CavPred 1.0.0

Examples

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

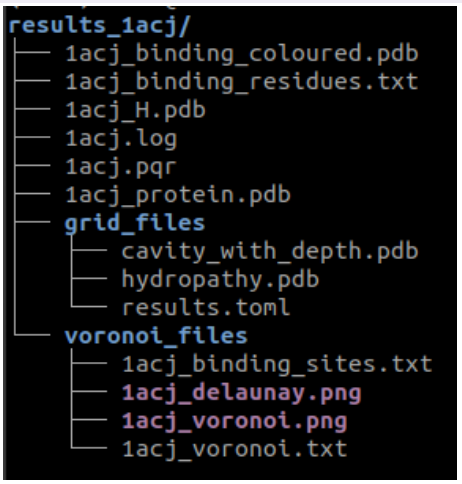
Master in Bioinformatics
for Health Sciences

Contents

Example 1: detailed analysis of results for the protein 1acj

The pdb file for protein 1acj was downloaded from the RCSB PDB database (https://files.rcsb.org/download/1ACJ.pdb). Then the program is run with the following command (in the same folder where it has been installed):

```
python3 CavPred-1.0.0/cavpred/get_binding.py 1acj.pdb
```



After running the program, many files are generated. We will run through each one of them to better understand the pipeline of CavPred and the results it gives.

a. Preprocessing step

The first part of the program consists in modulating the initial pdb file so it can be processed correctly by the prediction algorithms.

1acj_protein.pdb

```
ATOM      1  N   SER A   4     -12.503  89.084  35.130  1.00 66.28           N
ATOM      2  CA  SER A   4     -12.189  87.877  35.866  1.00 63.52           C
ATOM      3  C   SER A   4     -11.066  88.196  36.842  1.00 59.52           C
ATOM      4  O   SER A   4     -11.260  89.101  37.633  1.00 57.70           O
ATOM      5  CB  SER A   4     -12.025  86.720  34.856  1.00 67.07           C
ATOM      6  OG  SER A   4     -13.195  86.792  34.018  1.00 70.02           O
ATOM      7  N   GLU A   5      -9.920  87.515  36.768  1.00 54.99           N
ATOM      8  CA  GLU A   5      -8.763  87.673  37.662  1.00 47.42           C
ATOM      9  C   GLU A   5      -9.082  86.841  38.872  1.00 43.05           C
ATOM     10  O   GLU A   5      -8.332  85.956  39.236  1.00 44.40           O
```

This file is generated after extracting only the protein information from the pdb file, in other words, all lines starting with the word “ATOM”. As a result, if the initial file had a ligand (which is our case), it is no longer present in this one.

1acj_H.pdb

```
ATOM      1  N   SER A   4     -12.503  89.084  35.130  1.00 66.28           N
ATOM      2  CA  SER A   4     -12.189  87.877  35.866  1.00 63.52           C
ATOM      3  C   SER A   4     -11.066  88.196  36.842  1.00 59.52           C
ATOM      4  O   SER A   4     -11.260  89.101  37.633  1.00 57.70           O
ATOM      5  CB  SER A   4     -12.025  86.720  34.856  1.00 67.07           C
ATOM      6  OG  SER A   4     -13.195  86.792  34.018  1.00 70.02           O
ATOM      7  H   SER A   4     -13.404  89.434  35.403  1.00  0.00           H
ATOM      8  HA  SER A   4     -12.990  87.612  36.397  1.00  0.00           H
ATOM      9  HB2 SER A   4     -11.238  86.880  34.317  1.00  0.00           H
ATOM     10  HB3 SER A   4     -12.030  85.880  35.336  1.00  0.00           H
```

This file has Hydrogen molecules added to the pdb file, which is useful for the analysis later on.

1acj.pqr

```
ATOM      1  N   SER A   4     -12.503  89.084  35.130  0.1849 1.8240
ATOM      2  CA  SER A   4     -12.189  87.877  35.866  0.0567 1.9080
ATOM      3  C   SER A   4     -11.066  88.196  36.842  0.6163 1.9080
ATOM      4  O   SER A   4     -11.260  89.101  37.633 -0.5722 1.6612
ATOM      5  CB  SER A   4     -12.025  86.720  34.856  0.2596 1.9080
ATOM      6  OG  SER A   4     -13.195  86.792  34.018 -0.6714 1.7210
ATOM      7  H   SER A   4     -13.404  89.434  35.403  0.1898 0.6000
ATOM      8  HA  SER A   4     -12.990  87.612  36.397  0.0782 1.1000
ATOM      9  HB2 SER A   4     -11.238  86.880  34.317  0.0273 1.3870
ATOM     10  HB3 SER A   4     -12.030  85.880  35.336  0.0273 1.3870
```

This file is the exact same as 1acj_H.pdb but converted to pqr format.

b. Grid step

This step consists in predicting binding sites thanks to geometrical information and some additional information on hydrophobicity.

In total, the grid step found a total number of 16 cavities. The most voluminous and deep one being the cavity named KAI by the program.

1acj_cavity_with_depth.pdb

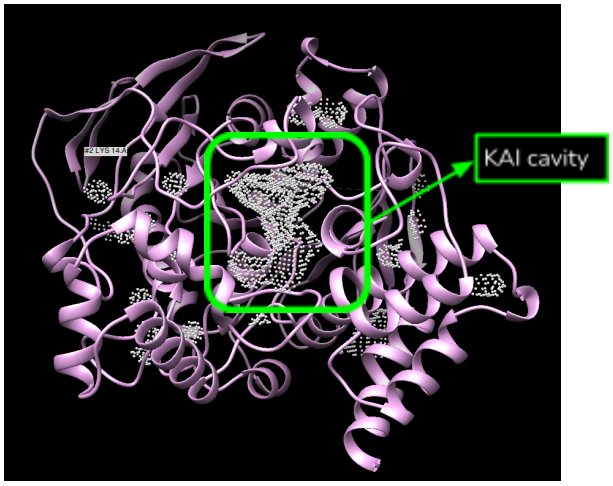
```
ATOM      1  HA  KAA   259     -17.630  68.333  50.112  1.00  0.00
ATOM      2  HA  KAA   259     -17.630  68.933  49.512  1.00  0.00
ATOM      3  HA  KAA   259     -17.030  67.733  50.112  1.00  0.85
ATOM      4  HA  KAA   259     -17.030  67.733  51.312  1.00  1.34
ATOM      5  HA  KAA   259     -17.030  68.333  49.512  1.00  0.85
ATOM      6  H   KAA   259     -17.030  68.333  50.112  1.00  0.60
ATOM      7  HA  KAA   259     -17.030  68.333  50.712  1.00  0.60
ATOM      8  HA  KAA   259     -17.030  68.333  51.312  1.00  0.85
ATOM      9  HA  KAA   259     -17.030  68.333  51.912  1.00  1.34
ATOM     10  HA  KAA   259     -17.030  68.933  48.912  1.00  0.85
```

This file stores the coordinates for the predicted cavities based on geometrical and depth information about the 3D disposition of the amino acids. It only contains information on the cavities, not the actual protein, so, when visualized with any graphic software, such as Chimera, it has to be visualized together with 1acj_protein.pdb.

1acj_hydropathy.pdb

```
ATOM      1  HA  KAA   259     -17.630  68.333  50.112  1.00 -0.12
ATOM      2  HA  KAA   259     -17.630  68.933  49.512  1.00 -0.12
ATOM      3  HA  KAA   259     -17.030  67.733  50.112  1.00 -0.12
ATOM      4  HA  KAA   259     -17.030  67.733  51.312  1.00 -0.12
ATOM      5  HA  KAA   259     -17.030  68.333  49.512  1.00 -0.12
ATOM      6  HA  KAA   259     -17.030  68.333  50.712  1.00 -0.12
ATOM      7  HA  KAA   259     -17.030  68.333  51.312  1.00 -0.12
ATOM      8  HA  KAA   259     -17.030  68.333  51.912  1.00 -0.12
ATOM      9  HA  KAA   259     -17.030  68.933  48.912  1.00 -0.12
ATOM     10  HA  KAA   259     -17.030  68.933  51.312  1.00 -0.12
```

Similarly to the previous file, this one contains additional information on hydropathy for each predicted cavity. Again, it only contains the coordinates of the cavities, so it has to be visualized together with the 1acj_protein.pdb file.



Visualization of hydropathy.pdb (white) and 1acj_protein.pdb (pink) in the Chimera software. The most relevant cavity (KAI) is highlighted in green.

1acj_results.toml

```
# pyKVFinder results

[FILES]
INPUT = "/home/anna/Downloads/results_1acj/1acj_H.pdb"
OUTPUT = "/home/anna/Downloads/results_1acj/grid_files/cavity_with_depth.pd

[PARAMETERS]
STEP = 0.6

[RESULTS.VOLUME]
KAA = 19.01
KAB = 29.16
KAC = 6.7
KAD = 12.31
KAE = 12.31
KAF = 16.85
KAG = 30.89
KAH = 12.31
KAI = 604.15
KAJ = 6.05
```

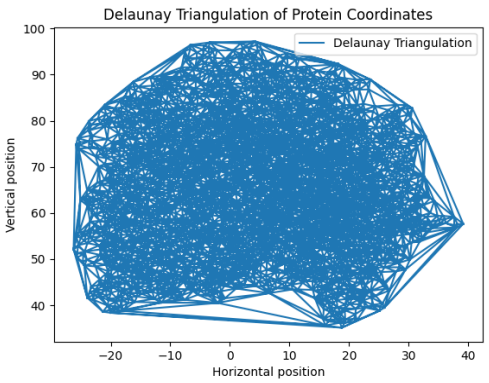
```
[RESULTS.FREQUENCY.KAI.RESIDUES]
ARG = 1
ASN = 1
ASP = 1
GLN = 1
GLU = 2
GLY = 6
HIS = 1
ILE = 2
LEU = 2
MET = 1
PHE = 4
PRO = 1
SER = 6
TRP = 3
TYR = 6
VAL = 1
```

This file stores information about all the cavities, with volume, depth and area information. Additionally, it contains information about which residues are found in each of the cavities.

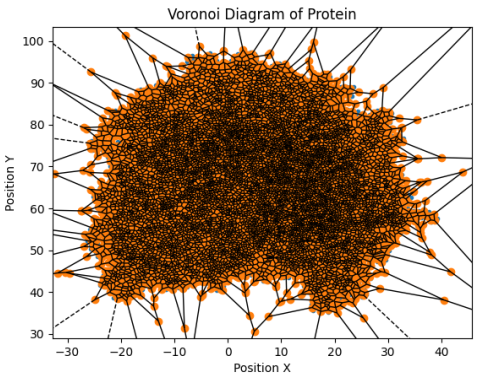
c. Voronoi step

This step found a total of 357 residues to be possible protein pockets.

1acj_delaunay.png



1acj_voronoi.png



These two files contain the graphical results of the Voronoi analysis, which are later used to calculate potential binding sites.

1acj_binding_sites.txt

```
GLU 5
LEU 7
VAL 8
ASN 9
LYS 11
SER 12
GLY 13
LYS 14
VAL 15
MET 16
```

This file contains all residues considered potential binding sites by the Voronoi algorithm.

d. Combination step

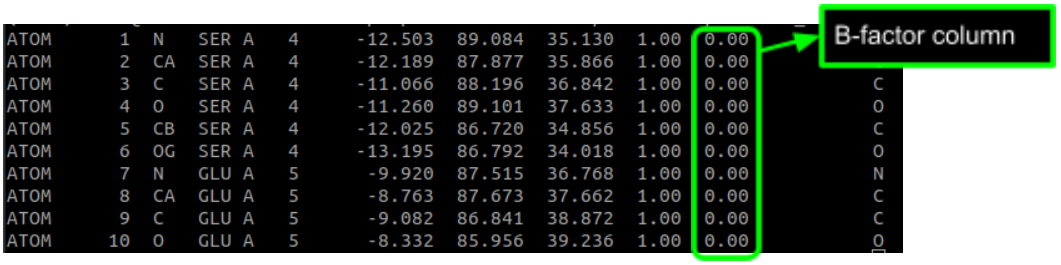
Lastly, the program keeps only the aminoacid positions that are considered potential cavities by both the grid and Voronoi calculations. In the case of 1acj, 60.51% of the amino acids predicted to be in a binding site according to the grid method are also likely to be in a binding site according to the Voronoi method.

1acj_binding_residues.txt

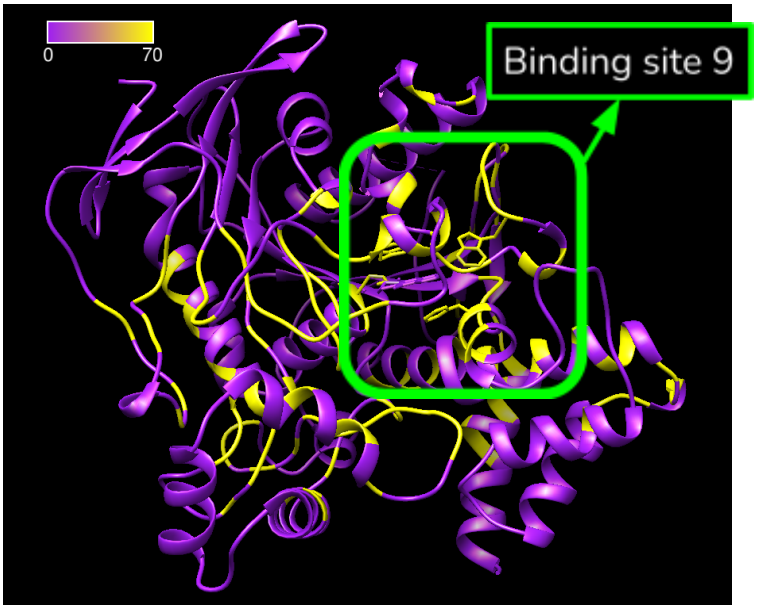
```
### Each amino acid is represented in a line under its binding site.
### Columns: position   chain   3letter_code
>> Binding site 1
34      A       PRO
36      A       ALA
38      A       PRO
51      A       LYS
93      A       ASP
94      A       CYS
95      A       LEU
```

This file contains all the different cavities found in the protein and the list of residues corresponding to them. The previously identified KAI cavity is now binding site 9.

1acj_binding_coloured.pdb



This file contains all the residues in the protein, and those potential to be part of a binding site have b-factor value (column 11) set to 70, while the rest have 0. This way, these residues can be visually differentiated with visualization softwares.

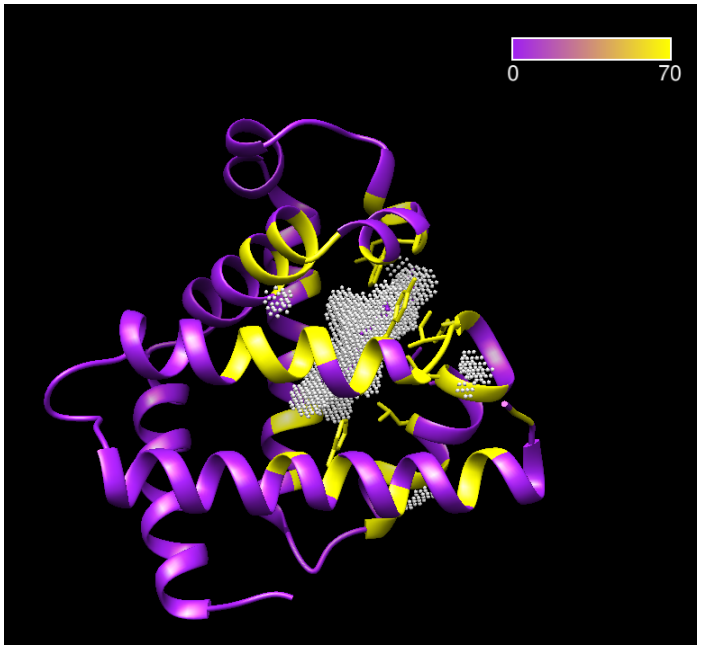


Visualization of1acj_binding_coloured.pdb in the Chimera software. The yellow-colored residues are those which have b-factor 70, so these correspond to potential binding sites. Binding site 9 (highlighted in green), which is the largest, correpsonds to the part where the ligand interacts in the original protein.

Example 2: Results for protein 1mbn

The pdb file for protein 1mbn was downloaded from the RCSB PDB database (https://files.rcsb.org/download/1MBN.pdb). This is a small protein, which helps in viusalizing the results.

```
python3 CavPred-1.0.0/cavpred/get_binding.py 1mbn.pdb
```



Visualization of 1mbn_binding_coloured.pdb together with 1mbn_hydropathy.pdb in the Chimera software. The yellow-colored residues are those which have b-factor 70, corresponding to potential binding sites. As it can be seen, the predicted binding sites correspond to where the ligand binds in the protein.