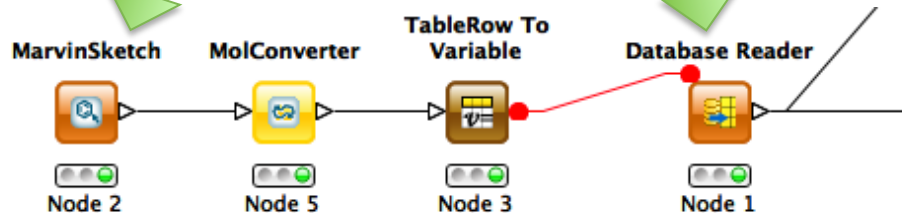
```
[
- {
  Molecular Weight: "111.144",
  LogP: "0.105",
  Lipinski H-Bond Acceptors: "2",
  Lipinski H-Bond Donors: "1",
  Number of atoms: "17",
  Number of heavy atoms: "8",
  Number of rotatable bonds: "1",
  Number of Heteroatoms: "2",
  Number of Rings: "1",
  Topological Polar Surface Area: "29.1"
},
]
```

```
print=Morgan&method=Tanimoto
```


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 — 142x30
hms-verdun:RDKit_UGM2 georgep$ tail demo.csv
1361091, O=C(NC1(CC1)C#N) [C@@H] 2CCCC[C@H] 2C(=O)N3CCN(CC3)c4nc5cccc5s4
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(sc1C1)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)c5cccc5
1364236, C0c1ccc(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(C1)cc4)c5ncc[nH]5) c3c2
1368073, C0c1ccc(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

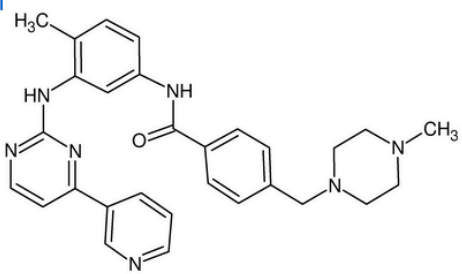
and these

```
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

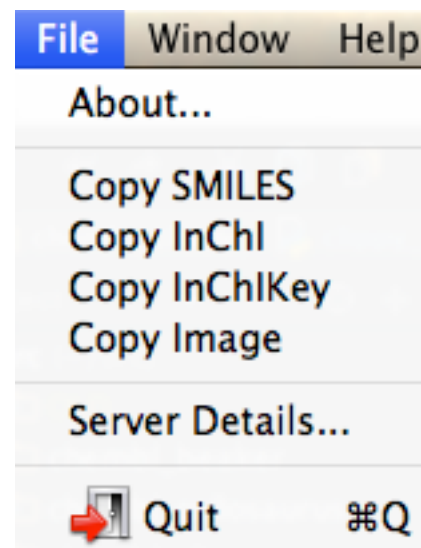
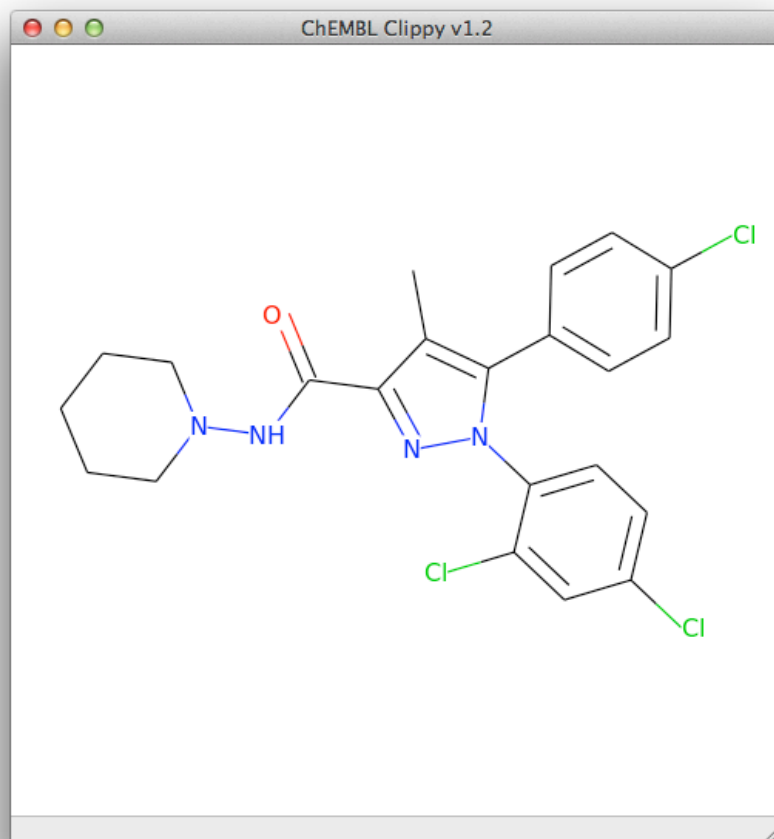
The answer: Clippy

- 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



Technologies

clippy



wxPython

beaker



OSRA

← smiles2ctab, ctab2smiles →

← ctab2inchi, inchi2inchikey →

← image2ctab, ctab2image →

← chembl2smiles →



ChEMBL WS

Packaging



PyInstaller

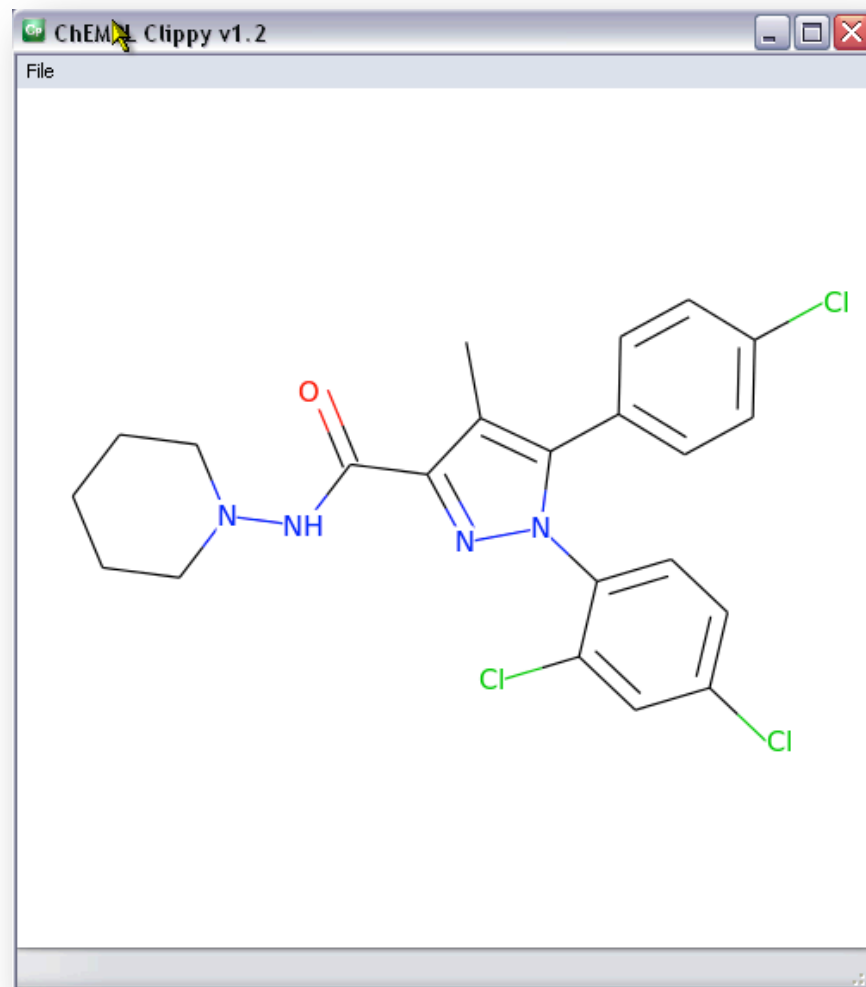
```
#!bin/env python  
import os  
import sys
```

```
111100011100 010011100011100  
1111111101110100100011110111010  
111111111000111110000111100011
```

```
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

Clippy: A cross-platform wxPython app

George Papadatos, PhD

Senior Technical Officer

ChEMBL group

georgep@ebi.ac.uk

EMBL-EBI

