GDC
CADD

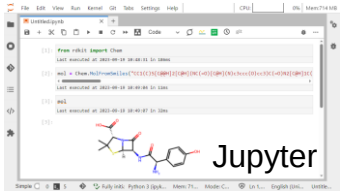
RDKit Office Add-in

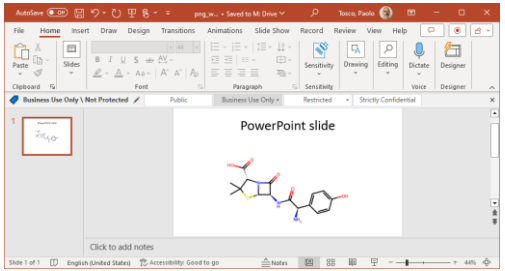
Paolo Tosco
13th RDKit UGM, Zurich
September 12, 2024

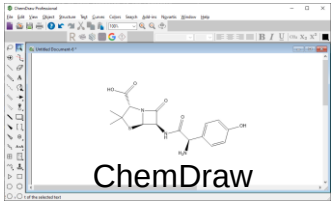
NOVARTIS | Reimagining Medicine

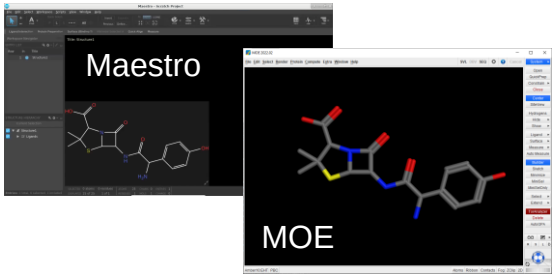
TL;DR: RDKit Office Add-in

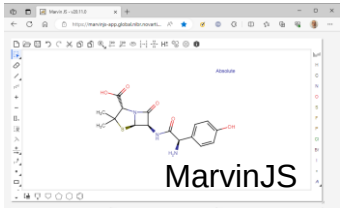
- Exchange molecule PNGs between Office documents and other apps through the clipboard retaining chemical structure information
- Embed molecule PNGs with chemical structure information into Office documents generated by your JS web app (e.g., Jupyter)











NOVARTIS | Reimagining Medicine

RDKit can deal with PNG metadata, but Office can't

[1]:
```
from rdkit import Chem
from rdkit.Chem.Draw import rdDepictor, rdMolDraw2D
from IPython.display import Image
```
Last executed at 2024-09-01 18:47:15 in 198ms

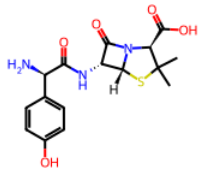
[2]:
```
amoxicillin_smi = "CC1([C@@H](N2[C@H](S1)[C@@H](C2=O)NC(=O)[C@@H](C3=CC=C(C=C3)O)N)C(=O)O)C"
amoxicillin = Chem.MolFromSmiles(amoxicillin_smi)
rdDepictor.SetPreferCoordGen(True)
rdDepictor.Compute2DCoords(amoxicillin);
```
Last executed at 2024-09-01 18:47:15 in 17ms

By default, RDKit writes molecule metadata (CXSMILES, CTAB and pickled mol) to molecule PNG strings:

[3]:
```
drawer = rdMolDraw2D.MolDraw2DCairo(250, 150)
drawer.DrawMolecule(amoxicillin)
drawer.FinishDrawing()
```
Last executed at 2024-09-01 18:47:22 in 7ms

[4]:
```
png = drawer.GetDrawingText()
Image(png)
```
Last executed at 2024-09-01 18:47:25 in 32ms

[4]:



RDKit's Python API has convenient functions to access PNG metadata as a `dict`:

[5]:
```
png_metadata = Chem.MetadataFromPNGString(png)
print("\n".join(png_metadata.keys()))
```
Last executed at 2024-09-01 18:47:25 in 27ms

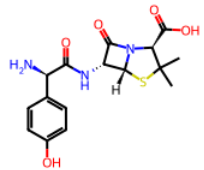
```
rdkitPKL rdkit 2024.09.1pre
MOL rdkit 2024.09.1pre
SMILES rdkit 2024.09.1pre
```

RDKit stores structure metadata in the PNG (pickle, CXSMILES, CTAB)

The original amoxicillin molecule can be reproduced from the CXSMILES metadata embedded in the PNG image:

[6]:
```
drawer = rdMolDraw2D.MolDraw2DCairo(250, 150)
drawer.DrawMolecule(Chem.MolFromSmiles(png_metadata["SMILES rdkit 2024.09.1pre"].decode()))
drawer.FinishDrawing()
Image(drawer.GetDrawingText())
```
Last executed at 2024-09-01 18:47:28 in 25ms

[6]:

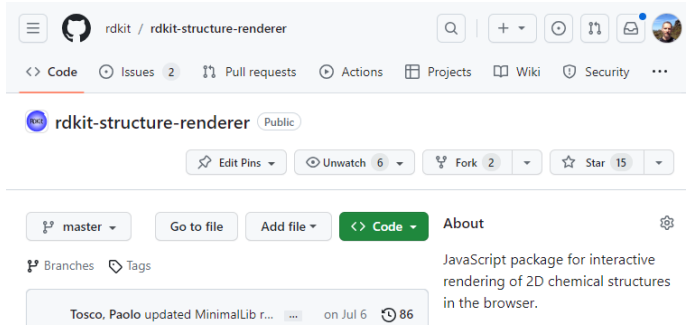


Unfortunately, Office strips metadata from PNG upon exporting/importing images to/from clipboard and files

rdkit-office-addin

```
PPTX_Presentation
├── [Content_Types].xml
├── docProps
│   ├── app.xml
│   └── core.xml
├── ppt
│   ├── charts
│   │   └── _rels
│   ├── embeddings
│   ├── media
│   │   ├── image1.png
│   │   └── image2.png
│   ├── notesMasters
│   │   ├── notesMaster1.xml
│   │   └── _rels
│   │       └── notesMaster1.xml.rels
│   ├── notesSlides
│   │   ├── notesSlide1.xml
│   │   ├── notesSlide2.xml
│   │   └── _rels
│   │       ├── notesSlide1.xml.rels
│   │       └── notesSlide2.xml.rels
│   ├── presentation.xml
│   ├── presProps.xml
│   ├── _rels
│   │   └── presentation.xml.rels
│   ├── slideLayouts
│   │   ├── _rels
│   │   │   └── slideLayout1.xml.rels
│   │   └── slideLayout1.xml
│   ├── slideMasters
│   │   ├── _rels
│   │   │   └── slideMaster1.xml.rels
│   │   └── slideMaster1.xml
│   ├── slides
│   │   ├── _rels
│   │   │   ├── slide1.xml.rels
│   │   │   └── slide2.xml.rels
│   │   ├── slide1.xml
│   │   └── slide2.xml
```

npm package depending on `rdkit-structure-renderer`

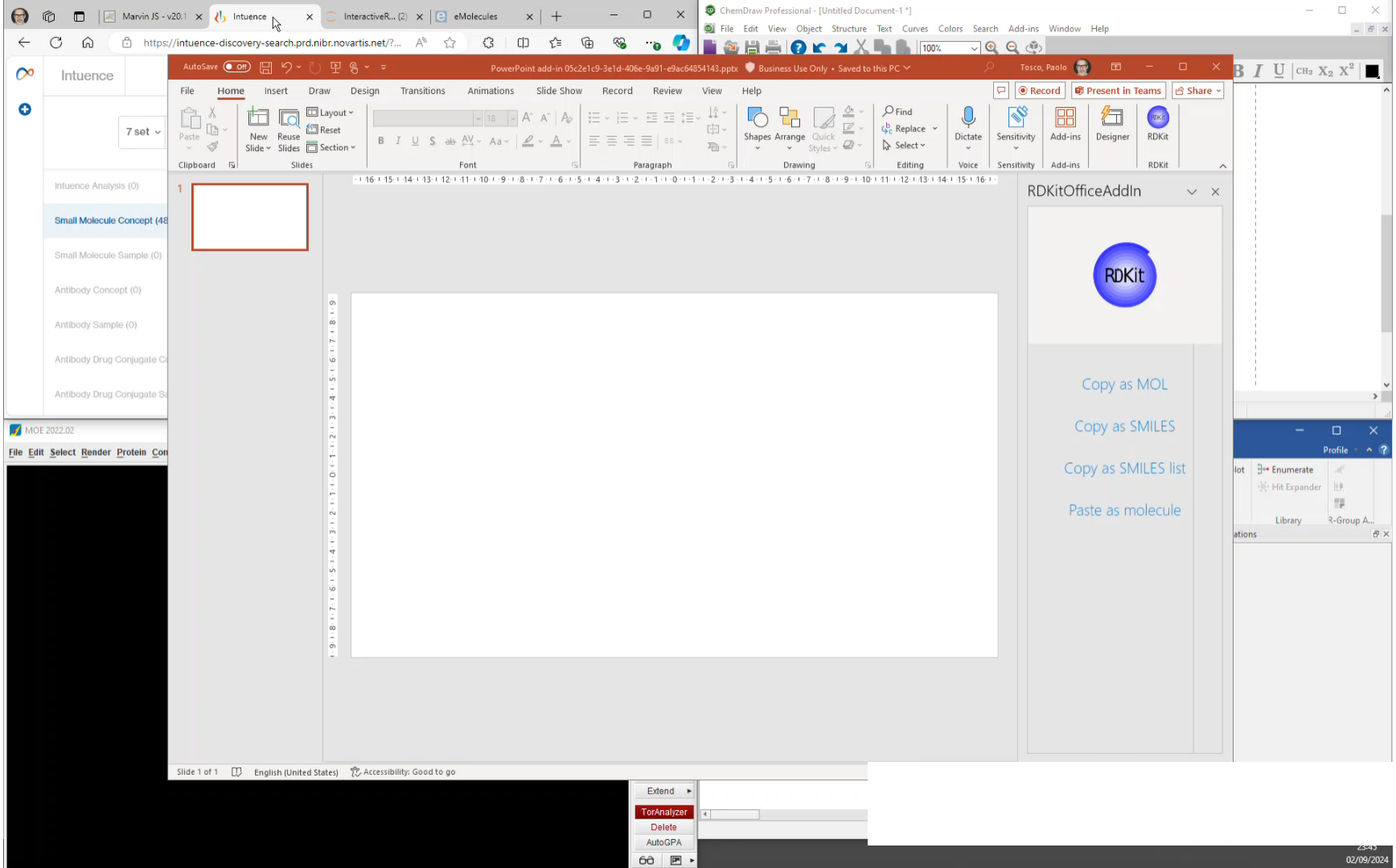


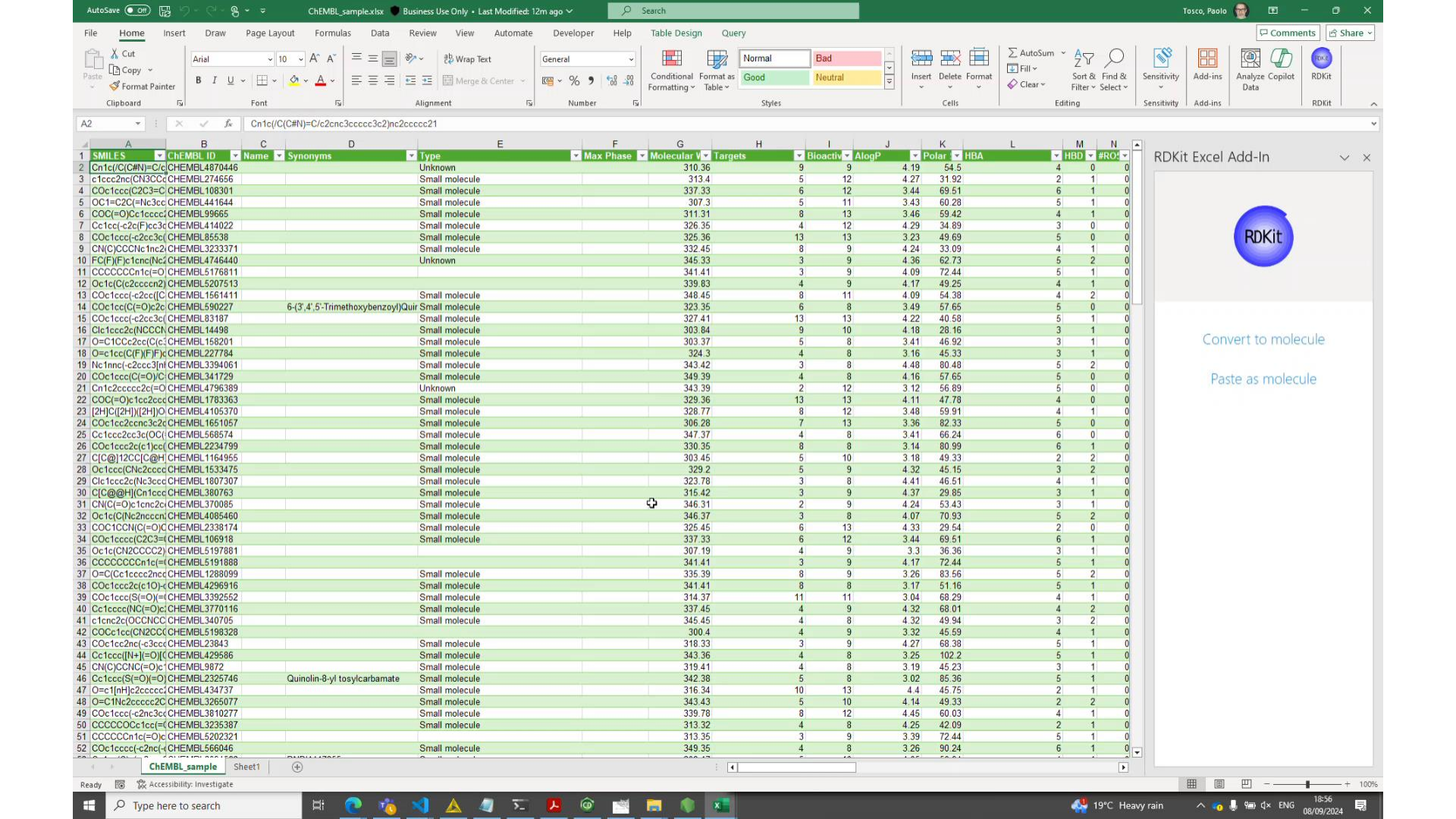
When a selection is to be copied to clipboard, the add-in looks for `/ppt/media/image###.png` objects in the Open XML tree

RDKit metadata are extracted and exchanged on the clipboard as plain text by the JS Office add-in

```
function getOOXml(doc, platform);
async function getSelectedImages(slideId);
function getMolSizeAngstrom(xyzArray);
function extractMolFromPngBlob(rdkitModule, data);
function getMolAndCentroidFromImage(rdkitModule, {
left, top, width, height, data }, molToImageRatio);
async function getCombinedMolFromImages(images)
async function getSelectedImagesWithData(event);
async function getSmilesArrayFromImages(images);
async function onCopyCommon(event);
async function onCopySmiles(event, sep, term);
async function onPasteSmilesOrMolBlock(event);
```

RDKit Office Add-In demo


Marvin JS - v20.1
Intuence
InteractiveR... (2)
eMolecules
https://intuence-discovery-search.prd.nibr.novartis.net/?...
Intuence
7 set
Intuence Analysis (0)
Small Molecule Concept (48
Small Molecule Sample (0)
Antibody Concept (0)
Antibody Sample (0)
Antibody Drug Conjugate C
Antibody Drug Conjugate Sa
ChemDraw Professional - [Untitled Document-1 *]
File Edit View Object Structure Text Curves Colors Search Add-ins Window Help
100%
AutoSave
PowerPoint add-in 05c2e1c9-3e1d-406e-9a91-e9ac64854143.pptx
Business Use Only • Saved to this PC
Tosco, Paolo
File Home Insert Draw Design Transitions Animations Slide Show Record Review View Help
Record
Present in Teams
Share
Paste
Clipboard
New Slide
Reuse Slides
Layout
Reset
Section
Slides
Font
Paragraph
Shapes
Arrange
Quick Styles
Drawing
Find
Replace
Select
Editing
Dictate
Voice
Sensitivity
Add-ins
Designer
RDKit
1
Slide 1 of 1
English (United States)
Accessibility: Good to go
RDKitOfficeAddIn
RDKit
Copy as MOL
Copy as SMILES
Copy as SMILES list
Paste as molecule
MOE 2022.02
File Edit Select Render Protein Con
Extend
TorAnalyzer
Delete
AutoGPA
Profile
Enumerate
Hit Expander
Library
R-Group A...
02/09/2024


AutoSave
ChEMBL_sample.xlsx
Business Use Only • Last Modified: 12m ago
Search
Tosco, Paolo
File
Home
Insert
Draw
Page Layout
Formulas
Data
Review
View
Automate
Developer
Help
Table Design
Query
Comments
Share
Cut
Copy
Format Painter
Paste
Arial
10
Wrap Text
Merge & Center
General
Normal
Bad
Good
Neutral
Conditional Formatting
Format as Table
Insert
Delete
Format
AutoSum
Fill
Clear
Sort & Filter
Find & Select
Sensitivity
Add-ins
Analyze Data
Copilot
RDKit
Clipboard
Font
Alignment
Number
Styles
Cells
Editing
A2
Cn1c(/C(C#N)=C/c2cnc3ccccc3c2)nc2ccccc21
SMILES
ChEMBL_ID
Name
Synonyms
Type
Max Phase
Molecular W
Targets
Bioactiv
AlogP
Polar
HBA
HBD
#RO5
Cn1c(/C(C#N)=C/c
CHEMBL4870446
Unknown
310.36
9
9
4.19
54.5
4
0
0
c1ccc2nc(CN3CCc
CHEMBL274656
Small molecule
313.4
5
12
4.27
31.92
2
1
0
COc1ccc(C2C3=C
CHEMBL108301
Small molecule
337.33
6
12
3.44
69.51
6
1
0
OC1=C2C(=Nc3cc
CHEMBL441644
Small molecule
307.3
5
11
3.43
60.28
5
1
0
COC(=O)Cc1cccc
CHEMBL99665
Small molecule
311.31
8
13
3.46
59.42
4
1
0
Cc1cc(-c2c(F)cc3c
CHEMBL414022
Small molecule
326.35
4
12
4.29
34.89
3
0
0
COc1ccc(-c2cc3c(
CHEMBL85538
Small molecule
325.36
13
13
3.23
49.69
5
0
0
CN(C)CCCNc1nc2
CHEMBL3233371
Small molecule
332.45
8
9
4.24
33.09
4
1
0
FC(F)(F)c1cnc(Nc2
CHEMBL4746440
Unknown
345.33
3
9
4.36
62.73
5
2
0
CCCCCCCn1c(=O
CHEMBL5176811
341.41
3
9
4.09
72.44
5
1
0
Oc1c(C(c2cccn2)
CHEMBL5207513
339.83
4
9
4.17
49.25
4
1
0
COc1ccc(-c2cc([C
CHEMBL1561411
Small molecule
348.45
8
11
4.09
54.38
4
2
0
COc1cc(C(=O)c2c
CHEMBL590227
6-(3',4',5'-Trimethoxybenzoyl)Quir
Small molecule
323.35
6
8
3.49
57.65
5
0
0
COc1ccc(-c2cc3c(
CHEMBL83187
Small molecule
327.41
13
13
4.22
40.58
5
1
0
Clc1ccc2c(NCCCN
CHEMBL14498
Small molecule
303.84
9
10
4.18
28.16
3
1
0
O=C1CCc2cc(C(c
CHEMBL158201
Small molecule
303.37
5
8
3.41
46.92
3
1
0
O=c1cc(C(F)(F)F)c
CHEMBL227784
Small molecule
324.3
4
8
3.16
45.33
3
1
0
Nc1nnc(-c2ccc3[nH
CHEMBL3394061
Small molecule
343.42
3
8
4.48
80.48
5
2
0
COc1ccc(C(=O)/C
CHEMBL341729
Small molecule
349.39
4
8
4.16
57.65
5
0
0
Cn1c2ccccc2c(=O
CHEMBL4796389
Unknown
343.39
2
12
3.12
56.89
5
0
0
COC(=O)c1cc2ccc
CHEMBL1783363
Small molecule
329.36
13
13
4.11
47.78
4
0
0
[2H]C([2H])([2H])O
CHEMBL4105370
Small molecule
328.77
8
12
3.48
59.91
4
1
0
COc1cc2ccnc3c2c
CHEMBL1651057
Small molecule
306.28
7
13
3.36
82.33
5
0
0
Cc1ccc2cc3c(OC(
CHEMBL568574
Small molecule
347.37
4
8
3.41
66.24
6
0
0
COc1ccc2c(c1)cc(
CHEMBL2234799
Small molecule
330.35
8
8
3.14
80.99
6
1
0
C[C@]12CC[C@H
CHEMBL1164955
Small molecule
303.45
5
10
3.18
49.33
2
2
0
Oc1ccc(CNc2ccc
CHEMBL1533475
Small molecule
329.2
5
9
4.32
45.15
3
2
0
Clc1ccc2c(Nc3ccc
CHEMBL1807307
Small molecule
323.78
3
8
4.41
46.51
4
1
0
C[C@@H](Cn1ccc
CHEMBL380763
Small molecule
315.42
3
9
4.37
29.85
3
1
0
CN(C(=O)c1cnc2c(
CHEMBL370085
Small molecule
346.31
2
9
4.24
53.43
3
1
0
Oc1c(C(Nc2nccn
CHEMBL4085460
Small molecule
346.37
3
8
4.07
70.93
5
2
0
COC1CCN(C(=O)C
CHEMBL2338174
Small molecule
325.45
6
13
4.33
29.54
2
0
0
COc1cccc(C2C3=(
CHEMBL106918
Small molecule
337.33
6
12
3.44
69.51
6
1
0
Oc1c(CN2CCCC2)
CHEMBL5197881
307.19
4
9
3.3
36.36
3
1
0
CCCCCCCCn1c(=(
CHEMBL5191888
341.41
3
9
4.17
72.44
5
1
0
O=C(Cc1cccc2ncc
CHEMBL1288099
Small molecule
335.39
8
9
3.26
83.56
5
2
0
COc1ccc2c(c1O)-c
CHEMBL4296916
Small molecule
341.41
8
8
3.17
51.16
5
1
0
COc1ccc(S(=O)(=(
CHEMBL3392552
Small molecule
314.37
11
11
3.04
68.29
4
1
0
Cc1cccc(NC(=O)c
CHEMBL3770116
Small molecule
337.45
4
9
4.32
68.01
4
2
0
c1cnc2c(OCCNCC
CHEMBL340705
Small molecule
345.45
4
8
4.32
49.94
3
2
0
COCc1cc(CN2CC(
CHEMBL5198328
300.4
4
9
3.32
45.59
4
1
0
COc1cc2nc(-c3ccc
CHEMBL23843
Small molecule
318.33
3
9
4.27
68.38
5
1
0
Cc1ccc([N+](=O)[(
CHEMBL429586
Small molecule
343.36
4
8
3.25
102.2
5
1
0
CN(C)CCNC(=O)c1
CHEMBL9872
Small molecule
319.41
4
8
3.19
45.23
3
1
0
Cc1ccc(S(=O)(=O)
CHEMBL2325746
Quinolin-8-yl tosylcarbamate
Small molecule
342.38
5
8
3.02
85.36
5
1
0
O=c1[nH]c2ccccc2
CHEMBL434737
Small molecule
316.34
10
13
4.4
45.75
2
1
0
O=C1Nc2ccccc2C
CHEMBL3265077
Small molecule
343.43
5
10
4.14
49.33
2
2
0
COc1ccc(-c2nc3cc
CHEMBL3810277
Small molecule
339.78
8
12
4.45
60.03
4
1
0
CCCCCOc1cc(=(
CHEMBL3235387
Small molecule
313.32
4
8
4.25
42.09
2
1
0
CCCCCCn1c(=O)c
CHEMBL5202321
313.35
3
9
3.39
72.44
5
1
0
COc1cccc(-c2nc(-(
CHEMBL566046
Small molecule
349.35
4
8
3.26
90.24
6
1
0
RDKit Excel Add-In
RDKit
Convert to molecule
Paste as molecule
ChEMBL_sample
Sheet1
Ready
Accessibility: Investigate
100%
Type here to search
19°C Heavy rain
ENG
18:56
08/09/2024

Thank you

NOVARTIS | Reimagining Medicine