PyWATER

Version 1.0

User Manual

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

Installation

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

Dependencies:

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

PyWATER 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 PyWATER plugin in PyMOL by following the path: `Plugins -> Manage Plugins -> install`
- Restart 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 the Graphical User Interface or command line or PyMOL's python API.

Using Graphical User Interface:

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

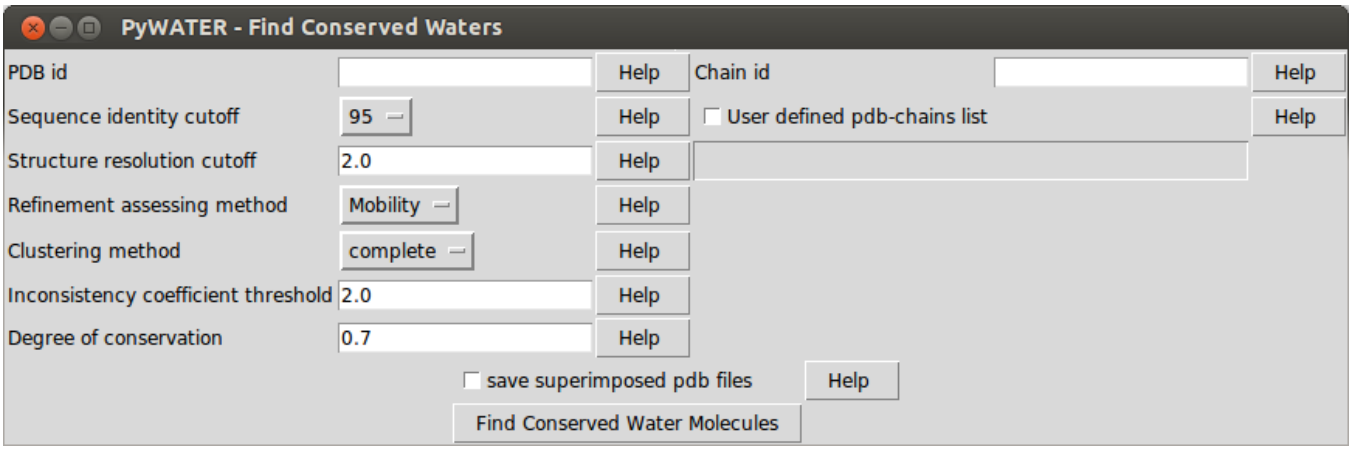

Figure 1. Snapshot of PyWATER plugin

- Enter required input parameters. *PDB* and *chain identifiers* are mandatory for specification. The remaining 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 Molecules'

Using from command line in PyMOL:

```
PyWATER [PDB id , Chain id [, sequence identity cutoff [, resolution cutoff [, refinement assessing method [, user defined proteins list [, linkage method [, inconsistency coefficient threshold [, degree of conservation]]]]]]]
```

Using from PyMOL's python API:

```
From pymol import cmd

cmd.pywater(PDB id , Chain id [, sequence identity cutoff [, resolution cutoff [, refinement assessing method [, user defined proteins list [, linkage method [, inconsistency coefficient threshold [, degree of conservation]]]]]]])
```

Table 1: Input parameters and default values.

Value	Default value	Explanation
PDB id	--	The PDB identifier of the protein for which you want to find conserved waters.
Chain id	--	The chain identifier of the protein for which you want to find conserved water molecules in the above mentioned PDB.
sequence identity	95%	The sequence identity cutoff to find similar proteins clustered by BlastClust.
resolution cutoff	2.0 Å	All the protein structures 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' or 'No refinement' as criteria to assess the refinement quality of crystal structure. Program will filter out the water molecules with bad refinement quality.
user defined proteins list	--	Give a custom list of protein structures to superimpose. Specifying this list will disable 'sequence identity' and 'resolution cutoff' parameters.
linkage method	complete	Linkage method for hierarchical clustering. Choose one from single, complete, average, weighted, median centroid or ward.
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 log is saved in a directory named ‘PyWATER_outdir’ under the home directory.

For each run the result file and optionally all intermediate superimposed pdb files are saved in a directory named ‘pdbid_chainid’ under the ‘PyWATER_outdir’ directory.

As result, the query protein structure is saved in the PDB file format with only conserved water molecules identified by PyWATER. 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 ‘pywater.log’ with all input parameters, program messages, warning and errors is saved.

An additional file is saved showing the degree of conservation of each cluster with atom numbers of water molecules from each superimposed pdb structure. These data are useful to analyze the conserved water molecules in more detail or to manipulate the input parameters. For example, user can analyze the surroundings of a water molecule which is conserved in most proteins but not present in some. Rotameric conformations of side chains of nearby residues may result in displacement of water molecule.