PyCWMs

Version 1.0

User Manual

PyCWMs is a tool to find conserved water molecules in X-ray protein structures by clustering. For detailed method kindly refer the article.

Installation

PyCWMs script and this tutorial are available at https://github.com/hiteshpatel379/PyCWMs. PyMOL is available at http://www.pymol.org/ or http://sourceforge.net/projects/pymol/.

Dependencies:

Scipy and Numpy python modules installed on Python 2.7 are required to run PyCWMs.

PyCWMs should be installed as a PyMOL plugin.

- Run PyMOL (on some systems you need to be an administrator to install a PyMOL plugin)
- Install the PyCWMs plugin in PyMOL by following the path: `Plugins -> Manage Plugins -> install`
- Restart the PyMOL to use plugin (Be sure **NOT** to be an administrator this time. Otherwise log file and result files are created with administrator rights only.)

Usage:

After installation as plugin, it can be run from Graphical User Interface or Command line or PyMOL's python API.

Using Graphical User Interface:

- Start PyMOL (NOT as an administrator)
- Open PyCWMs plugin from PyMWOL Tcl-Tk GUI: `Plugin` → `PyCWMs`
- Figure 1 shows the snapshot of PyCWMs as PyMOL plugin.

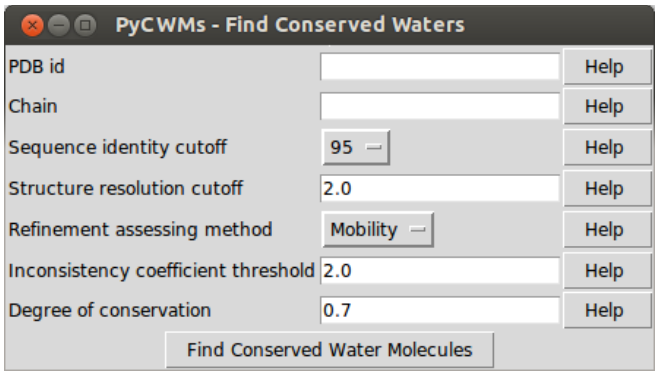

Figure 1. Snapshot of PyCWMs plugin

- Enter required input parameters. *PDB id* and *chain* are mandatory to specify. The rest parameters are optional. Consult table 1 for more details of all input parameters and their default values.
- Change the default values if desired and click on 'Find Conserved Water Moelcules'

Using from command line in PyMOL:

```
pycwms [PDB id , Chain id [, sequence identity cutoff [, resolution cutoff [, refinement assessing method [, inconsistency coefficient threshold [, degree of conservation]]]]]]
```

Using from PyMOL's python API:

```
From pymol import cmd

cmd.pycwms(PDB id , Chain id [, sequence identity cutoff [, resolution cutoff [, refinement assessing method [, inconsistency coefficient threshold [, degree of conservation]]]]])
```

Table 1: Input parameters and default values.

Value	Default value	Explanation
PDB id	--	The PDB id of the protein for which you like to find conserved waters.
Chain id	--	The chain identifier of the protein for which you like to find conserved waters in above mentioned PDB.
sequence identity	95%	All the protein structures, clustered by BlastClust, having sequence identity more than given cutoff will be superimposed to find the conserved water molecules in query protein chain.
resolution cutoff	2.0 Å	All the protein structures to be superimposed will be filtered first according to the structure resolution cutoff. Only structures with better resolution than given cutoff will be used further.
refinement assessing method	Mobility	Choose either 'Mobility' or 'Normalized B-factor' as criteria to assess the refinement quality of crystal structure. Program will filter out the water molecules with bad refinement quality.
inconsistency coefficient threshold	2.0 Å	Any two clusters of water molecules will not be closer than given inconsistency coefficient threshold. Value ranges from 0 to 2.4.
degree of conservation	0.7	Water molecules will be considered CONSERVED if their probability of being conserved is above given cutoff. Value ranges from 0.4 to 1.

Result:

The result file and log file are saved in a directory named ‘PyCWMs_outdir’ under the home directory.

As result, the query protein structure is saved in the PDB file format with only conserved water molecules identified by PyCWMs. And, a PyMOL session is presented showing all conserved water molecules, with H-bonds in between them, with protein and/or the ligand. All conserved water molecules are colored according to their degree of conservation.

A log file ‘pycwms.log’ with all input parameters, program messages, warning and errors is saved. These data are useful to statistically analyze the conserved water molecules in more detail or to manipulate the input parameters.