ConservedWaterSearch

Release 0.0.1

Jelena Tosovic, Domagoj Fijan, Marko Jukic, Urban Bren

Jun 18, 2022
The conservedwatersearch Python library uses density based clustering approach to detect conserved waters from simulation trajectories. Conserved water molecules can be further classified into 3 distinct conserved water types based on their hydrogen orientation: Fully Conserved Waters (FCW), Half Conserved Waters (HCW) and Weakly conserved waters (WCW) - see the figure below for examples. We support many different density based clustering approaches using standard OPTICS and HDBSCAN procedures as well as multi stage re-clustering approach using either of the two algorithms for very precise (and slow) determination of conserved water molecules.
IMPORTANT LINKS

- Documentation: hosted on Read The Docs
- GitHub repository: source code/contribute code
- Issue tracker: Report issues/ request features
• WaterNetworkAnalysis: prepare trajectories and analyse results for/from conserved water search
Coming soon.
The easiest ways to install `ConservedWaterSearch` is using pip:

```
pip install ConservedWaterSearch
```

Conda builds will be available soon.
The easiest way to use CWS is by calling WaterNetworkAnalysis (WNA) package. However, sometimes users might want to explicitly classify conserved water molecules. A simple python code can be used to classify waters into categories given an array of 3D oxygen coordinates and their related relative hydrogen orientations:

```python
import ConservedWaterSearch.hydrogen_orientation as HO
# load some example
orientations = np.loadtxt("tests/data/conserved_sample_FCW.dat")
# Run classification
res = HO.hydrogen_orientation_analysis(
    orientations,
)
# print the water type
print(res[0][2])
```

For more complex usecases, please refer to the WaterNetworkAnalysis.

### 6.1 ConservedWaterSearch

#### 6.1.1 ConservedWaterSearch package

**Submodules**

**ConservedWaterSearch.hydrogen_orientation module**

ConservedWaterSearch.hydrogen_orientation.__return_normalized_orientation_pair

```python
```

returns normalized orientations.
ConservedWaterSearch.hydrogen_orientation.find_fully_conserved_orientations(orientations: ndarray, pct_size_buffer: float = 0.85, kmeans_ang_cutoff: float = 120, kmeans_inertia_cutoff: float = 0.4, angdiff_cutoff: float = 5, angstd_cutoff: float = 17.0, xi: float = 0.03, njobs: int = 1, verbose: int = 0, debugH: int = 0, plotreach: bool = False) → List

Checks if given oxygen cluster can be considered as a fully conserved water based on hydrogen orientations. Fully conserved water is one which has well defined hydrogen orientations in two distinctive groups (ie strongly hydrogen bonded for both hydrogens). To check if water is conserved, one first checks if k means clustering of hydrogen orientations gives two distinctive clusters with low inertia and required angle between the clusters. Afterwards more rigorous check is carried out with OPTICS clustering where again the spread of orientations and angle is considered.

**Parameters**

orientations (numpy.ndarray) – orientations of hydrogen atoms around studied oxygen cluster

**Returns**

waters – returns list of lists which contains orientations of hydrogen atom orientations (in xyz) wrt to oxygen and string conserved

**Return type**

List

ConservedWaterSearch.hydrogen_orientation.find_half_conserved_orientations(orientations: ndarray, pct_size_buffer: float = 0.85, min_samp_data_size_pct: float = 0.35, angdiff_cutoff: float = 5, angstd_cutoff: float = 17.0, xi: float = 0.01, njobs: int = 1, verbose: int = 0, debugH: int = 0, plotreach: bool = False) → List

Checks if given oxygen cluster can be considered as a half conserved water based on hydrogen orientations. Half conserved water is one which has one well defined hydrogen orientation (ie one strongly hydrogen bonded hydrogen). To check if water is half conserved, one calculates OPTICS clustering of hydrogen orientations. One
then loops over clusters in an attempt to find a hydrogen orientation cluster which is the size of oxygen cluster and weather the angle between that cluster with all other orientations is of right angle and if spread of orientations is sufficiently low.

**Parameters**
- **orientations** (*numpy.ndarray*) – orientations of hydrogen atoms around studied oxygen cluster

**Returns**
- **waters** – returns list of lists which contains orientations of hydrogen atom orientations (in xyz) wrt to oxygen and string conserved type

**Return type**
- List

ConservedWaterSearch.hydrogen_orientation.find_weakly_conserved_orientations(orientations: ndarray, pct_size_buffer: float = 0.85, lower_bound_pct_buffer: float = 0.35, min_samp_data_size_pct: float = 0.15, pct_explained: float = 0.7, angdiff_cutoff: float = 15, angstd_cutoff: float = 20.0, xi: float = 0.01, njobs: int = 1, verbose: int = 0, debugH: int = 0, plotreach: bool = False) → List

Checks if given oxygen cluster can be considered as a weakly conserved water based on hydrogen orientations. weakly conserved water is one which has no well defined hydrogen orientation (ie no strongly hydrogen bonded hydrogen) but still has distinct hydrogen orientational clusters. To check if water is weakly conserved, one calculates OPTICS clustering of hydrogen orientations. One then loops over clusters in an attempt to find a pair of hydrogen orientation clusters which is of the same size and weather the angle between the two clusters is of right angle and if spread of orientations is sufficiently low. Additionally triplets are checked as well. Here we do the same check but we are looking at cluster one vs two other clusters combined.

**Parameters**
- **orientations** (*numpy.ndarray*) – orientations of hydrogen atoms around studied oxygen cluster

**Returns**
- **waters** – returns list of lists which contains orientations of hydrogen atom orientations (in xyz) wrt to oxygen and string conserved type

**Return type**
- List
ConservedWaterSearch.hydrogen_orientation.hydrogen_orientation_analysis(orientations: ndarray, pct_size_buffer: float = 0.85, kmeans_ang_cutoff: float = 120, kmeans_inertia_cutoff: float = 0.4, conserved_angdiff_cutoff: float = 5, conserved_angstd_cutoff: float = 17, min_samp_data_size_pct: float = 0.15, non_con_angdiff_cutoff: float = 15, half_con_angstd_cutoff: float = 17, wigcon_angstd_cutoff: float = 20, weakly_explained: float = 0.7, xiFW: List = [0.03], xiHCW: List = [0.05, 0.01], xiWCW: List = [0.05, 0.001], njobs: int = 1, verbose: int = 0, debugH: int = 0, plotreach: bool = False, which=['FCW', 'HCW', 'WCW'], normalize_orientations=True) → List

High level function that does hydrogen orientation analysis.

Checks if the water cluster belongs into one of the following groups by analyzing hydrogen orientations:

FCW (Fully Conserved Water) - hydrogens are strongly oriented in two directions with angle of 104.5

HCW (Half Conserved Water) - one set (cluster) of hydrogens is oriented in certain directions and
other are spread into different orientations with angle of 104.5

WCW (weakly Conserved Water) - several orientation combinations exist with satisfying angles

If orientations don’t satisfy the criteria for any of the waters, an empty list is returned.

Parameters

- `orientations (np.ndarray)` – array of hydrogen orientations for given oxygen clustering
- `pct_size_buffer (float, optional)` – [description], by default 0.85
- `kmeans_ang_cutoff (float, optional)` – [description], by default 120
- `kmeans_inertia_cutoff (float, optional)` – [description], by default 0.4
- `conserved_angdiff_cutoff (float, optional)` – [description], by default 5
- `conserved_angstd_cutoff (float, optional)` – [description], by default 17
• **min_samp_data_size_pct** *(float, optional)* – [description], by default 0.15
• **noncon_angdiff_cutoff** *(float, optional)* – [description], by default 15
• **noncon_angstd_cutoff** *(float, optional)* – [description], by default 17
• **xi** *(float, optional)* – [description], by default 0.01
• **njobs** *(int, optional)* – [description], by default 1
• **verbose** *(int, optional)* – [description], by default 0
• **debugH** *(int, optional)* – [description], by default 0
• **plotreach** *(bool, optional)* – [description], by default False
• **which** *(list, optional)* – list of strings denoting which waters to search for. Options: any combination of FCW (fully conserved waters), HCW (half conserved waters) and WCW (weakly conserved waters), by default ["FCW", "HCW", "WCW"]

**normalize_orientations** *(bool, optional)* – [description], by default False

**Returns**

- **waters** – returns list with valid clustered hydrogen orientations

**Return type**

List
CHAPTER
SEVEN

INDICES AND TABLES

• genindex
• modindex
• search
C
ConservedWaterSearch.hydrogen_orientation, 11
Symbols

__return_normalized_orientation_pair()  
    (in module ConservedWaterSearch.hydrogen_orientation), 11

C
ConservedWaterSearch.hydrogen_orientation
    module, 11

F
find_fully_conserved_orientations() (in module ConservedWaterSearch.hydrogen_orientation), 11
find_half_conserved_orientations() (in module ConservedWaterSearch.hydrogen_orientation), 12
find_weakly_conserved_orientations()  
    (in module ConservedWaterSearch.hydrogen_orientation), 13

H
hydrogen_orientation_analysis() (in module ConservedWaterSearch.hydrogen_orientation), 13

M
module
    ConservedWaterSearch.hydrogen_orientation, 11