

# BME Chemical Toolkit 1.0

Created by Patrick Chirdon 2020

with the support of Dr.Douglas Goetz, Dr.Sumit Sharma, Dr.Colin Kruse, and Athena Ghiasi  
at Ohio University

Requires Python

Spark

Java

Anaconda

Virtualbox

## 1.Test Pickle

Input the pickle name into the first box. The file must end in .pkl. Then input a list of smiles into the second box, click test pickle and then click submit. The output is a target probability.

The screenshot shows a Linux desktop with a background image of a molecular structure. The desktop contains several application icons: Terminal, Task, DataWarrior, Trash, Chromium, LibreOffice 6.4 Writer, datawarrior\_linux, Anaconda, LibreOffice 6.4 Calc, jre1.8.0\_251, Jupyter, LibreOffice 6.4 Draw, LibreOffice 6.4.2 Linux..., test.xlsx, Autodock, best.smi, Aldrugdiscovery-master3, Chrome Remote Desktop, PCManFM, Autodock, Synaptic, and Bmetoolkit. A terminal window is open, displaying a warning message: "WARNING:tensorflow:site-packages/tensorflow/python/framework/op\_def\_library.py:262: ModuleNotFoundError: No module named 'tensorflow.python.framework.op\_def\_library'. Please check your TensorFlow installation. Instructions for installing TensorFlow from source: https://www.tensorflow.org/install/source". The Biomedical Engineering Chemical Toolkit 1.0 window is the main focus. It has a menu bar with "Help", "Report Bug", and "Request Solution". The main area contains a form for creating a bioassay or screening compounds. The form has a text input field for the molecule population name, which is currently "Phospholipidosis.pkl". Below this is a list of chemical structures (SMILES) for the population. The list includes: CC(C)NCC(COC1=CC=C(C=C1)CCCC)O, CC(C)NCC(COC1=CC=C(C=C1)CCCC)O, and CC(C)NCC(COC1=CC=C(C=C1)CCCC)O. There is a "Test H5" button and a "Submit" button. A progress bar is shown at the bottom of the window, indicating 0% completion. A large banner at the bottom of the toolkit window reads "AI for Medicine". The system tray at the bottom shows the taskbar with icons for the terminal, toolkit, and other applications, along with the system clock showing "ES 17:00".

Biomedical Engineering Chemical Toolkit

Help Report Bug Request Solution

Create a Bioassay or Screen Your Compounds!

Enter the name of your molecule population here

Phospholipidosis.pkl

CC(C)NCC(COC1=CC=C(C=C1)CCCC)O  
CC(C)NCC(COC1=CC=C(C=C1)CCCC)O  
CC(C)NCC(COC1=CC=C(C=C1)CCCC)O

Test H5

Target Lookup  
Download Compounds  
Build Regression Model  
Library Creation  
Molecular Descriptors  
Mesh Molecules

Target Probability

CC(C)NCC(COC1=CC=C(C=C1)CCCC)O 0.3251430883868382

Target Probability

CC(C)NCC(COC1=CC=C(C=C1)CCCC)O 0.3251430883868382

Target Probability

Submit

Clear Images View Draw SDF 3DSDF Spreadsheet

0%

AI for Medicine

**BME Chemical Toolkit 1.0**

Warning is r  
models wit  
warnings.  
Using Tens  
2020-06-15  
ader.cc:55]  
r.so.6: can  
2020-06-15  
ader.cc:55]  
bnvinfer\_pl  
2020-06-15  
30] Cannot  
with TensorR  
ed properly  
Unna  
10438  
12476  
WARNING:ten  
site-packag  
riables (fr  
moved in a  
Instruction  
non-resourc

## 2. Test H5

Type the name of the .h5 in the first box. Type a list of SMILES into the second box and click submit.  
The output is a probability.

The screenshot shows a Linux desktop with various application icons on the left, including Terminal, Task, DataWarrior, Trash, Chromium, LibreOffice 6.4 Writer, datawarrior\_linux, Anaconda, LibreOffice 6.4 Calc, jre1.8.0\_251, Jupyter, LibreOffice 6.4 Draw, LibreOffice 6.4.2 Linux, test.xlsx, Autodock, best.smi, Aldrugdiscovery-master3, Chrome Remote Desktop, PCManFM, Autodock, Synaptic, and Bmetoolkit. The background features a molecular structure visualization. The Biomedical Engineering Chemical Toolkit application is open in the center, displaying a window titled "Biomedical Engineering Chemical Toolkit". The window contains a menu bar with "Help", "Report Bug", and "Request Solution". Below the menu bar is a text input field labeled "Enter the name of your molecule population here" with the text "classifier.h5" entered. Below this is a list of SMILES strings: CC(C)NCC(COC1=CC=C(C=C1)CCOC)O, CC(C)NCC(COC1=CC=C(C=C1)CCOC)O, and CC(C)NCC(COC1=CC=C(C=C1)CCOC)O. Below the list is a "Submit" button. Below the "Submit" button is a row of buttons: "Clear", "Images", "View", "Draw", "SDF", "3DSDF", and "Spreadsheet". Below the buttons is a progress bar showing 0%. At the bottom of the window is a large black banner with the text "AI for Medicine" and a microscope icon. The taskbar at the bottom shows several open applications: Aldrugdiscovery-master3, Untitled 1 - LibreOffice..., bmetoolkit1.0.sh, Biomedical Engineering..., and best.csv - LibreOffice... The system clock in the bottom right corner shows "ES 19:02".

### 3. Target Lookup

Type in the name of the protein target into the first box.

The output will be a ChEMBL ID. ie

C[C@@H](C(=O)N[C@@H](CO)C(=O)N1CCC[C@H]1C(=O)N[C@@H](CCCN=C(N)N)C(=O)N[C@@H](CCCCN)C(=O)O)NC(=O)[C@H](CC2=CN=CN2)NC(=O)[C@H](CC3=CN=CN3)N

Unnamed: 0            0            1            2

10435	10435	CHEMBL3038469	Homo sapiens	CDK2/Cyclin A
10436	10436	CHEMBL3038470	Homo sapiens	CDK2/Cyclin A1
10483	10483	CHEMBL3038517	Homo sapiens	CDK2/CDK4
12444	12444	CHEMBL4106152	Homo sapiens	CDK2/Cyclin O
12445	12445	CHEMBL4106153	Homo sapiens	CDK2/Cyclin A
12477	12477	CHEMBL4106185	Mus musculus	CDK2/CDK4
12478	12478	CHEMBL4106186	Mus musculus	CDK2/CDK9

Biomedical Engineering Chemical Toolkit

Help Report Bug Request Solution

Create a Bioassay or Screen Your Compounds!

Enter the name of your molecule population here

CDK2

Enter up to 100 smiles here, One smile only for toxicity map.

Test Pickle  
Test H5  
Target Lookup  
Download Compounds  
Build Regression Model  
Library Creation  
Molecular Descriptors  
Mash Molecules

Submit

Clear Images View Draw SIF 3DSIF Spreadsheet

0%

AI for Medicine

BME CHEMICAL TOOLKIT 1.0

Terminal Task DataWarrior

Trash Chromium LibreOffice 6.4 Writer

datawarrior\_li nux Anaconda LibreOffice 6.4 Calc

jre1.8.0\_251 Jupyter LibreOffice 6.4 Draw

LibreOffice 6.4.2 Linux\_... test.xlsx

Autodock best.smil

LibreOffice 6.4.2 Linux\_... Chrome Remote Des...

PCManFM Autodock

Synaptic Bmetoolkit

[(2) Pcmanfm] [Untitled 1 - LibreOffice...] (2) Lxterminal (2) Biomedical Enginee... [best.csv - LibreOffice ...]

ES 19:27

## 4. Download Compounds



Input the ChEMBL ID into the first box and click submit. The output will be a list of smiles and their corresponding ChEMBLID for the compound

Biomedical Engineering Chemical Toolkit

Help Report Bug Request Solution

Create a Bioassay or Screen Your Compounds!

Enter the name of your molecule population here

CHEMBL3038471

Enter up to 100 smiles here. One smile only for toxicity map.

Test Pickle  
Test HS  
Target Lookup  
Download Compounds  
Build Regression Model  
Library Creation  
Molecular Descriptors  
Wash Molecules

There are 87 compounds initially identified as active on the known targets, e.g.  
["CHEMBL3716375", "CHEMBL3714855"]  
ChEMBLID, smiles

0, CHEMBL308978, CN[C@@H]1C[C@H]2O[C@H](C)([C@@H]1OC)n1c3ccccc3c3c4c(c5c6ccccc6n2c5c31)C(=O)NC4  
1, CHEMBL1944638, CN1C/C=C/C/COC2CCCC(C2)=C2C=CC(C2)C1  
2, CHEMBL2377825, Nc1nc(Nc2ccc(S(N)(=O)=O)cc2)sc1C(=O)c1ccccc1[N+](=O)[O-]

Submit  
Clear Images View Draw SDF 3DSDF Spreadsheet

0%

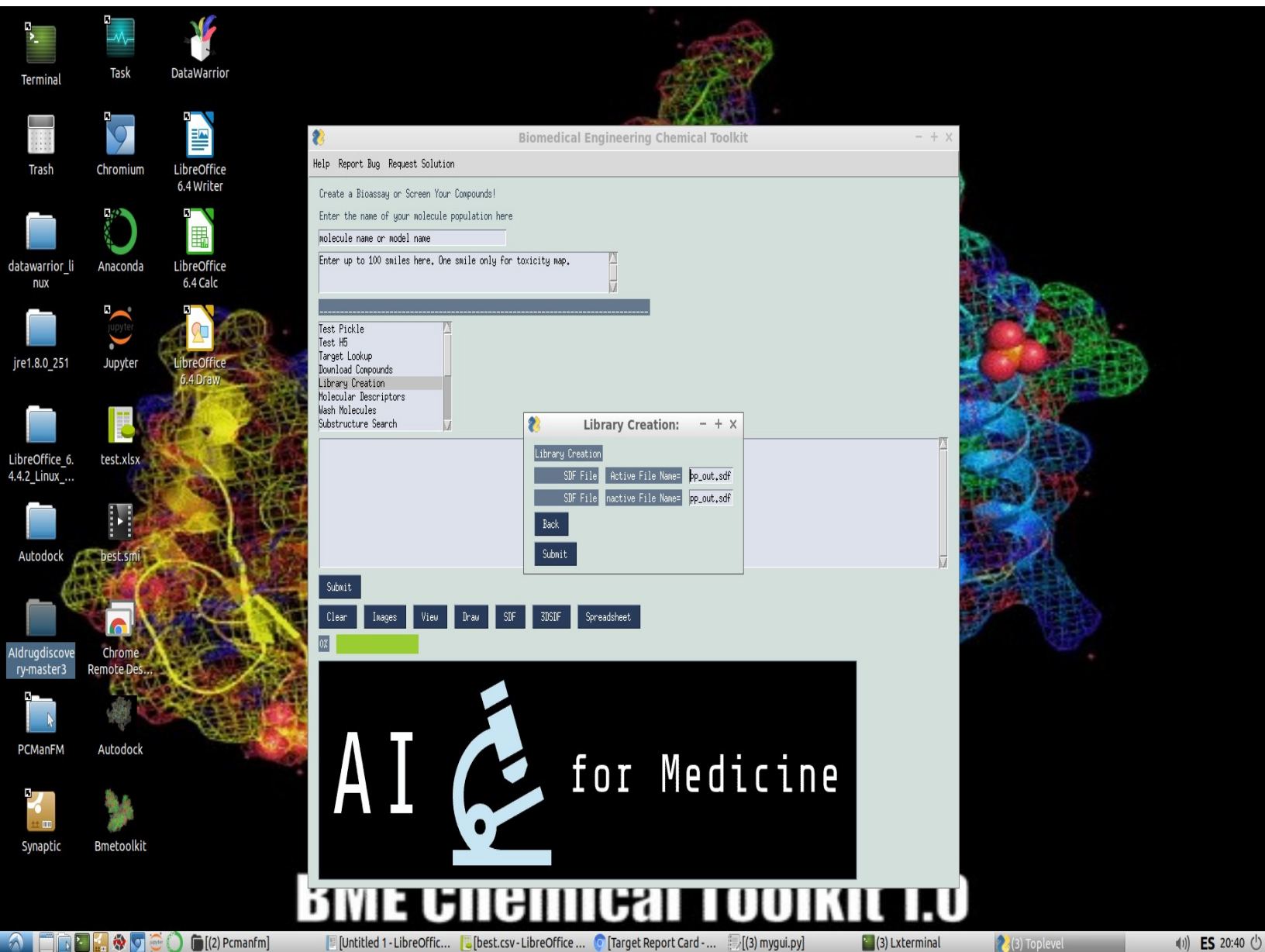
AI for Medicine

Medical Toolkit 1.0

Warning is raised when loading pickled models, models with scikit-learn 0.21+.  
warnings.warn(msg, category=DeprecationWarning)  
Using TensorFlow backend.  
2020-06-15 20:09:19.723891: W tensorflow/stream\_executor.cc:55] Could not load dynamic library 'libnvinfer.so.6: cannot open shared object file: No such file or directory.  
2020-06-15 20:09:19.724050: W tensorflow/stream\_executor.cc:55] Could not load dynamic library 'libnvinfer\_plugin.so.6: cannot open shared object file: No such file or directory.  
2020-06-15 20:09:19.724093: W tensorflow/compiler/xrt/compiler.cc:30] Cannot dlopen some TensorRT libraries. If you are using the XRT compiler, please make sure the missing libraries are properly loaded.  
Unnamed: 0  
10438 10438 CHEMBL3038472 Homo sapiens  
12476 12476 CHEMBL4106184 Homo sapiens  
WARNING:tensorflow:From /home/lubuntu/anaconda3/lib/python3.7/site-packages/tensorflow\_core/python/compat/v2\_compat.py:107: Dlopen of variables (from tensorflow.python.ops.variables) is deprecated and will be removed in a future version.  
Instructions for updating:  
Use tf.compat.v2.nn.conv2d instead  
non-resource variables are not supported in the long term

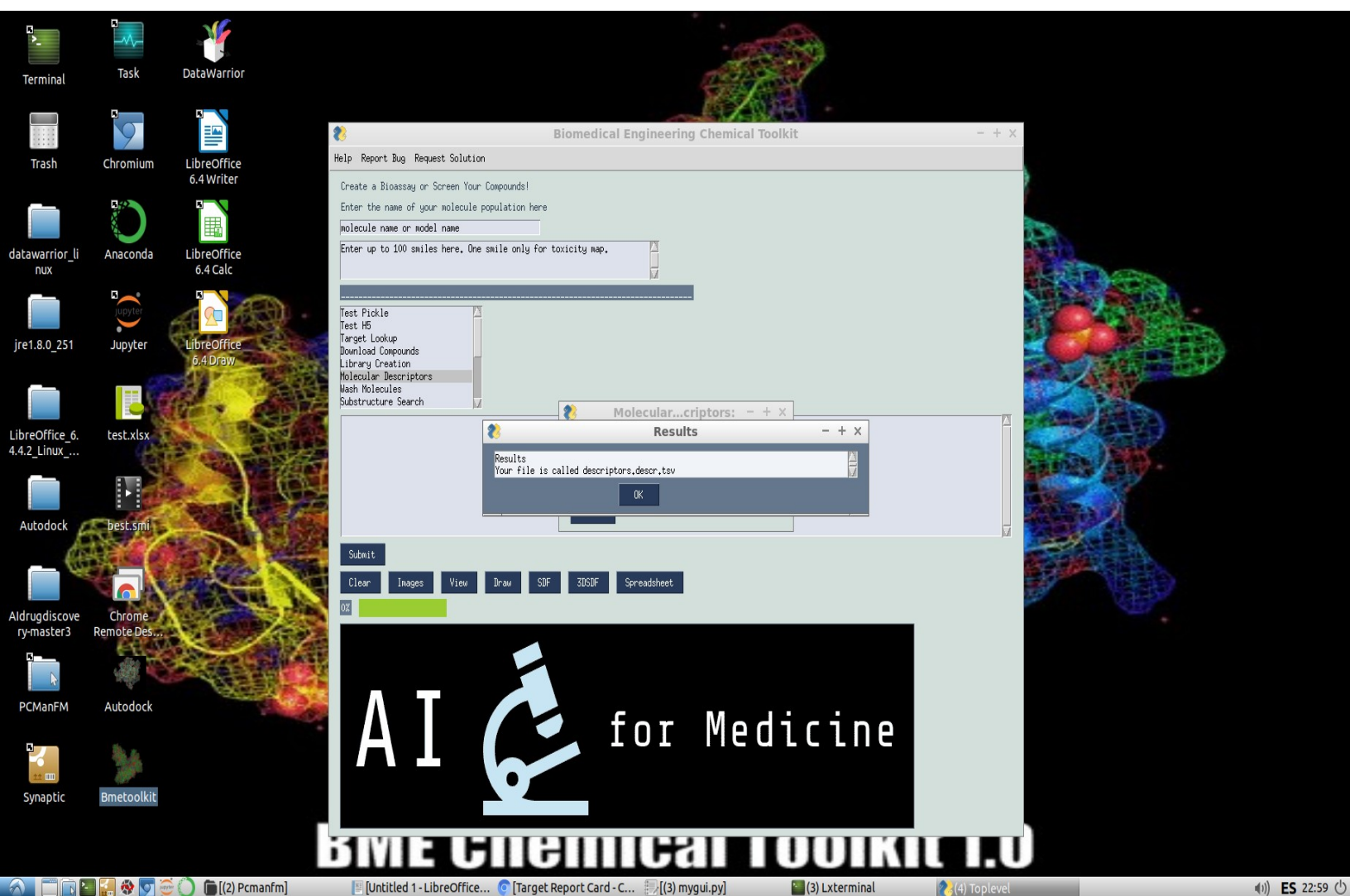
## 5. Library Creation

Create a library of compounds labeled active, 1, or inactive, 0 as a SDF file. Input an active file name containing compounds that are known to bind a target and input an inactive file name containing compounds that do not bind (could also be just random compounds)



## 6. Molecular Descriptors

Input a SDF file. The output file will be called descriptors.descr.tsv. The output will be an excel file with chemical properties.





## 7. Wash Molecules

Input an SDF molecule. The program will remove some metal atoms and ions, remove small fragments, adjust formal charges, and rebalance hydrogen counts.

The screenshot displays the Biomedical Engineering Chemical Toolkit (BME Chemical Toolkit 1.0) interface. The main window is titled "Biomedical Engineering Chemical Toolkit" and contains a menu bar (Help, Report Bug, Request Solution) and a toolbar (Clear, Images, View, Draw, SDF, 3DSDF, Spreadsheet). The main area is divided into two panes. The left pane, titled "Create a Bioassay or Screen Your Compounds!", contains a text input field for "Enter the name of your molecule population here" (containing "metoprolo") and a list of actions: Test Pickle, Test H5, Target Lookup, Download Compounds, Library Creation, Molecular Descriptors, Wash Molecules (selected), and Substructure Search. The right pane, titled "Wash Molecules:", contains a text input field for "SDF File To Wash" (containing "bp\_out.sdf") and buttons for "Back" and "Submit". A progress bar at the bottom of the main window shows 0% completion. A large black banner at the bottom of the screen features the text "AI for Medicine" and a microscope icon. The desktop background shows various application icons (Terminal, Task, DataWarrior, LibreOffice, Anaconda, Jupyter, etc.) and a molecular structure visualization.

Biomedical Engineering Chemical Toolkit

Help Report Bug Request Solution

Create a Bioassay or Screen Your Compounds!

Enter the name of your molecule population here

metoprolo

CC(C)NCC(COC1=CC=C(C=C1)COC)O  
CC(C)NCC(COC1=CC=C(C=C1)COC)O  
CC(C)NCC(COC1=CC=C(C=C1)COC)O

Test Pickle  
Test H5  
Target Lookup  
Download Compounds  
Library Creation  
Molecular Descriptors  
Wash Molecules  
Substructure Search

Wash Molecules:

Wash Molecules

SDF File SDF File To Wash bp\_out.sdf

Back Submit

Submit

Clear Images View Draw SDF 3DSDF Spreadsheet

0%

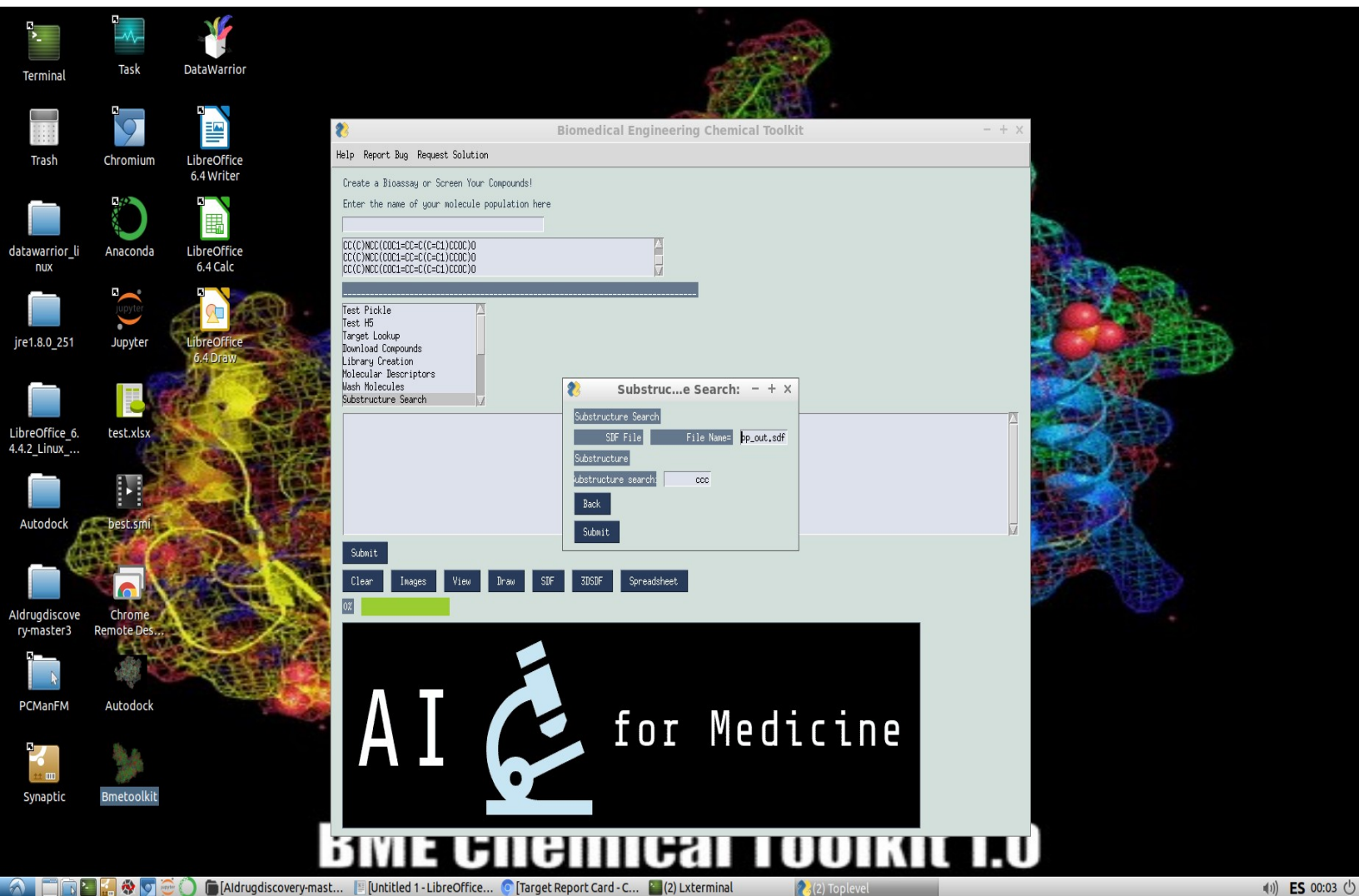
AI for Medicine

BME CHEMICAL TOOLKIT 1.0

[Aldrugdiscovery-mast... [Untitled 1 - LibreOffice... [Target Report Card - C... (2) Lxterminal (2) Toplevel ES 23:37

## 8. Substructure Search

For a substructure search input a SDF file and search for a substructure. The results output will be a list of molecules containing the substructures.



The screenshot displays the Biomedical Engineering Chemical Toolkit (BME Chemical Toolkit 1.0) interface. The main window is titled "Biomedical Engineering Chemical Toolkit" and contains a menu bar (Help, Report Bug, Request Solution) and a toolbar (Submit, Clear, Images, View, Draw, SDF, 3DSDF, Spreadsheet). The main area has a text input field for "Enter the name of your molecule population here" and a list of molecules. A substructure search window is open, titled "Substructure Search", with fields for "SDF File" (File Name: bp\_out.sdf), "Substructure", and "Substructure search" (ccc). The background features a 3D molecular model of a protein-ligand complex. A large watermark "AI for Medicine" is overlaid on the bottom right, and "BME CHEMICAL TOOLKIT 1.0" is visible at the bottom.

Terminal Task DataWarrior

Trash Chromium LibreOffice 6.4 Writer

datawarrior\_linux Anaconda LibreOffice 6.4 Calc

jre1.8.0\_251 Jupyter LibreOffice 6.4 Draw

LibreOffice 6.4.2 Linux... test.xlsx

Autodock best.smi

Aldrugdiscovery-master3 Chrome Remote Desktop

PCManFM Autodock

Synaptic Bmetoolkit

Biomedical Engineering Chemical Toolkit

Help Report Bug Request Solution

Create a Bioassay or Screen Your Compounds!

Enter the name of your molecule population here

CC(C)NCC(COC1=CC=C(C=C1)COC)O  
CC(C)NCC(COC1=CC=C(C=C1)COC)O  
CC(C)NCC(COC1=CC=C(C=C1)COC)O

Test Pickle  
Test H5  
Target Lookup  
Download Compounds  
Library Creation  
Molecular Descriptors  
Mash Molecules  
Substructure Search

Substructure Search

SDF File File Name: bp\_out.sdf

Substructure

Substructure search: ccc

Back Submit

Submit

Clear Images View Draw SDF 3DSDF Spreadsheet

0%

AI for Medicine

BME CHEMICAL TOOLKIT 1.0

[Aldrugdiscovery-master3] [Untitled 1 - LibreOffice...] [Target Report Card - C...] (2) Lxterminal (2) Toplevel ES 00:03

## 9. Pharmacophore

Input two smiles and click pharmacophore. The output will be the pharmacophore similarity. 1 means that the molecules are the same. The scale is from 0 to 1.

The screenshot displays the Biomedical Engineering Chemical Toolkit 1.0 interface. The desktop background features a molecular structure visualization. The application window, titled "Biomedical Engineering Chemical Toolkit", contains a menu bar with "Help", "Report Bug", and "Request Solution". Below the menu bar, there is a section titled "Create a Bioassay or Screen Your Compounds!" with a text input field for "Enter the name of your molecule population here". A list box contains two chemical structures: CC(C)NCC(COC1=CC=C(C=C1)CCOC)O and CCCCOC(1)NCC(COC1=CC=C(C=C1)CCOC)O. A dropdown menu is open, showing options: "Mash Molecules", "Substructure Search", "Pharmacophore" (selected), "PAINS", "Chembl Prediction", "Toxicity Prediction", "Toxicity Map", and "Build a SAR model". Below the list box, a text area displays the similarity score: "0.29411764705882354". A "Submit" button is located below the text area. At the bottom of the window, there is a row of buttons: "Clear", "Images", "View", "Draw", "SIF", "3DSIF", and "Spreadsheet". A progress bar shows "0%". A large black banner at the bottom of the window features the text "AI for Medicine" with a microscope icon. The system tray at the bottom of the screen shows the taskbar with various icons and the system clock displaying "ES 00:54".

## 10. PAINS

### PAN Assay Interference

"The program will tell you what functional groups for each compound were responsible for a positive PAINFUL test result. The program also tells you the fraction of SP3 hybridized carbons. Compounds with scores > .47 are more selective binders. Note that double bonds reduce the fraction of sp3 hybridization, as they make the compound more flat. See Escape from flatland: increasing saturation as an approach to improving clinical success. Pains are defined as follows:

Doveston R, et al. A Unified Lead-oriented Synthesis of over Fifty Molecular Scaffolds. *Org Biomol Chem* 13 (2014) 859D65. doi:10.1039/C4OB02287D

Jadhav A, et al, Quantitative Analyses of Aggregation, Autofluorescence, and Reactivity Artifacts in a Screen for Inhibitors of Thiol Protease. *J Med Chem* 53 (2009) 37D51. doi:10.1021/jm901070c

## 11.Chembl Prediction

### Chembl Target Prediction

#### Multiclass Classifier

Number of unique targets 560

Ion channel 5

kinase 96

nuclear receptor 21

GPCR 180

Others 258

accuracy .87

auc .92

sensitivity .76

specificity .92

precision .82

225/225 100% correct mechanism on independent test set. Note-- 1 is considered positive and zero is negative for a given target.

Terminal Task DataWarrior

Trash Chromium LibreOffice 6.4 Writer

datawarrior\_linux Anaconda LibreOffice 6.4 Calc

jre1.8.0\_251 Jupyter LibreOffice 6.4 Draw

LibreOffice\_6.4.2\_Linux\_Autodock test.xlsx howto.doc

best.smi

Aldrugdiscove ry-master3 Chrome Remote Des...

PCManFM Autodock

Synaptic Bmetoolkit

Biomedical Engineering Chemical Toolkit

Help Report Bug Request Solution

Create a Bioassay or Screen Your Compounds!

Enter the name of your molecule population here

CC(C)NCC(COC1=CC=C(C=C1)CCOC)O

Library Creation  
Molecular Descriptors  
Mash Molecules  
Substructure Search  
Pharmacophore  
PAINS  
Chembl Prediction  
Toxicity Prediction

Target Probability

"Acyl-CoA:cholesterol acyltransferase (Fragment)" 1.0

"Alpha-1A adrenergic receptor (Alpha-1A adrenoreceptor) (Alpha-1A adrenoreceptor) (Alpha-1C adrenergic receptor) (Alpha-adrenergic receptor 1c)" 1.0

"Melanocortin receptor 4 (MC4-R)" 1.0

Submit

Clear Images View Draw SDF 3DSDF Spreadsheet

0%

AI for Medicine

BME CHEMICAL TOOLKIT 1.0

[howto.doc - LibreOffice...] (2) Lxterminal [mygui.py] Biomedical Engineerin...

ES 01:42



## 12. Target Prediction

Support Vector Machine

46 models. Metrics for test data:

accuracy 97.59 +/- 2.41

sensitivity 91.9 +/- 8.1

specificity 98.6 +/- 1.4

215/247 87% correct mechanism on independent test set.

Terminal Task DataWarrior

Trash Chromium LibreOffice 6.4 Writer

datawarrior\_linux Anaconda LibreOffice 6.4 Calc

jre1.8.0\_251 Jupyter LibreOffice 6.4 Draw

LibreOffice 6.4.2 Linux... test.xlsx howto.doc

Autodock best.smi

Aldrugdiscover-master3 Chrome Remote Desktop

PCManFM Autodock

Synaptic Bmetoolkit

Biomedical Engineering Chemical Toolkit

Help Report Bug Request Solution

Create a Bioassay or Screen Your Compounds!

Enter the name of your molecule population here

CC(C)NCC(COC1=CC=C(C=C1)COC)O  
CC(C)NCC(COC1=CC=C(C=C1)COC)O  
CC(C)NCC(COC1=CC=C(C=C1)COC)O

Pharmacophore  
PAINS  
ChEMBL Prediction  
Toxicity Prediction  
Toxicity Map  
Build a SAR model  
Predict solubility  
Fragment Molecules

Target Probability

"modulator of Beta-1 adrenergic receptor" 0.9842084815501039  
"induce Phospholipidosis" 0.9251430883868382  
Phospholipidosis 0.9251430883868382  
"drug Induced Liver Injury" 0.7604610472657531

Submit

Clear Images View Draw SDF 3DSIF Spreadsheet

0%

AI for Medicine

BME CHEMICAL TOOLKIT 1.0

[howto.doc - LibreOffice...] (2) Lxterminal [mygui.py] Biomedical Engineerin... [2020-06-16-015629\_18...]

ES 02:02

## 13. Toxicity Map

Input One SMILE. The output will be 46 toxicity maps for the 46 models available from toxicity prediction.

The screenshot displays the Biomedical Engineering Chemical Toolkit (BME Chemical Toolkit 1.0) interface. The window title is "Biomedical Engineering Chemical Toolkit". The main area contains a text input field for a molecule name, with the SMILES string CC(C)NCC(COC1=CC=C(C=C1)CCOC)O entered. Below the input field is a list of available models, including Pharmacophore, PRINS, Chembl Prediction, Toxicity Prediction, Toxicity Map (which is selected), Build a SAR model, Predict solubility, and Fragment Molecules. A "Submit" button is located below the list. At the bottom of the window, there is a row of buttons: "Clear", "Images", "View", "Draw", "SDF", "3DSDF", and "Spreadsheet". A progress bar shows 39% completion. A large black banner at the bottom of the window features the text "AI for Medicine" with a microscope icon. The desktop background is a molecular structure visualization. The taskbar at the bottom shows various open applications, including LibreOffice, a terminal, and the BME Chemical Toolkit.



## 14. Build a SAR Model

The screenshot displays the Biomedical Engineering Chemical Toolkit 1.0 desktop environment. The desktop background features a molecular structure visualization. The taskbar at the bottom shows several open applications: [howto.doc - LibreOffice...], (2) Lxterminal, [mygui.py], [2020-06-16-015629\_18...], and (2) Toplevel. The system clock indicates ES 02:32.

The main window, titled "Biomedical Engineering Chemical Toolkit", contains a "Machine Learning Front End" sub-window. This sub-window is used for configuring machine learning models. It includes the following sections:

- Create a Bio**: A section for entering molecule names and IDs.
- Machine Learning Command Line Parameters**: A section for searching available assays and finding associated compounds.
- Layers**: A section for configuring the neural network layers, including the number of nodes and drop out rates.
- Pharmacophore**: A section for configuring the pharmacophore model, including the number of layers and drop out rates.
- Stochastic Gradient Descent**: A section for configuring the learning rate, momentum, and decay.
- Compile**: A section for selecting the compilation method (Cross-Entropy, Hinge, or MSE(L1/L2)).
- Support Vector Machine**: A section for configuring the SVM model, including the penalty and model name.

The "Machine Learning Front End" window is currently displaying the "Machine Learning Command Line Parameters" section, which includes a search bar for assays and a list of parameters for the neural network layers.

Toggle the parameters. There are 4 models for the SVM and 4 models for the neural network

cross entropy- default loss function for binary classification problems. Summarizes the average difference between the actual and predicted probability.

hinge- alternative to cross entropy binary classification developed with SVM models used with support vector machine models

mse-default loss to use for regression problems. calculated as the average of the squared differences between the predicted and actual values

mae-for regression problems. used in cases where there are outliers. average of the absolute difference between actual and predicted values



## 15. Solubility

### Solubility

Predicts log S. Log S greater than -4 is soluble.

Root mean square error of 1.27 on a scale from -4 to 4.

linear regression

Terminal Task DataWarrior

Trash Chromium LibreOffice 6.4 Writer

datawarrior\_linux Anaconda LibreOffice 6.4 Calc

jre1.8.0\_251 Jupyter LibreOffice 6.4 Draw

LibreOffice 6.4.2 Linux... test.xlsx howto.doc

Autodock best.smi

Aldrugdiscover-master3 Chrome Remote Desktop

PCManFM Autodock

Synaptic Bmetoolkit

Biomedical Engineering Chemical Toolkit

Help Report Bug Request Solution

Create a Bioassay or Screen Your Compounds!

Enter the name of your molecule population here

CC(C)NCC(COC1=CC=C(C=C1)OCOC)O

Pharmacophore  
PAINS  
ChEMBL Prediction  
Toxicity Prediction  
Toxicity Map  
Build a SAR model  
Predict solubility  
Fragment Molecules

[-0.9572903]

Submit

Clear Images View Draw SDF 3DSDF Spreadsheet

0%

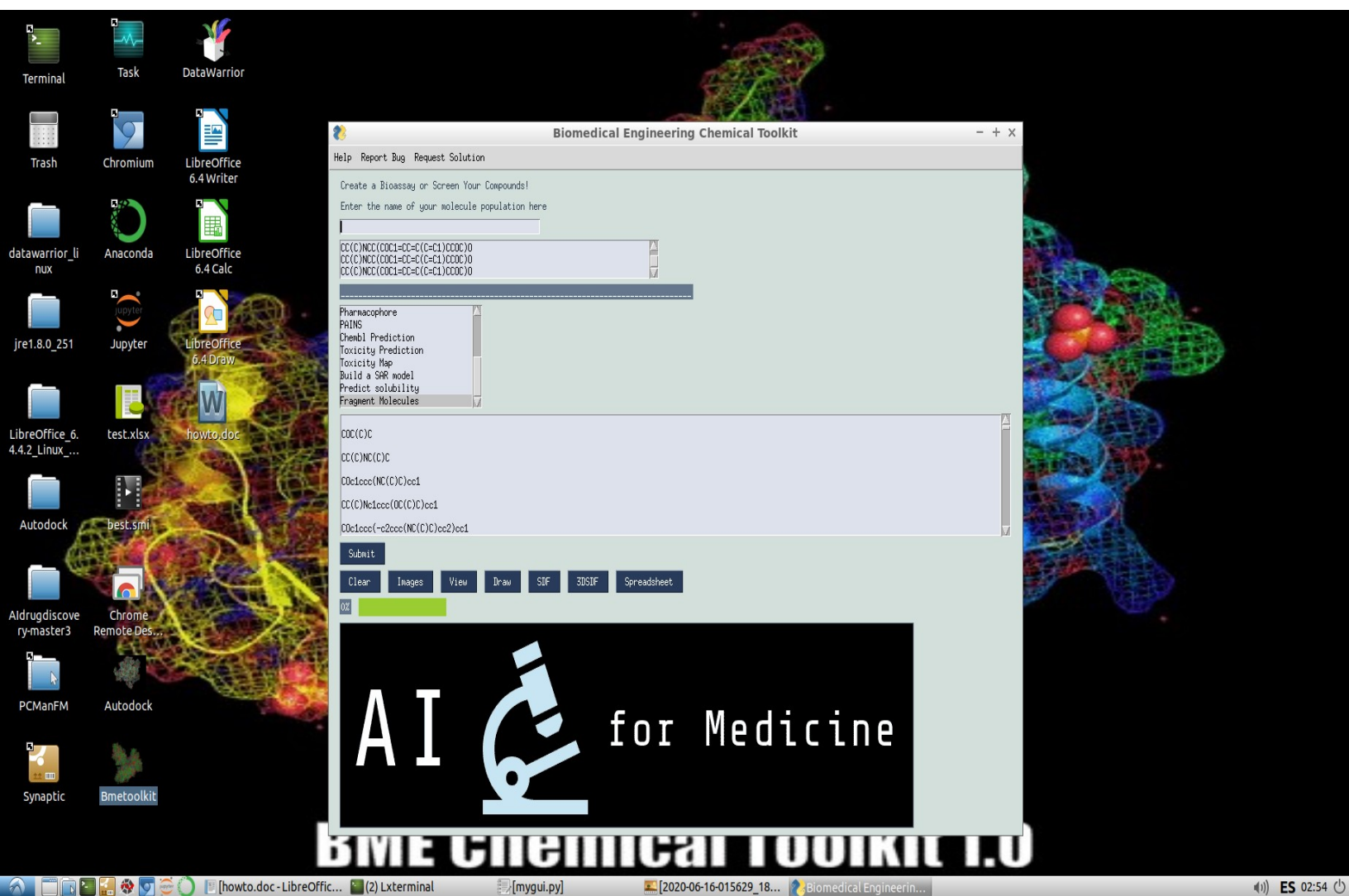
AI for Medicine

BME CHEMICAL TOOLKIT 1.0

[howto.doc - LibreOffice...] (2) Lxterminal [mygui.py] [2020-06-16-015629\_18...] Biomedical Engineerin... ES 02:43

## 16. Fragmenter

Input a list of smiles. These will be recombined into new combinations. When you take the lowest energy ligands from a docking program and recombine these there may be some compounds that bind with lower energy than the original.



## 17. View

Input a list of SMILES and click Generate to view

Smi2Depict: Generate 2D x

Not secure | cdb.ics.uci.edu/cgi-bin/Smi2DepictWeb.py

### ChemDB Chemoinformatics Portal

Smi2Depict: Generate 2D Images from Molecule Files

Applications Tools

Home - Smi2Depict

My Clip Board

#### Depiction Parameters

Sketch

**SMILES**

```
c1ccccc1O  
NC(=O)C(=O)O  
O/C=C/N  
CC=CC(=O)O  
OCC[C@H](O)CC
```

**Options**

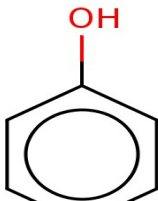
width:800,extraImageSetting:amap,height:400,arrows

**Image Module** ChemAxon Marvin

**Display** ☐ SMILES ☐ Label ☐ Output Only

Generate Images

#### Molecule Images



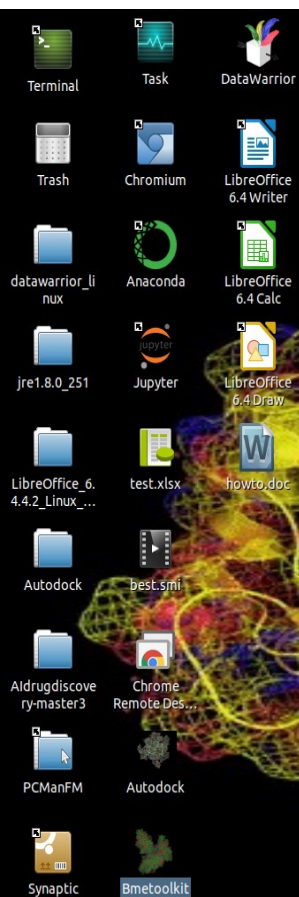
The screenshot shows a web browser window with the URL 'cdb.ics.uci.edu/cgi-bin/Smi2DepictWeb.py'. The page is titled 'ChemDB Chemoinformatics Portal' and 'Smi2Depict: Generate 2D Images from Molecule Files'. It features a navigation bar with 'Home - Smi2Depict' and 'My Clip Board'. The main content area is divided into sections: 'Depiction Parameters' and 'Molecule Images'. In the 'Depiction Parameters' section, there is a text input field for SMILES containing a list of chemical structures: 'c1ccccc1O', 'NC(=O)C(=O)O', 'O/C=C/N', 'CC=CC(=O)O', and 'OCC[C@H](O)CC'. Below this is an 'Options' section with a text input field containing 'width:800,extraImageSetting:amap,height:400,arrows'. There is also an 'Image Module' dropdown menu set to 'ChemAxon Marvin' and a 'Display' section with checkboxes for 'SMILES', 'Label', and 'Output Only'. A 'Generate Images' button is located at the bottom right of the 'Depiction Parameters' section. The 'Molecule Images' section displays a chemical structure of Phenol (C1=CC=CC=C1O) in the center. The browser's address bar shows the URL, and the status bar at the bottom indicates the time as 02:58.

18. Draw

Screenshot of the PubChem Sketcher V2.4 web interface. The browser tabs show "Smi2Depict: Generate 2D" and "PubChem Sketcher V2.4". The address bar displays "pubchem.ncbi.nlm.nih.gov/edit3/index.html". The interface includes a toolbar with various chemical drawing tools (New, Undo, Cut, Style, Del, Copy, Paste, S/A, D/A, S/D, etc.), a periodic table, and buttons for Export (MDL Molfile, Done), Hydrogen (Keep As Is, Help), and Import (Choose File, No file chosen). The main drawing area is empty.

## 19. SDF

Input a list of molecules. Click SDF. The output will be a 2D SDF file called output.sdf



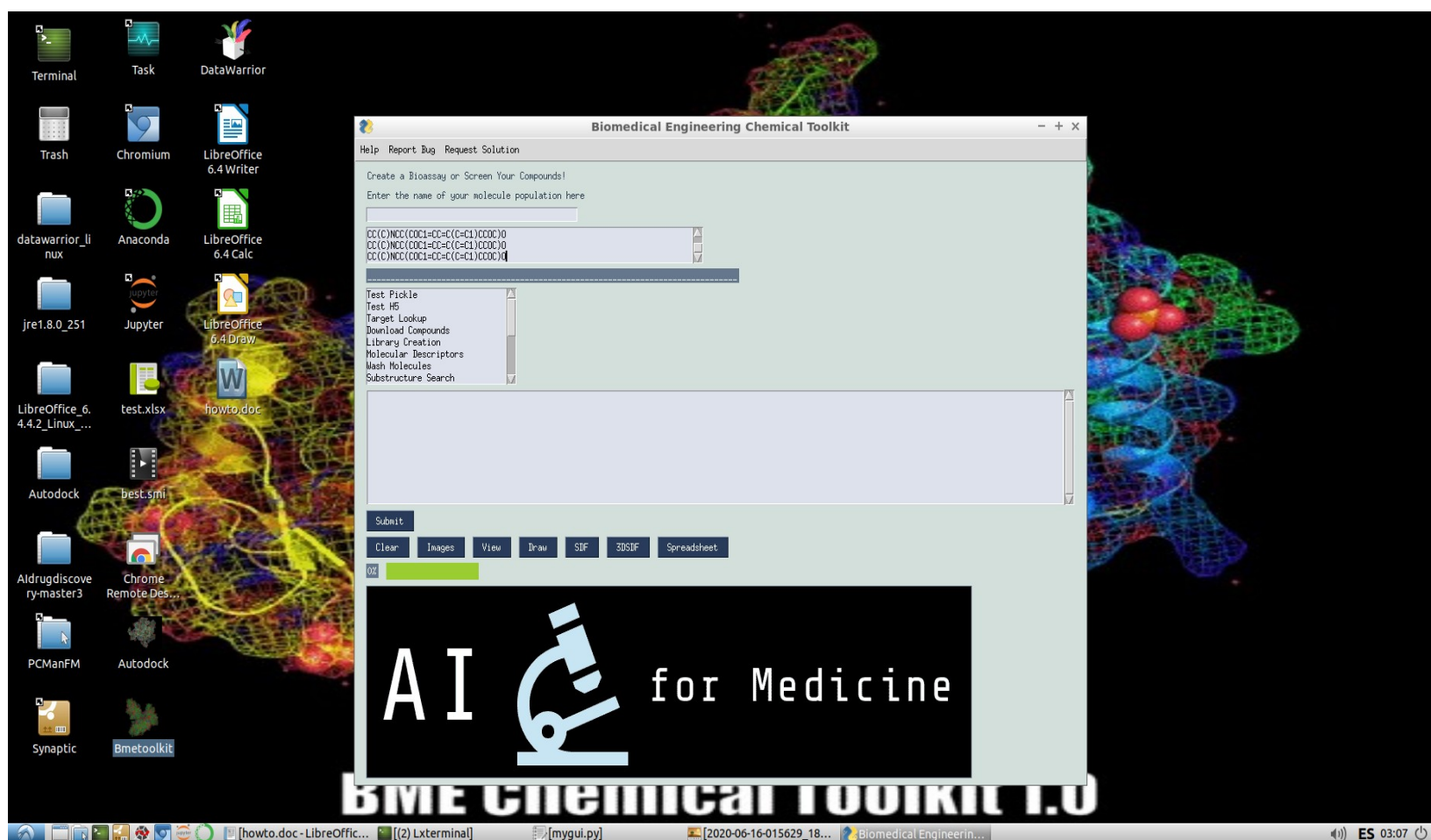
AI  for Medicine

BME Chemical Toolkit 1.0



## 20. 3D SDF

Input a list of SMILES and the output will be a 3D SDF file called output.sdf



## 21. Spreadsheet

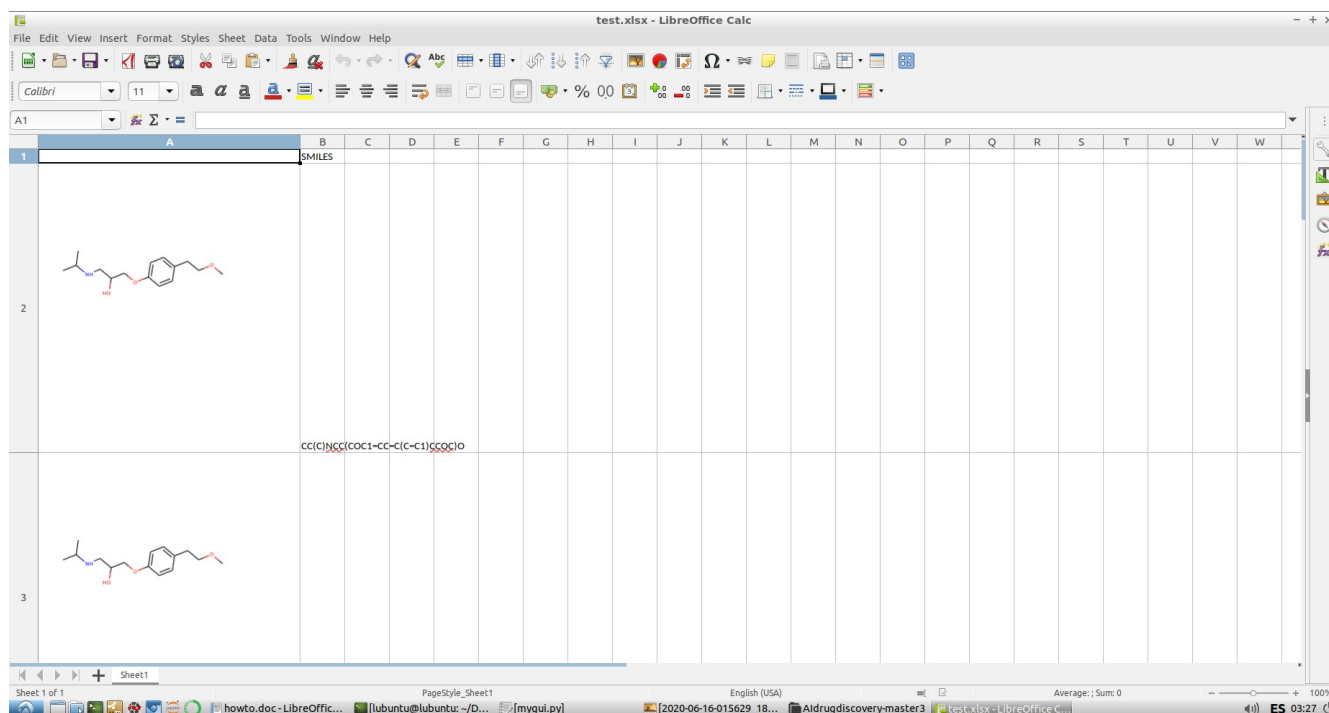
The screenshot displays a Linux desktop environment with a dark background and a colorful molecular structure. The desktop is populated with various application icons, including Terminal, Task, DataWarrior, Trash, Chromium, LibreOffice 6.4 Writer, datawarrior\_linux, Anaconda, LibreOffice 6.4 Calc, jre1.8.0\_251, Jupyter, LibreOffice 6.4 Draw, LibreOffice 6.4.2 Linux, test.xlsx, howto.doc, Autodock, best.smi, Aldrugdiscover-master3, Chrome Remote Desktop, PCManFM, Autodock, Synaptic, and Bmetoolkit.

The central application window is titled "Biomedical Engineering Chemical Toolkit". It features a menu bar with "Help", "Report Bug", and "Request Solution". The main interface includes a text input field for "Enter the name of your molecule population here", a list of chemical structures (SMILES strings), and a list of actions: "Test Pickle", "Test H5", "Target Lookup", "Download Compounds", "Library Creation", "Molecular Descriptors", "Mash Molecules", and "Substructure Search". Below these are buttons for "Submit", "Clear", "Images", "View", "Draw", "SIF", "3DSIF", and "Spreadsheet". A progress bar shows 0% completion.

At the bottom of the window, there is a large black banner with the text "AI for Medicine" and a stylized microscope icon. Below the banner, the text "BME CHEMICAL TOOLKIT 1.0" is visible.

The system tray at the bottom shows the date and time as "2020-06-16-015629\_18..." and the user as "ES 03:07".

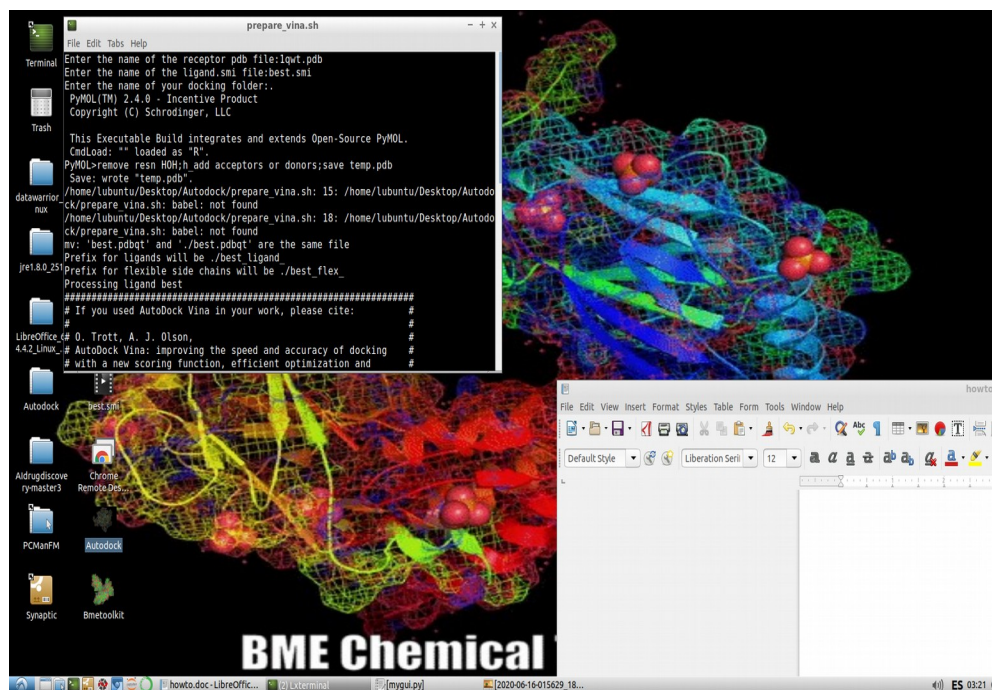
Input a list of SMILES. The output will be a spreadsheet called test.xlsx with pictures and SMILES next to them.



## 22. Prepare Autodock VINA

1. Enter the name of the receptor pdb
2. Enter the name of the ligand.smi file
3. Enter the docking folder as

The output will be a separate directory for each receptor ligand complex with a log file.



There will be an output file called results table.csv in the autodock folder that will contain the docking energy in kJ for each of the compounds. The lower the energy the better the binder.

Additional Pre Installed Packages



Python

pip install for new python packages

Anaconda

Conda install for new packages

Data Warrior for viewing and creating libraries.

Java

Spark

Open Office

Chrome Remote Desktop for sharing the virtualbox with others using the chromium web browser

Google Colab for sharing a jupyter notebook with others. Also gives you the ability to use the GPU

.....The password for lubuntu when installing new packages is ridiculous

Type sudo apt install package name

cd spark-3.0.0-preview2-bin-hadoop2.7

source activate my-rdkit-env

cd bin

sh makemolecules.sh

Follow the instructions. You can edit scaffold.smi  
using <https://pubchem.ncbi.nlm.nih.gov/edit3/index.html>

Use the ? symbol to mark R groups

```
Activities Terminal Mon 21:49
patrickchirdon@system76-pc: ~/Desktop/spark-3.0.0-preview2-bin-hadoop2.7/bin

File Edit View Search Terminal Help
(base) patrickchirdon@system76-pc:~/Desktop/spark-3.0.0-preview2-bin-hadoop2.7/bin$ source activate my-rdkit-env
(my-rdkit-env) patrickchirdon@system76-pc:~/Desktop/spark-3.0.0-preview2-bin-hadoop2.7/bin$ sh makemolecules.sh
Make a compound library from a scaffold and substitute R groups
See scaffold.smi as an example. You can create a scaffold using
https://pubchem.ncbi.nlm.nih.gov/edit3/index.html
and using the ? symbol in place of the R groups you want to substitute
Enter the SMILES file.smi here:scaffold.smi
makemolecules.sh: 8: makemolecules.sh: source: not found
20/06/15 21:49:38 WARN Utils: Your hostname, system76-pc resolves to a loopback address: 127.0.1.1; using 192.168.1.109 instead (on interface wlp3s0)
20/06/15 21:49:38 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.spark.unsafe.Platform (file:/home/patrickchirdon/Desktop/spark-3.0.0-preview2-bin-hadoop2.7/jars/spark-unsafe_2.12-3.0.0-preview2.jar) to constructor java.nio.DirectByteBuffer(long,int)
WARNING: Please consider reporting this to the maintainers of org.apache.spark.unsafe.Platform
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
20/06/15 21:49:39 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
20/06/15 21:49:42 INFO SparkContext: Running Spark version 3.0.0-preview2
20/06/15 21:49:42 INFO ResourceUtils: =====
20/06/15 21:49:42 INFO ResourceUtils: Resources for spark.driver:

20/06/15 21:49:42 INFO ResourceUtils: =====
20/06/15 21:49:42 INFO SparkContext: Submitted application: sample scaffolds
20/06/15 21:49:42 INFO SecurityManager: Changing view acls to: patrickchirdon
20/06/15 21:49:42 INFO SecurityManager: Changing modify acls to: patrickchirdon
20/06/15 21:49:42 INFO SecurityManager: Changing view acls groups to:
20/06/15 21:49:42 INFO SecurityManager: Changing modify acls groups to:
20/06/15 21:49:42 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(patrickchirdon); groups with view permissions: Set(); users with modify permissions: Set(patrickchirdon); groups with modify permissions: Set()
20/06/15 21:49:42 INFO Utils: Successfully started service 'sparkDriver' on port 41113.
20/06/15 21:49:43 INFO SparkEnv: Registering MapOutputTracker
20/06/15 21:49:43 INFO SparkEnv: Registering BlockManagerMaster
20/06/15 21:49:43 INFO BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information
20/06/15 21:49:43 INFO BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up
20/06/15 21:49:43 INFO SparkEnv: Registering BlockManagerMasterHeartbeat
20/06/15 21:49:43 INFO DiskBlockManager: Created local directory at /tmp/blockmgr-e9a48e90-8340-44c7-9ec6-6ccf235f79ed
20/06/15 21:49:43 INFO MemoryStore: MemoryStore started with capacity 4.6 GiB
20/06/15 21:49:43 INFO SparkEnv: Registering OutputCommitCoordinator
20/06/15 21:49:43 INFO Utils: Successfully started service 'SparkUI' on port 4040.
```

Activities LibreOffice Calc Mon 21:55 generated\_molecules.csv - LibreOffice Calc

File Edit View Insert Format Styles Sheet Data Tools Window Help

Libertation Sans 10

A1	B	C	D	E	F	G	H	I	J	K	L	M	N
1	smiles	scaffold	decorations	count									
2	0CC(O)N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C(C)C	16									
3	1CCCCN1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]CCC; 1: [O]	3									
4	2CCN1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	5									
5	3CN1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]CCC; 1: [O]N1CCCCC1	4									
6	4(O)C1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	3									
7	5C(=O)N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	1									
8	6CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	1									
9	7NC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	2									
10	8CCN1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	28									
11	9CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	1									
12	10CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	4									
13	11OC1C(=O)C1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	2									
14	12CN1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	12									
15	13O(=N)C1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	1									
16	14CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	1									
17	15CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	39									
18	16CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	5									
19	17CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	3									
20	18NC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	1									
21	19CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	1									
22	20CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	1									
23	21CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	1									
24	22CN1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	1									
25	23CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	1									
26	24CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	2									
27	25CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	16									
28	26CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	41									
29	27CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	6									
30	28CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	1									
31	29CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	1									
32	30CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	5									
33	31CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	21									
34	32CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	1									
35	33CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	55									
36	34CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	1									
37	35CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	1									
38	36CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	1									
39	37CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	10									
40	38CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	1									
41	39CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	1									
42	40CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	117									
43	41CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	1									
44	42CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	1									
45	43CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	8									
46	44CCOC1N1C(=S)N(C(=O)C1)O1c1cccc1	C2(C1=CC=CC=C1)N(C(N2)=S)*[O]O[H]*[O]	[O] [O]C; 1: [O]N1CCCCC1	1									

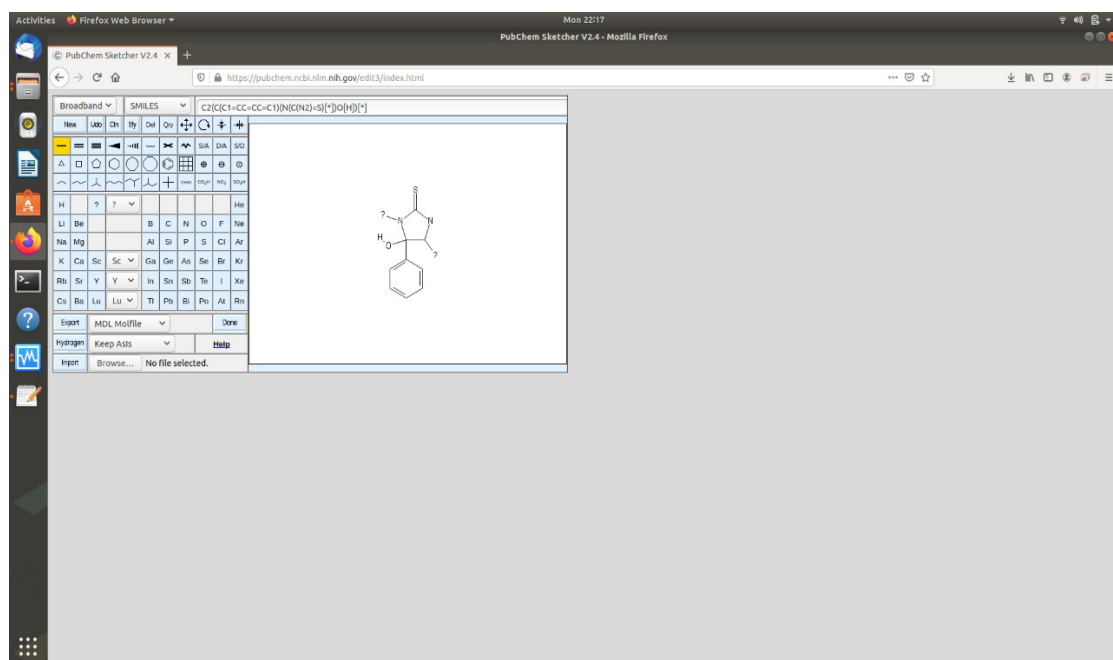
generated\_molecules

Find Find All Formatted Display Match Case

Sheet 1 of 1 Default English (USA) Average: Sum: 0 100%

The output will be a csv file called generated\_molecules.csv that contains the generated SMILES and the scaffold it was generated from

Here is an example scaffold.



Other useful packages pre installed fpocket, openbabel, and pymol. See documentation online

Data Warrior Library View allows you to store and edit libraries

**BME Chemical Toolkit 1.0**

Table	Structure	chemical name	Additional info detail	Synonym
1		Azaguanine-8		
2		Sulfaguanidine	Sulfonamide	
3		Nocodazole		
4		Lymecycline	Tetracycline	
5		Alfadolone acetate		
6		Alfaxalone	General anesthetic	
7		Azapropazone	NSAID	
8		Meptazinol hydrochloride	Opioid	

49 of 495 MB    Selected:    Visible:1585    Total:1585

chemical name  
contains  
CAS number  
contains

Column Name    Data    Value

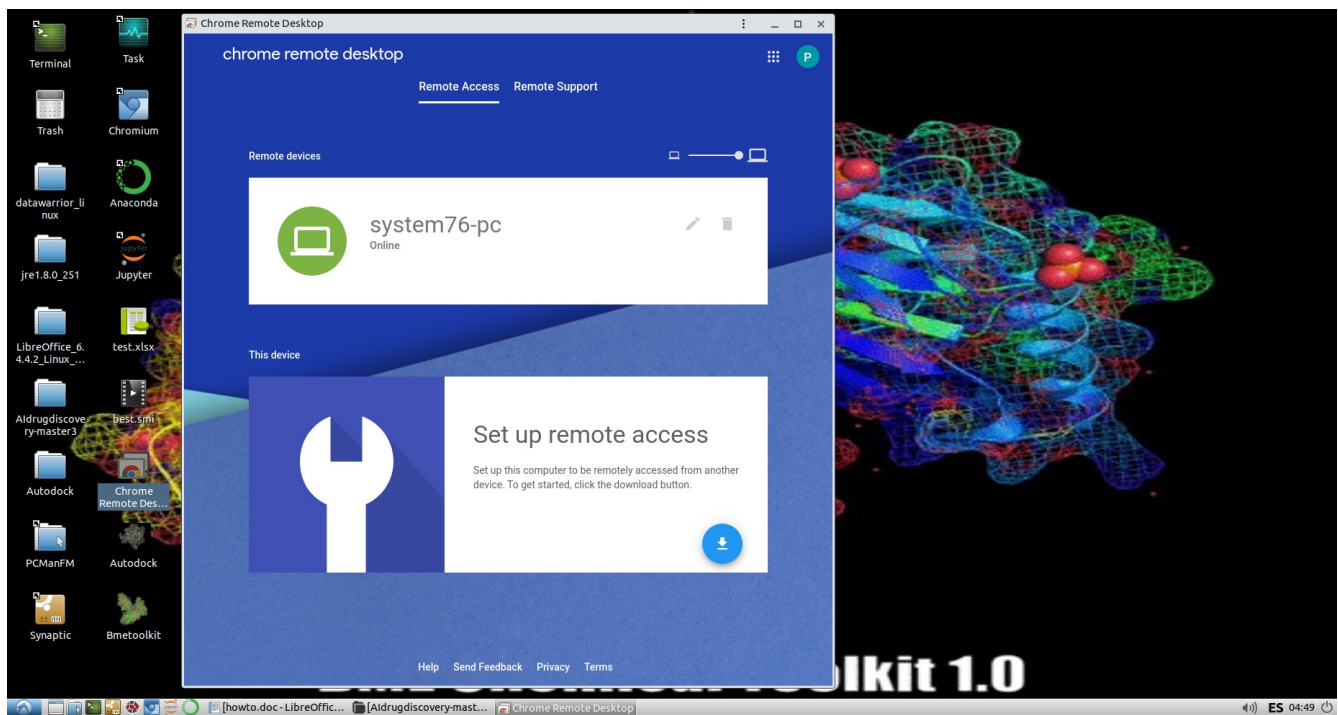
Column Name	Data	Value
CAS number		57-67-0
chemical name		Sulfaguanidine
name IUPAC		4-amino-N(aminomethyl)pyrimidin-2-amine
merck index ref		M1, 13, 8993
literature ref		Contact
patent date		23.05.1950
launched date		
FDA approved code ATC		A07AB03
solubility		DMSO
Additional info type		Chemical Structure

Generate an Evolutionary Library





Chrome Remote Desktop allows you to share your screen remotely with someone through the browser



Open babel

open the shell type babel and you can convert between formats

fpocket

Allows you to find binding sites to use as docking coordinates for autodock

PyMOL

Allows you to view images

