ProBiS H2O

version 0.91

2017

User's Guide
&
TUTORIAL

May 1st, 2017

INSILAB
Support: marko.jukic@ffa.uni-lj.si
Collaborations&support: konc@cmm.ki.si

Background

In order to help in identification of conserved water sites, we developed ProBiS H2O workflow that supports the complete process of data collection, extraction, identification of conserved water locations and result visualisation. ProBiS H2O workflow adopts the available experimental pdb data deposited in online databases or local user experimental data collections, collects similar macromolecular systems, performs local/specific binding site superimposition with the computational speed and accuracy of ProBiS algorithm, collects the experimental water location data reported in the parent macromolecule systems and transposes gathered data to the examined system, its specific chain, binding site or individual water. Water location data is then clustered to identify discrete spaces with high conservation of water molecules and visualised in the context of studied system. ProBiS H2O workflow is a robust, transparent and fast methodology with emphasis on experimental data that can identify trends in water localization intra/inter macromolecule or macromolecule-small molecule interfaces. For further info please refer to the relevant articles by the authors of ProBiS H2O.

Download

ProBiS H2O and this document can be downloaded from: http://insilab.org/probis-h2o/
Python software is available at: http://www.python.org
Scikit-learn Machine Learning in Python library is available at: http://scikit-learn.org/stable/
If you already have a working installation of numpy and scipy, the easiest way to install scikit-learn is using pip/conda:

```
$pip install -U scikit-learn
```

or

```
$conda install scikit-learn
```

PyMOL software is available at: http://www.pymol.org/ or http://sourceforge.net/projects/pymol/

In Ubuntu type the following command in terminal to install all required libraries:

```
$sudo apt-get install python-numpy python-scipy python-sklearn
```

Installation Instructions (Linux)

Dependencies & required libraries:

- **Python 2.7**
- NumPy (>= 1.6.1) python library
- SciPy (>= 0.9) python library
- scikit-learn (>= 1.18) python library
- PyMOL software
- ProBiS algorithm binary (installed during ProBiS H2O setup)

ProBiS H2O is installed as a PyMOL plugin

- Run PyMOL
- Install the plugin in PyMOL by following the path: Plugin > Plugin Manager > Install New Plugin
- Restart PyMOL!

First time Usage

After plugin is installed, it can be run from PyMOL by following Plugin>ProBiS H2O and used through PyMOL Tcl-Tk GUI where plugin window supplements PyMOL default control and display windows as depicted in Figure 1.

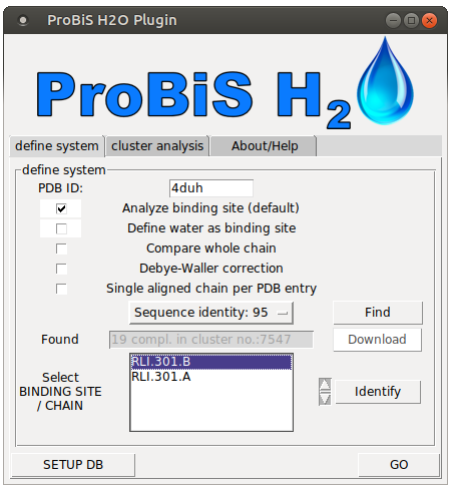

Figure 1: ProBiS H2O plugin main window

In the lower left corner of the ProBiS H2O main window resides SETUP DB button. Before usage ProBiS H2O plugin requires up-to-date information on RCSB sequence clustering and ProBiS algorithm binary file. When SETUP DB button is pressed ProBiS H2O plugin creates /Probis_H2O/ folder in default working directory of PyMOL (default is user home directory) and downloads recent sequence clustering data from RCSB PDB Database. ProBiS H2O plugin also creates /Probis_H2O/.pro/ folder where ProBiS algorithm binary file is downloaded to achieve a final workspace:

```
~/Probis_H2O/.pro/probis
~/Probis_H2O/clusters50.txt
~/Probis_H2O/clusters70.txt
~/Probis_H2O/clusters90.txt
~/Probis_H2O/clusters95.txt
~/Probis_H2O/bc-30.out
~/Probis_H2O/bc-40.out
~/Probis_H2O/bc-50.out
~/Probis_H2O/bc-70.out
~/Probis_H2O/bc-90.out
~/Probis_H2O/bc-95.out
~/Probis_H2O/bc-100.out
```

Final step is to grant probis binary executable rights (+x):

```
$sudo chmod +x ~/Probis_H2O/.pro/probis
```

Additional info on preculstering:

The bc-xx.out files are obtained with blastclust clustering on PDB database website (with parameters -c param_file.txt[-e 0.01] -p T -b T -S 30 for clustering at 30% sequence Identity).

- Basic local alignment search tool, S.F. Altschul, W. Gish, W. Miller, E.W. Myers, & D.J. Lipman (1990) J. Mol. Biol. 215:403-410.

The files clustersXX.txt are obtained by running cd-hit clustering on PDB database website (with parameters
-c 0.50 -n 3 -M 2000; The files clusters70.txt, clusters90.txt and clusters95.txt are obtained with parameter -c 0.70/0.90/0.95 -M 2000).

- Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences, Weizhong Li & Adam Godzik (2006) Bioinformatics, 22:1658-9.

Usage

Default workflow proceeds in following steps (all the data manipulation is performed in /Probis_H2O/ working directory that was configured using SETUP DB button):

/define system tab, Figure 2:

1. User inputs a desired PDB ID as examined system
2. User presses “Find” button to display identified structures in cluster
3. User downloads all relevant structures by pressing “Download” button (*this step is in parentheses because this step is optional if the user already analysed a particular protein cluster and has the structures already downloaded. If it is a first time experiment, the “Download” button downloads the set from the PDB website*)
4. User identifies binding sites or chains of the examined system (1.)
5. User selects desired binding site or chain from a list
6. User proceeds with conserved water identification by pressing GO button

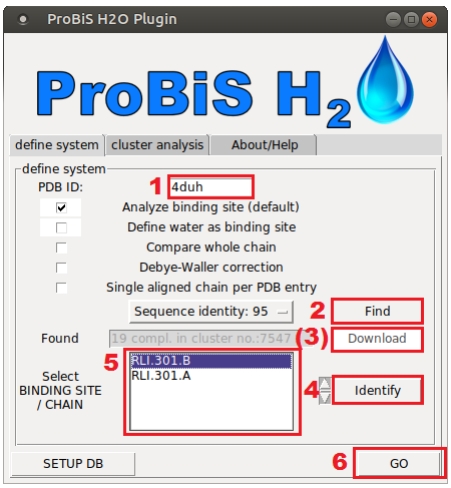

Figure 2: ProBiS H2O plugin define system tab workflow

ProBiS takes advantage of all present CPU cores on the computer and when the calculation is finished plugin switches to cluster analysis tab to present the results as depicted in Figure 2.

Input parameters and defaults:

All the optional input parameters (found in starting GUI define system tab on ProBiS H2O) are defined in Table 1:

Table 1: optional parameters

value	default	explanation
PDB ID:	n/a	The PDB identifier of user query structure
Analyze binding site	checked	Analyze waters in a box 4 Å around selected binding site, if not checked all waters will be used in calculation
Define water as binding site	unchecked	filter - When identifying binding sites, display crystal waters as binding sites (in the case a specific crystal water should be inspected for conservation)

Compare whole chain	unchecked	Compare whole chains instead of specific binding sites
Debye-Waller correction	unchecked	Inspect waters in the context of isotropic displacement – identified clusters will be surrounded by dotted spheres that indicate spatial uncertainty. Relevant if low quality structures are used and user is made aware a problematic structure is used. Further inspection of input data needed.
Only single aligned chain per entity	unchecked	Use only best aligned chain from each pdb entry when performing alignments and superimpositions.
Sequence identity	95	User selects sequence identity cluster too be used for initial data input: 30, 40, 50, 70, 90, 95, 100 or custom cluster that is prepared by the user. Described below.
Select BINDING SITE/CHAIN	n/a	User selects the relevant chain/binding site for examination

Custom cluster preparation:

A file clusters_custom.txt should be prepared and placed in:

```
~/Probis_H2O/clusters_custom.txt
```

File is formatted with arbitrary cluster name, numbering and a pdb identifier with a relevant chain as following:

```
1000   1   1H1S:B
1000   2   1Q5K:B
1000   3   1Q5K:A
1000   4   1XBB:A
.
.
.
```

Results and visualisation:

After the calculation is finished ProBiS H2O switches to the /cluster analysis tab and displays the results as depicted in Figure 3. Typical workflow is as follows (Figure 3):

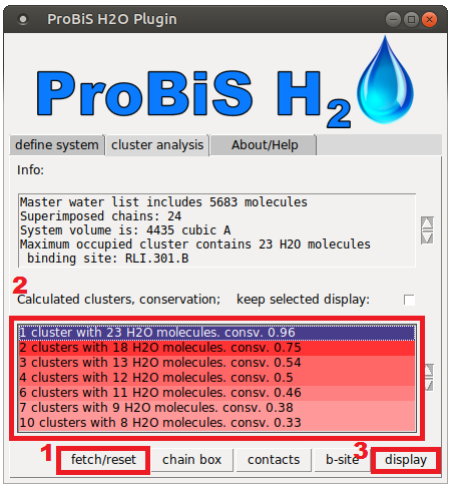

Figure 3: ProBiS H2O plugin cluster analysis tab workflow

1. User displays the studied system in PyMOL display window by clicking “fetch/reset” button. This button can be used for resetting the display for alternative clustering representation (Figure 4).

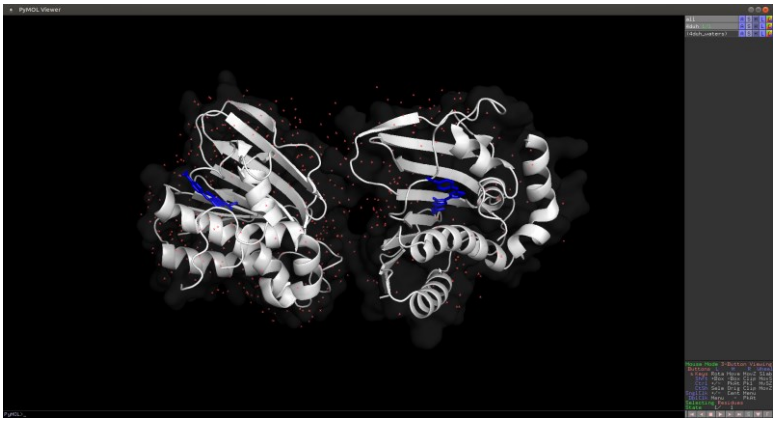

Figure 4: system display

2. User selects the cluster from calculated clusters window
3. User displays the relevant cluster with display button (Figure 5).

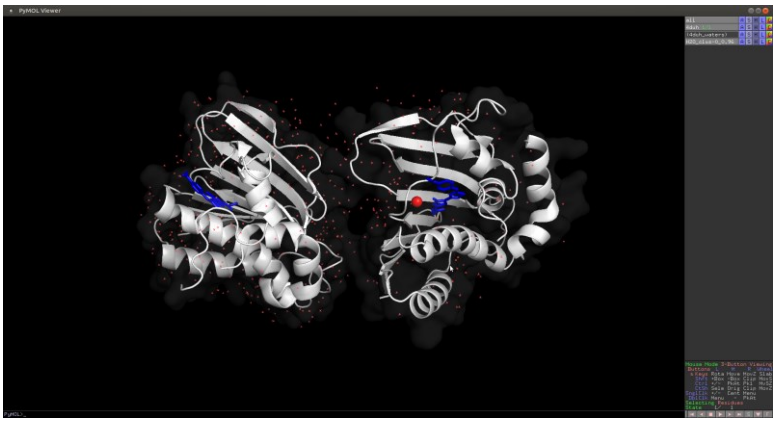

Figure 5: cluster display

Optional settings:

If **b-site** button (Figure 3) is clicked when clusters are displayed, the cluster surroundings will be displayed, colored and labeled for further inspection (Figure 6).

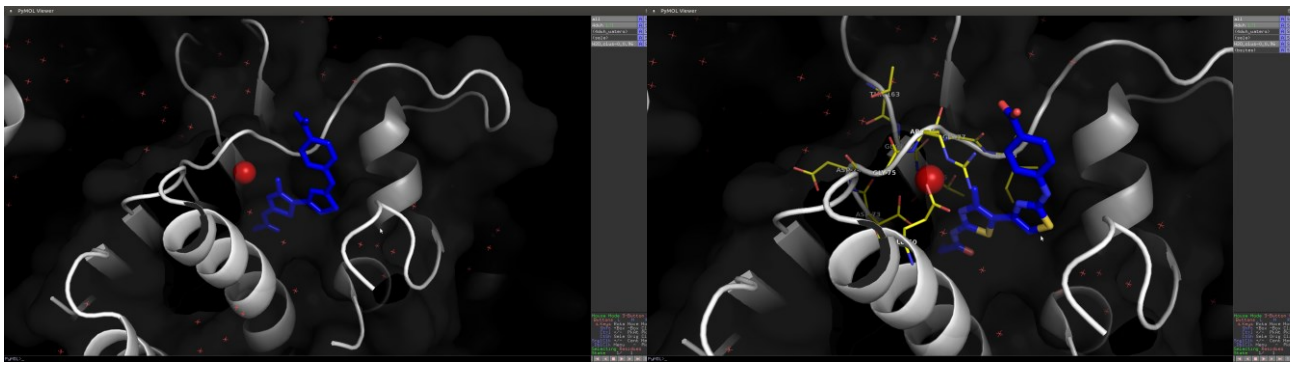

Figure 6: left: without b-site display; right:cluster b-site display

If **contacts** button (Figure 3) is clicked when clusters are displayed, the cluster center measurements to all H-bond donors and H-bond acceptors in vicinity (up to 4 Å) will be displayed, colored in magenta and labeled for further inspection (Figure 7).

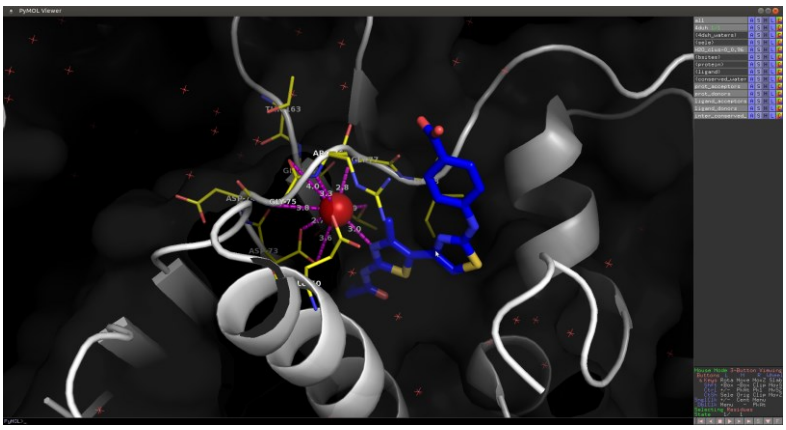

Figure 7: displayed contacts of examined water cluster

If **chain box** button (Figure 3) is clicked, a box (colored red) will be displayed to emphasize the studied system (protein chain), (Figure 8).

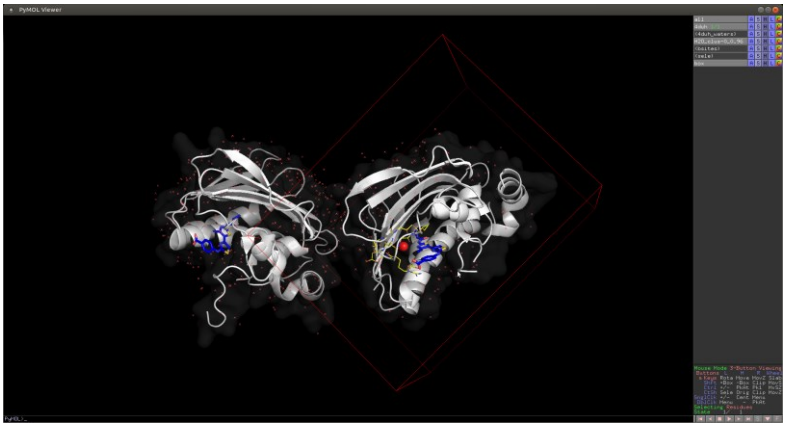

Figure 8: displayed box around the examined chain to emphasize the studied system.

If **keep selected display** (Figure 3) is checked, clusters are not deleted from PyMOL display window and conservation trends can be visualized as depicted in Figure 9. Clusters are colored in gradient form white to red according to their conservation with red color indicating most conserved clusters (Figure 9).

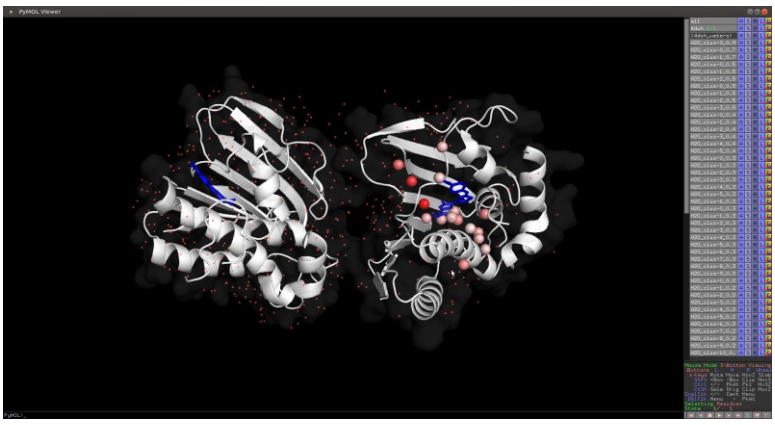

Figure 9: keep selected display for trend visualisation

If user opted for Debye-Waller correction (Figure 1, Figure 2), clusters are concatenated to a single pymol entry and surrounded by dotted spheres that indicate isotropic displacement of individual waters in the cluster. Here on should be mindful for large or disproportional dotted spheres indicating one of the waters used possesses high B-factor (input data should be inspected) (Figure 10).

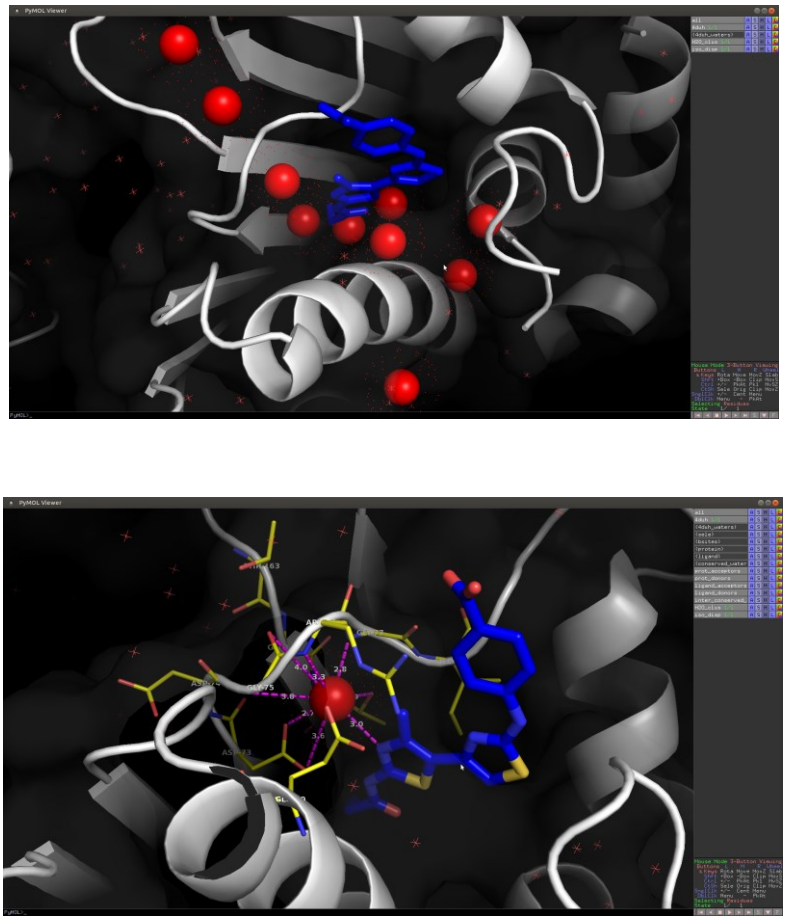

Figure 10: Debye-Waller correction on multiple clusters (top) and single displayed cluster (bottom)

Logging and report file:

ProBiS H2O plugin keeps a report file on each query structure in /Probis_H2O/ folder (Described in first time Usage in this tutorial). Report is named `./report_$pdb query.txt` and is structured as follows:

```
ProBiS H2O REPORT file
-------------------------

Examined complex: 4duh
Whole chain setting used: not
Whole chain selection: / (not used)
Binding site selection: RLI.301.B
Chain selection: B
Used PDB clusters with: Sequence identity: 95 %
Unique structures in identified cluster: ['4WUC', '5L3J', '5MMN', '4DUH', '3G7E', '1AJ6', '4ZVI',
'4WUD', '4KFG', '4PU9', '4PRV', '4XTJ', '1EI1', '4WUB', '4HYP', '1KZN', '5MMO', '4PRX', '5MMP']
```

19 compl. in cluster no.:7511 (sequence identity pre-cluster)

Maximum occupied cluster contains 23 H2O molecules

binding site: RLI.301.B

Master water list includes 5683 waters

System volume is: 4435 cubic A

IDENTIFIED CLUSTERS:

---

32 clusters with 2 H2O molecules. consv. 0.08 [*]

22 clusters with 3 H2O molecules. consv. 0.13 [*]

18 clusters with 4 H2O molecules. consv. 0.17 [**]

16 clusters with 5 H2O molecules. consv. 0.21 [**]

13 clusters with 6 H2O molecules. consv. 0.25 [***]

10 clusters with 7 H2O molecules. consv. 0.29 [***]

10 clusters with 8 H2O molecules. consv. 0.33 [***]

7 clusters with 9 H2O molecules. consv. 0.38 [****]

6 clusters with 10 H2O molecules. consv. 0.42 [****]

6 clusters with 11 H2O molecules. consv. 0.46 [*****]

4 clusters with 12 H2O molecules. consv. 0.5 [*****]

3 clusters with 13 H2O molecules. consv. 0.54 [*****]

2 clusters with 14 H2O molecules. consv. 0.58 [******]

2 clusters with 15 H2O molecules. consv. 0.63 [******]

2 clusters with 16 H2O molecules. consv. 0.67 [*******]

2 clusters with 17 H2O molecules. consv. 0.71 [*******]

2 clusters with 18 H2O molecules. consv. 0.75 [********]

1 cluster with 19 H2O molecules. consv. 0.79 [********]

1 cluster with 20 H2O molecules. consv. 0.83 [********]

1 cluster with 21 H2O molecules. consv. 0.88 [*********]

1 cluster with 22 H2O molecules. consv. 0.92 [*********]

1 cluster with 23 H2O molecules. consv. 0.96 [**********]

Binding site info (name, avg x, y, z, min x, max x, min y, max y, min z, max z; box 4 A around extremes):

['RLI.301.B', 30.654846153846158, 4.82126923076923, 4.8800384615384615, 27.574, 33.514, -1.586, 11.524, 2.077, 10.035]

Examined cluster with 23 H2O molecules

[[27.743, 6.381, 8.997], 0, 2.093375090685332]

[[27.504, 6.347, 9.49], 0, 1.9335846006351467]

[[27.523, 6.211, 9.331], 0, 2.0142432397577976]

.

.
.

Citing:

Please cite the relevant articles by the authors of ProBiS H2O.

Licencing: