CavPred 1.0.0

Tutorial

Authors: Anna Basquet, Juana Muñoz, Ester Muñoz
Courses: SBI, PYT 2022/2023

Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Contents

1. Program installation

Once the file *CavPred-1.0.0.tar.gz* is downloaded it should be moved to the desired folder, where it will be untared.

```
$ mv path/to/CavPred-1.0.0.tar.gz path/to/desired_folder
```

Now from within the desired folder, we will untar the file:

```
$ tar -xvf CavPred-1.0.0.tar.gz
```

The files and folders contained in CavPred-1.0.0 are:

```
CavPred-1.0.0
|__ CavPred.egg-info
|    |__ PKG-INFO
|    |__ SOURCES.txt
|    |__ dependency_links.txt
|    |__ requires.txt
|    |__top_level.txt
|
|__ PKG-INFO
|
|__ cavpred
|    |__ __init__.py
|    |__ get_binding.py
|    |__grid
|    |    |__ __init__.py
|    |    |__ ganalysis.py
|    |    |__ gresults.py
|    |__ preprocess
|    |    |__ __init__.py
|    |    |__ hydrogens.py
|    |    |__ protein.py
|    |__ voronoi
|         |__ __init__.py
|         |__ vanalysis.py
|         |__ vresults.py
|
|__ setup.cfg
|
|__ setup.py
```

Finally, moving into the CavPred-1.0.0 folder we are able to install the package.

```
$ cd CavPred-1.0.0
$ python3 setup.py install
```

* Note: the version of the packages used by pyhon3 must be the ones installed by the program in order for it to work.

2. Program execution

- Input: the program requires a path to a pdb file to be provided as input (i.e. 2aox.pdb).
- Output: all the created files will be in a folder named "results_*pdbcode*" (i.e. results_2aox).

From the CavPred-1.0.0 folder where we were standing in the previous section, the command would be the following:

```
$ python3 cavpred/get_binding.py path/to/2aox.pdb
```

In the results folder we will find a series of intermediate files, 2 folders corresponding to the Grid and Voronoi analysis containing intermediate files/visualizations, and the 2 binding result files named *pdbcode_binding_coloured.pdb* and *pdbcode_binding_residues.txt*.

3. Results visualization

In the *pdbcode_binding_coloured.pdb* file we can find a protein in pdb format which has the binding sites signaled through b factors. This allows the visualization of the binding sites using a molecular graphics software like Chimera or PyMOL. In both cases the file must be loaded and then the structure can be coloured by command line:

```
# Chimera
Command: rangecolor bfactor key 0 blue 70 red

# PyMOL
PyMOL>  spectrum b
```

In the case of the *pdbcode_binding_residues.txt*, it provides lists of the residues involved in each binding site with the following format:

```
### Each amino acid is represented in a line under its binding site.
### Columns: position    chain    3letter_code
>> Binding site 1
residue1_position       residue1_chain    residue1_aacode
residue2_position       residue2_chain    residue2_aacode
.
.
.
residueN_position       residueN_chain    residueN_aacode
>> Binding site 2
residue1_position       residue1_chain    residue1_aacode
residue2_position       residue2_chain    residue2_aacode
.
.
.
residueN_position       residueN_chain    residueN_aacode
.
.
.
.
.
.
>> Binding site n
residue1_position       residue1_chain    residue1_aacode
residue2_position       residue2_chain    residue2_aacode
.
.
.
residueN_position       residueN_chain    residueN_aacode
```

4. Example

In this example we will study the binding sites of the protein Histamine Methyltransferase (Primary Variant T105) Complexed with the Acetylcholinesterase Inhibitor and Alzheimer's Disease Drug Tacrine (2aox). Then we will visualize the results using PyMOL.

The first step is to download the program tared file *CavPred-1.0.0.tar.gz* from https://github.com/emunozdc/juannes_predictor/tree/main/dist and the pdb file for the protein (https://files.rcsb.org/download/2AOX.pdb).

To start we will create a directory called "testing" and copy the needed files there, then we will move to that directory to work from there.

```
$ mkdir testing
$ cp CavPred-1.0.0.tar.gz testing/
$ cp 2aox.pdb testing/
$ cd testing
```

* Note: the CavPred-1.0.0.tar.gz and 2aox.pdb files must be provided with its corresponding path.

The next step is to untar the file *CavPred-1.0.0.tar.gz* and install the program.

```
$ tar -xvf CavPred-1.0.0.tar.gz
$ cd CavPred-1.0.0
$ python3 setup.py install
```

Now we run the program with the following command:

```
$ python3 cavpred/get_binding.py ../2aox.pdb
```

* Note: it is important to remember that the pdb file is lying on the previous folder, so it needs the "../" to work properly.

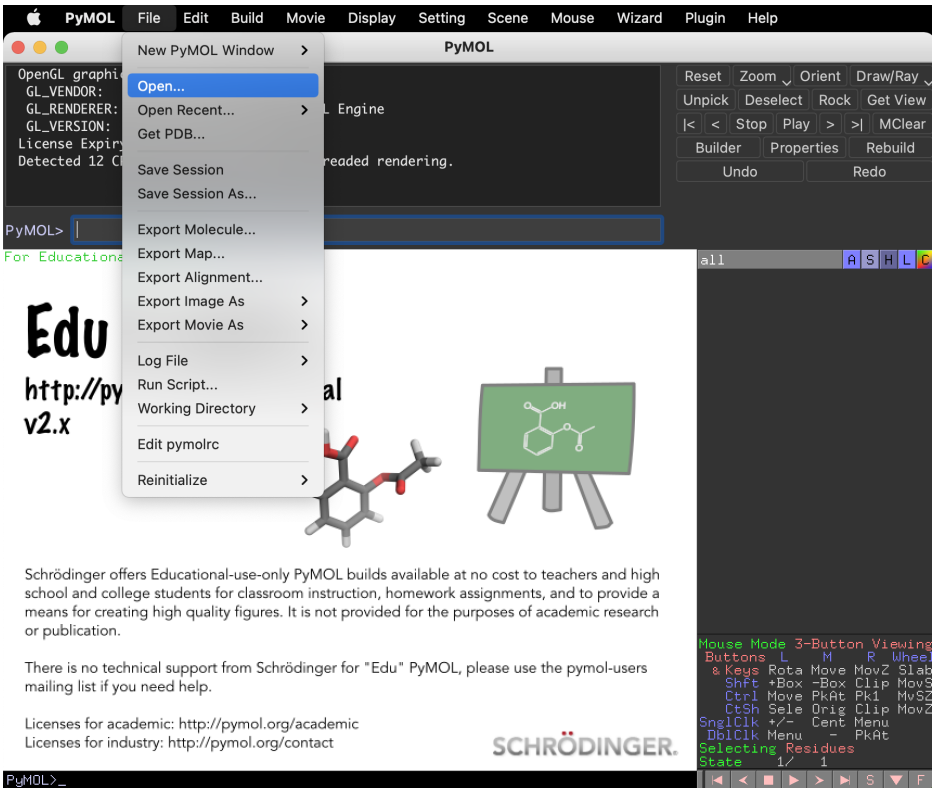
To visualize the binding sites amino acids we can use the following command:

```
$ less results_2aox/2aox_binding_residues.txt
```

This displays the file, and we can move up and down the file. When we are done with the visualization we can exit by pressing "q". The file is the following:

```
### Each amino acid is represented in a line under its binding site.
### Columns: position   chain   3letter_code
>> Binding site 1
127     B       MET
133     B       LEU
134     B       GLN
157     B       PHE
161     B       LEU
>> Binding site 2
127     A       MET
130     A       LYS
132     A       GLU
133     A       LEU
134     A       GLN
>> Binding site 3
39      A       LYS
174     A       SER
175     A       GLY
176     A       SER
177     A       SER
178     A       GLY
181     A       LYS
201     A       SER
202     A       ASP
205     A       THR
206     A       GLN
209     A       ASP
212     A       GLY
213     A       LEU
214     A       LYS
215     A       TYR
216     A       GLU
217     A       CYS
218     A       TYR
219     A       ASP
220     A       LEU
221     A       LEU
269     A       PRO
270     A       GLU
271     A       PHE
272     A       SER
273     A       ALA
282     A       ASN
284     A       THR
results_2aox/2aox_binding_residues.txt
```

Lastly, to visualize the binding sites in PyMOL we will open the application and then open the result pdb file *2aox_binding_coloured.pdb* which is sitting in the results_2aox/ folder.



Once we have loaded the pdb, we must type “spectrum b” in the command line in order to visualize the coloured binding sites.

The result has the potential binding sites coloured in red and the non-binding sites coloured in blue.