****

**White Paper- BME Chemical Toolkit 1.0**

**Problem**

-Toxicity Testing is expensive and time consuming.

-Anyone who uses chemicals has to be concerned about safety

-There’s a trend toward digitalization of the chemical industry but this is largely not readily available to small firms and labs

**Solution**

**-**We build software that can make predictions on protein target prediction. The software can be used to screen problematic compounds and create new ones before performing expensive experiments, potentially saving time and money. We hope we can be what office suites were for word processing or like what digital streaming services were for the movie industry, a subscription service toolkit for the chemical industry.

**Existing Alternatives**

**-**BioSolveIT

-PyRX

**-** In Silico Medicine

-Benevolent AI

- Big Pharma

-Chemicalize

-Schrodinger

-There is no AI toolkit that uses a GUI, only programming languages that may be inaccessible to people who do not have programming experience

**Key Metrics**

**-**number of compounds screened

-number of subscriptions

-number of downloads

**Cost Structure**

**-**cost of cloud

-potential cost of partnering with a cloud company and any training involved

-advertising-mailchimp

-cost of programmer

**Unique Value Proposition**

**-**low cost toolkit that is easy to use and offered as a subscription service on the cloud. It’s unique compared to previous chemical toolkits in that it uses AI and contains such a large volume of models

**High-Level Concept**

-scalable toolkit that can be billed on a per use basis

**Unfair Advantage**

**-**cloud marketplace on a major cloud partner’s site would make it accessible to a large user base

**-**anyone with a subscription can use the toolkit

-no partnership required

- like a Microsoft Word for drug discovery-- no such AI tool is publicly available

-low cost

-ease of use

**Channels**

**-**journal articles

**-**trade shows/conferences

**-**get a free hour of consulting with purchase of the toolkit

-testimonials

-referrals

-advertising on Linkedin/ Reddit

**Customer Segments**

-academics

-any company that uses chemicals needs to be concerned with safety

-pharmaceuticals

-patent attorneys

**Early Adopters**

**-**academics

-people who read our papers

-tradespeople at trade shows in the chemical industry

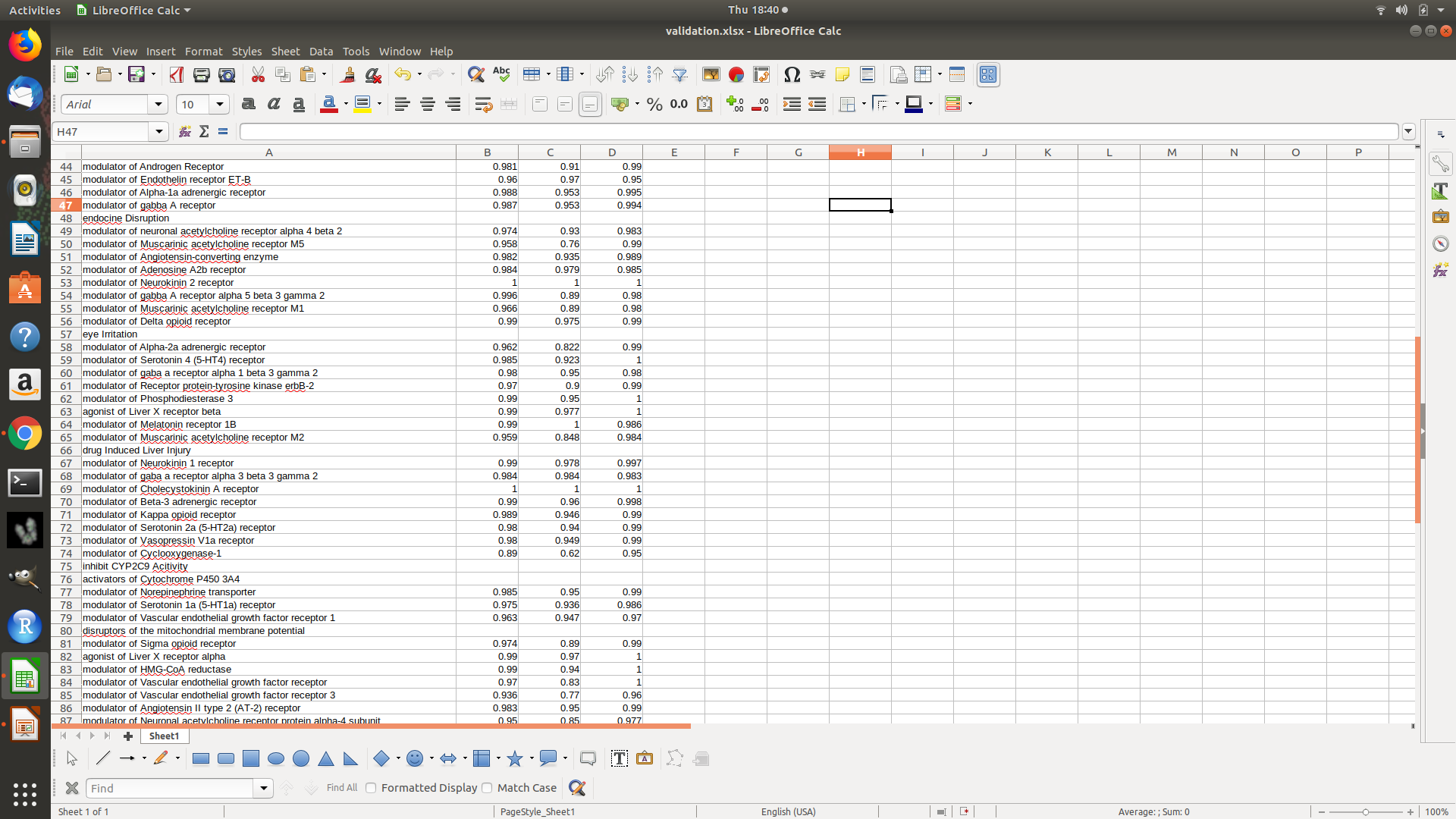
**Revenue Streams**

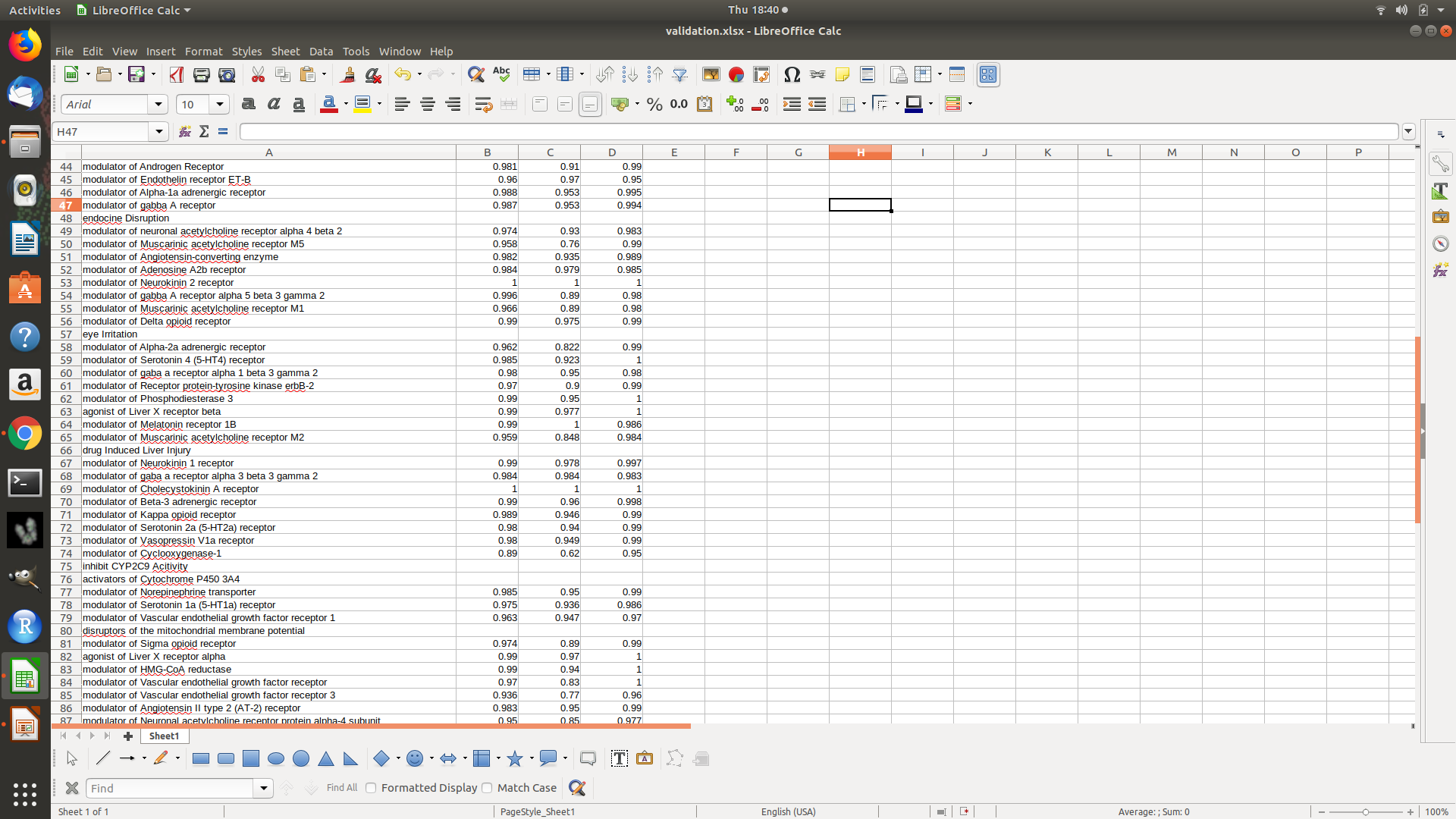
-video courses

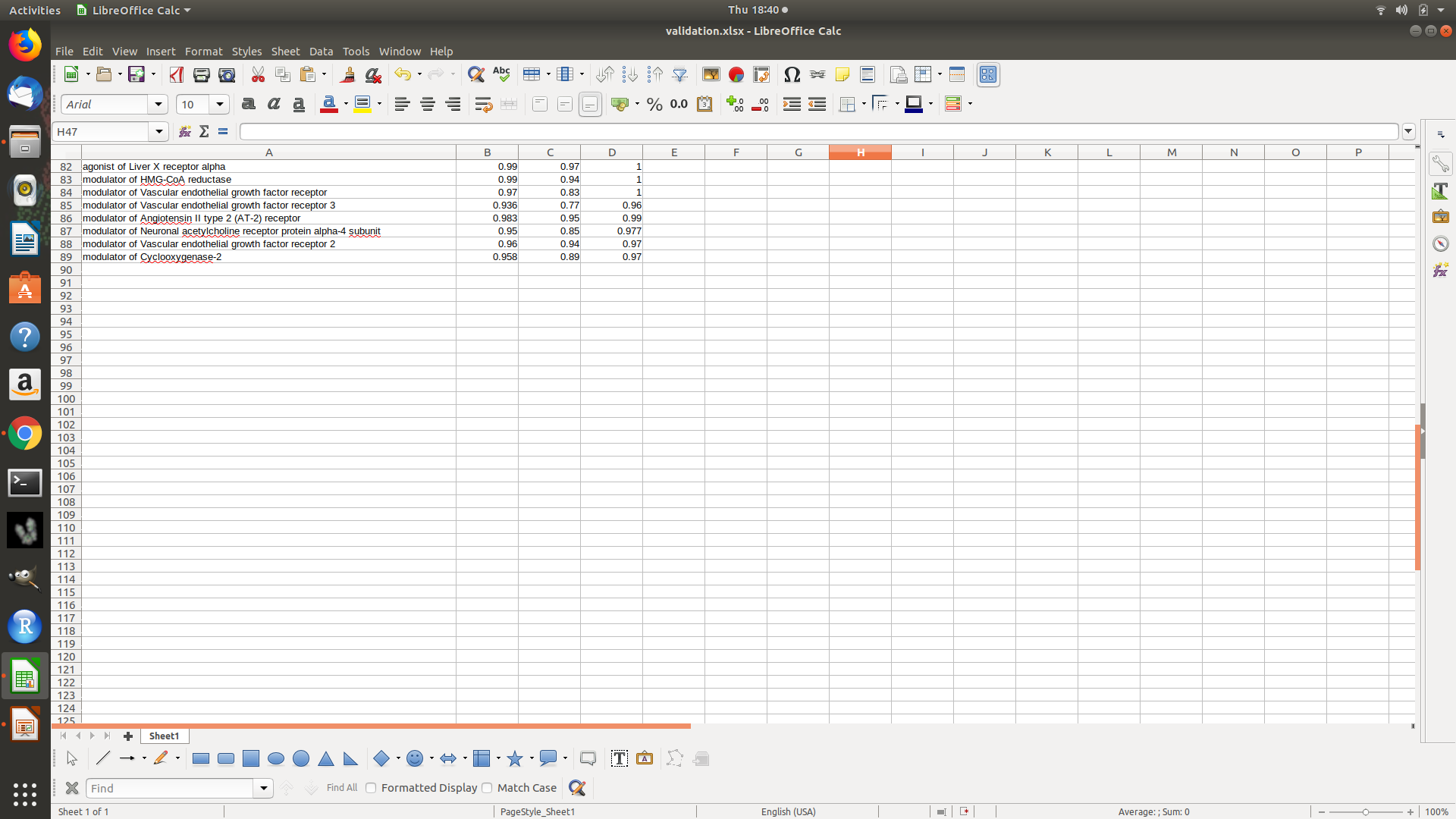
-consulting

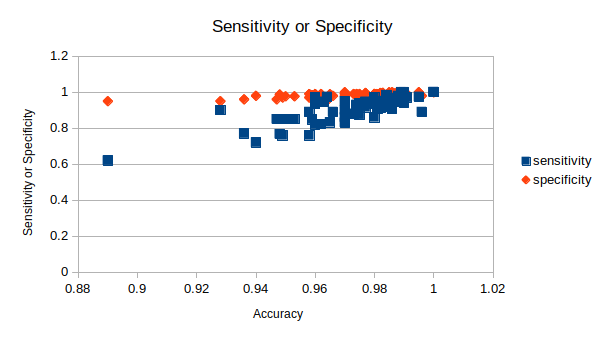
**Revenue Streams**

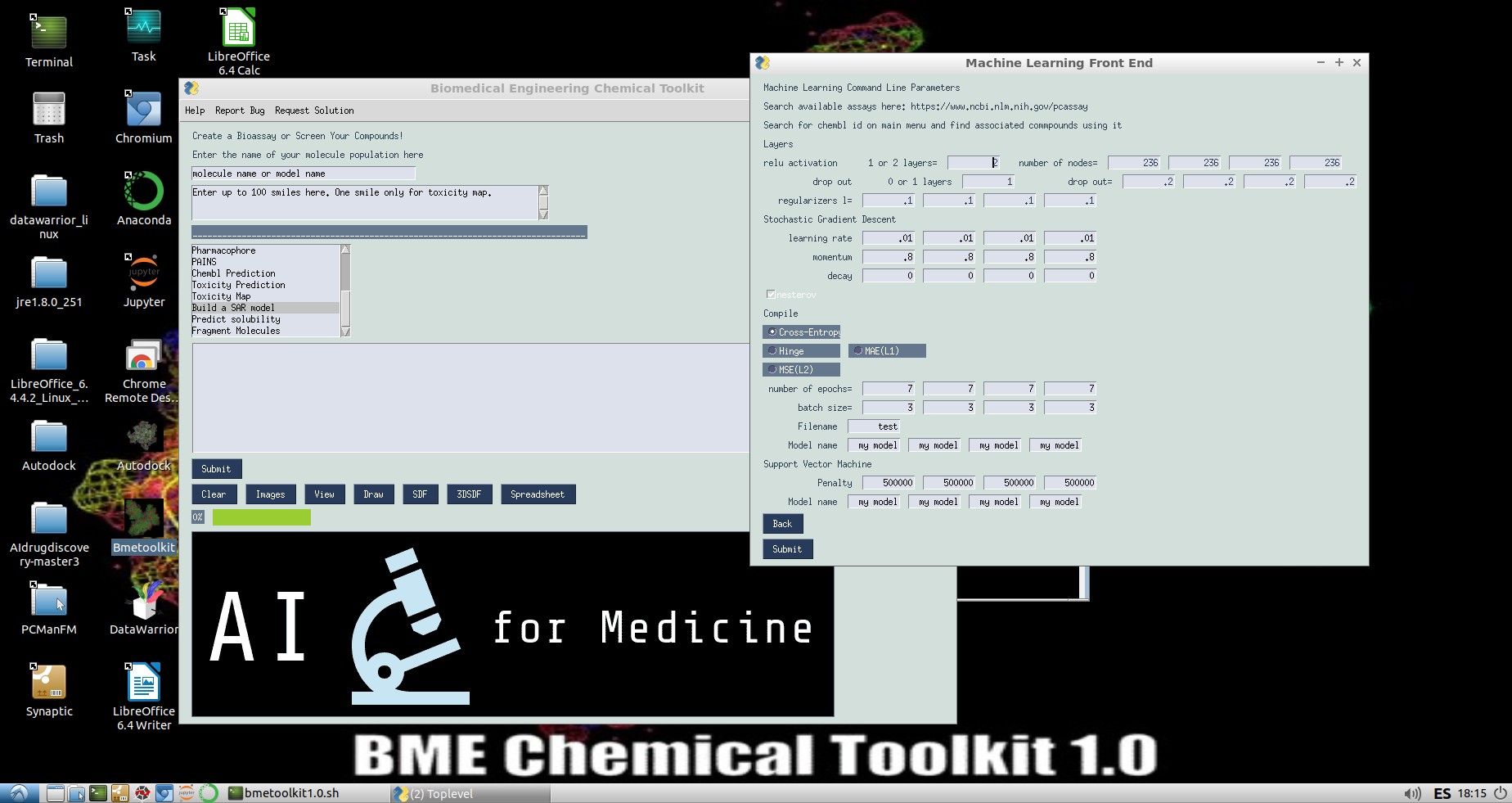
48 Toxicity Targets Available









SAR Model Builder

Chembl Target Prediction

For the 560 targets, the accuracy was the following:

accuracy: 0.8748522411706381, auc: 0.9207217594593, sens: 0.7683270155837822, spec: 0.9246036059820166, prec: 0.8263694343446385, mcc: 0.7071229238195414, f1: 0.796291935419024

Number of unique targets: 560 Ion channel: 5 Kinase: 96 Nuclear receptor: 21 GPCR: 180 Others: 258

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.

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.

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

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

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. 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

7. 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.

8. 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.

9. 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.

10. 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.

11. 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.

12. Toxicity Map

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

13. 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

14. 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.

15. View

Input a list of SMILES and click Generate to view

16. Draw- Input a SMILES, the output will be a molecule image

17. SDF- 2D or 3D

18. Spreadsheet

19. 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.