CONTENT

You have successfully installed AQUA-DUCT according to the protocol and want to learn how to use it, and what can you learn from it? This User Guide will get you familiar with the AQUA-DUCT software provided by Tunneling Group. You can run this User Guide step-by-step by using data samples provided at the AQUA-DUCT website in the Downloads section. During the course of this guide you will learn how to run your calculations, prepare visualization and how to perform standard analysis of your results. You will also learn how to generate pockets, hot-spots and energy profiles. Please also check our other tutorials at Tutorial section to be familiar with more advanced features of AQUA-DUCT.

AQUA-DUCT is a tool for tracking small molecules during MD simulation of large macromolecules, identifying of residues important for protein engineering and also, it can help you with drug design. All you need to provide as input files are your **trajectory** and **topology** files. Our sample data are carefully prepared to ensure all the calculations will be clear, fast and easy to reproduce.
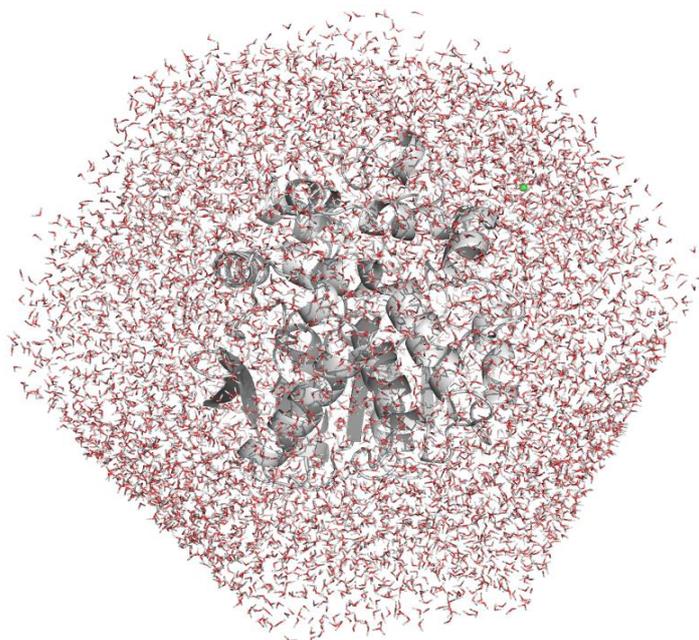
Figure 1. A snapshot of a system prepared for AQUA-DUCT calculations.

Before you start running AQUA-DUCT on your own data, please verify your trajectory files according to what's listed below:

- make sure that your trajectory **files contain information on both macromolecule** (e.g., protein) **and ligand(s)** (for example water molecules, ions, substrates, cosolvents etc.). If you don't have the ligands or solvents, how do you want to track them?
- check if your trajectory have fixed periodic boundaries conditions and the **protein stays in the centre** of the solvent box. Seriously, if you want to see something really useful, just put the protein in the solvent box centre.
- ensure that the all **rotational movements of the protein are reduced**. Unless you want to see some crazy stuff, just align all the structures to some reference and don't let it rotate.
- verify **how often do you save the coordinates to your trajectory file**. We strongly recommend saving coordinates every picosecond (1 ps) for simulations shorter than 100 ns, to have your picture clear. Otherwise you won't be able to see the molecule movements properly.
- we also recommend the combination of 3D-RISM and Placevent algorithm to place water molecules inside the protein's cavities and/or extend the equilibration step of the MD simulation to enable water molecules penetration of the protein.

TIP: For longer trajectories you can save the coordinates less often to save disk space and time, however try to keep ratio 1 ps/100 ns (e.g., for 500 ns do not exceed 5 ps sampling).

If your system looks similar to Figure 1 and you carefully verified your data, you're free to go with AQ!

EASY MODE ANALYSIS

In this User Guide we will be using an example data set of 10 ns MD simulation of mouse soluble epoxide hydrolase structure (msEH, PDB ID: 1cqz). The msEH is a relatively small protein with the active site buried in the protein core. Here we will be tracking all water molecules that visited the active site pocket during the simulation.

To start AQUA-DUCT GUI (Graphical User Interface) simply type in the terminal:

```
valveconfig_run
```
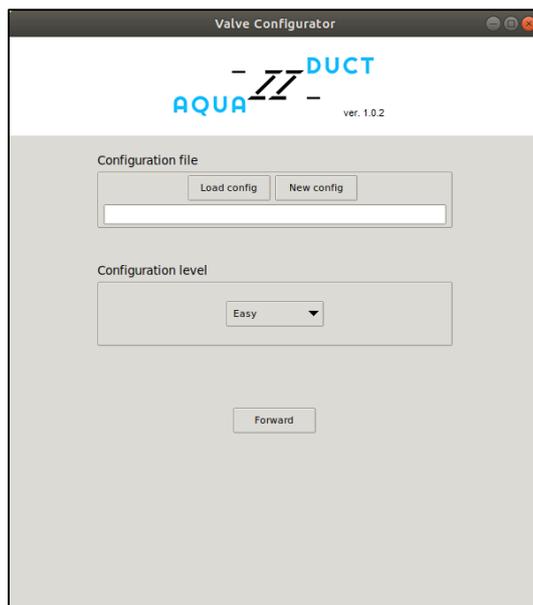
Figure 2. Valve Configurator window.

You will see the Valve Configurator window (Figure 2). Here you can prepare your config file or, if you already have one, simply load it and use for further analysis. Since we do not have any config file, and you want to learn how to prepare your own config file, let's run **Forward**. Or you can click the **New file** button to provide the precise location and name for your brand new config file (Figure 3).
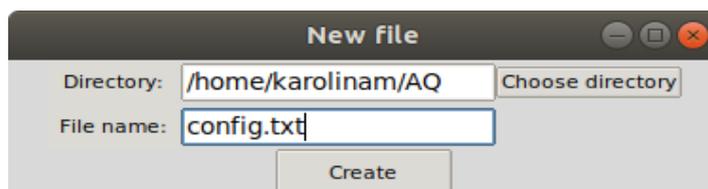
Figure 3. Valve Configurator: New file window.

One more thing before we go - please remember that our GUI can help you, if you'll stuck at some point. If you don't know what does a particular option or window stand for, simply point it (see Figure 4 for pointing). If you look closer at the GUI, you will also find yellow and blue squares - you can point them also to get some more information. In the first part of this User Guide we will use the **Easy mode**, which comprise the essential options for running AQUA-DUCT calculations, even when you don't know your system or AQUA-DUCT itself. The **Normal mode** is more advanced and will provide you some more time- and memory-consuming options to find out more about your system. And the **Expert mode** have all the options of AQUA-DUCT.
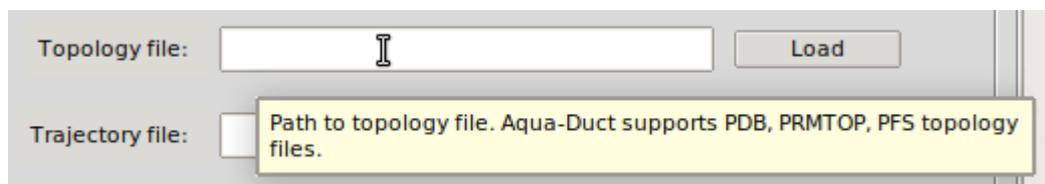
Figure 4. If you don't know what to do with some options, stick'em with the pointy end!
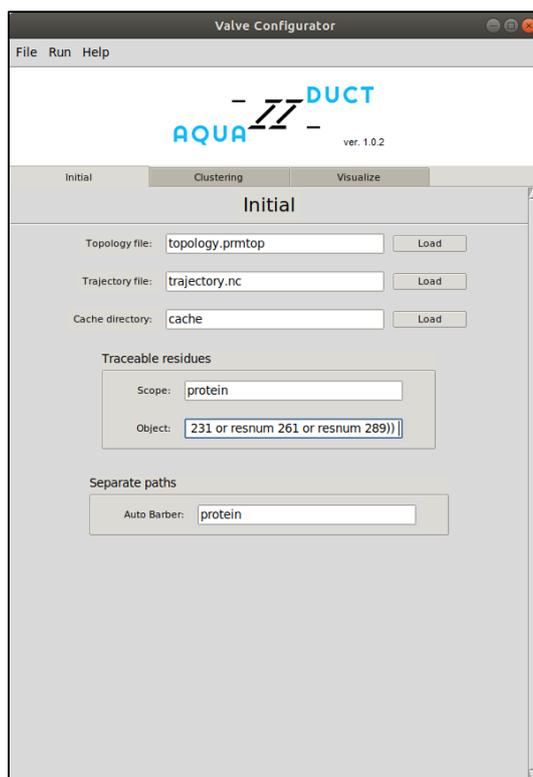
Figure 5. Easy mode - Initial options.

The Easy mode of the Valve Configurator consists of 3 tabs that correspond to the essential steps of AQUA-DUCT calculations: Initial step, Clustering and Visualization.

At the very beginning you need to tell AQUA-DUCT which topology and trajectory files it needs to read for your analysis. AQUA-DUCT supports multiple topology and trajectory files formats e.g., XPDB, MMTF, PRMTOP, TOP, PSF, PDBQT, MOL2, PQR, TRP, CRD, PARM7, GRO, XYZ, PDB for topology and CRD, MDCRD, XPDB, DCD, NC, NetCDF, TRJ, CRDBOX, TRZ, GRO, TRR, PDB, LAMMPS, MMTF, RESTRT, XTC if it comes to trajectories formats. For more information on available files formats please look at MDAnalysis documentation. Note that after you load trajectory and/or topology files, a new cells pop out, so you can add multiple trajectory and/or topology files. Also, you need to set a path to a *cache* directory, where AQUA-DUCT will store some essential files.

TIP: The multiple trajectory and topology files allows you to use advanced modes of analysis e.g., sandwich or consolidator mode.

For the completeness of your calculations you need to provide AQUA-DUCT with information on your system i.e., space where do you want to track your ligands, and the ligands combined with the space of your interests. AQUA-DUCT knows them as *Scope* and *Object*. In this example we wish to study water behaviour inside the enzyme core, however limited to those water molecules which were able to visit the active site. This way we can learn how the active site is connected with the surroundings, which residues are important for transportation, how the protein interior is changing in time etc.

TIP: Depending of your need, you can analyse several different features of your system simultaneously. You have access to information about ligands distribution in time and in space, evolution of your system during simulations, information about residues which can modify properties of your system and also information which can be used for pharmacophore design. All this data are available at once!

Since we are interested in tracking water molecules inside the whole protein, we'll set the *Scope* to `protein`. The definition of *Object* is somehow more complicated and consists of two essential information: the name of tracked ligands (e.g., water, other solvent, ions, or combination of several different small molecules) and the space which they must visit to be tracked, e.g., the active site, binding cavity or some other pocket. In our example we we'll set the *Object* definition as water molecules (resname WAT) within 6 Å sphere of the centre of geometry defined by active site amino acids (residues 99, 147, 231, 261 and 289):

```
(resname WAT) and (sphzone 6.0 (resnum 99 or resnum 147 or resnum 231
or resnum 261 or resnum 289))
```

TIP: AQUA-DUCT can handle really complicated Object *definition like a ligand that visited couple of separated pockets. Although the above example of* Object *definition would be most often use, you can check the [MDAnalysis Documentation](#) for more sophisticated selections.*

TIP: Please ensure that the Object *you are setting is inside the* Scope *of interests and is not sticking out. In such case you could have experience weird stuff during path visualization and your results might be wrong!*

All right, so you told AQUA-DUCT what are the ligands you want to track, and where do you want to track them. Now it's time to talk about paths. A path in AQUA-DUCT is the main source of information. A path shows where the molecule of interests have entered the *Scope*, visited the *Object* (left the *Object* and visit it again), and left the *Scope*. Sometimes the path can have less than those three parts and it's also OK. If the molecule you are tracking have entered and left the protein, the beginning and the end of each path are marked by inlets. But sometimes the path could end inside the protein, i.e. inside the Scope. Some paths may even begin inside the Scope. Such paths are then marked as N. We will come back to them later, during analysis. To make your picture clear, you want the inlets to stay on the protein's surface. This way they will mark the tunnels' entrances. So here comes the AutoBarber!

For the clarity of the results, AQUA-DUCT uses the ConvexHull approximation to describe the protein's surface. It's good for the calculations, they are fast, but not very good for accuracy, if it comes to analysing the paths. The inlets usually flows beyond the protein's surface and are widely dispersed. AutoBarber cuts the paths to fit them into the protein's surface. You can tell the barber, how short you want those paths to be. Typically when you set AutoBarber to `protein`, you'll get longer paths than when you set it to `backbone`. And if you want the shortest paths, you should try `name CA`. (For more information about using the AutoBarber, please visit our [Tips&Tricks - Trimming paths](#)).
Now let's move on to the next tab.
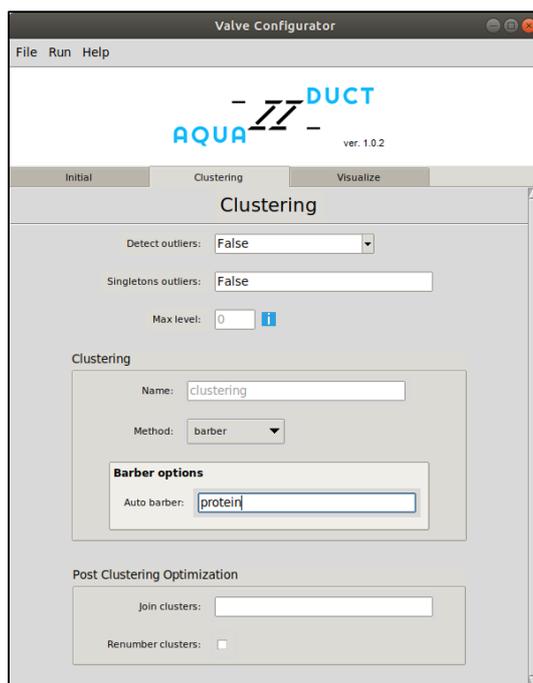
Figure 6. Easy mode - inlets clustering.

Here is the most tricky part of the AQUA-DUCT calculations - clustering the inlets. We want to group together those water molecules which enters or leave the protein interior by the same tunnel entrance. If the protein has more tunnels we can separate them, and see how often each of them is used.

Since you are probably doing this for the first time of your life (and trust me, it will not be the last one!), we strongly advise you to use the `barber` method at the very beginning, set the **Auto barber** option to `protein`, and let the other options default. This method will give you the first glance of your system, the water paths network and could also give you some hints on how to proceed with some additional clustering methods.

The AutoBarber procedure works here in a similar manner that it is during in the previous step, but that doesn't mean that you need to set the same value for this option. You can mix the AutoBarber parameters however you like - do not hesitate and check Tips&Tricks - Trimming paths if you haven't done that before!
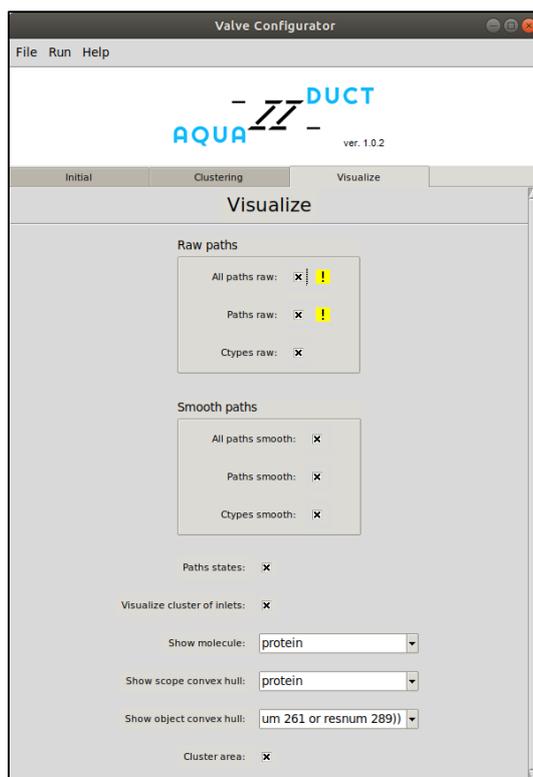
Figure 7. Easy mode - Visualization.

The last step before we finally run AQUA-DUCT is the visualization. Here we define specifically what aspects of calculations we want to visualize. Since the molecule, *Scope*, *Object* and clusters are quite important, we will select them. We would like also to have a quick peek on how the paths and inlets look like. So let's choose all of the available options. Please note that we changed the *Object* definition to (`resname *`) and (sphzone 6.0 (resnum 99 or resnum 147 or resnum 231 or resnum 261 or resnum 289)) for the visualization clarity.

TIP: The visualisation file (PyMOL session) can be really large, when you want to include all the available information (even over 1GB). Therefore, in the first round of analysis, you need to optimise clustering procedure, you will need only the most crucial items. We also advise not to include the raw paths option if you think you could have a lot of paths. However, the analysis of the raw paths could be necessary, when you are not sure if neighbouring clusters mark different tunnels' exits, or they can be merged.

Now let's save the config file. Select *File → Save as* and set the directory and name of your configuration .txt file.
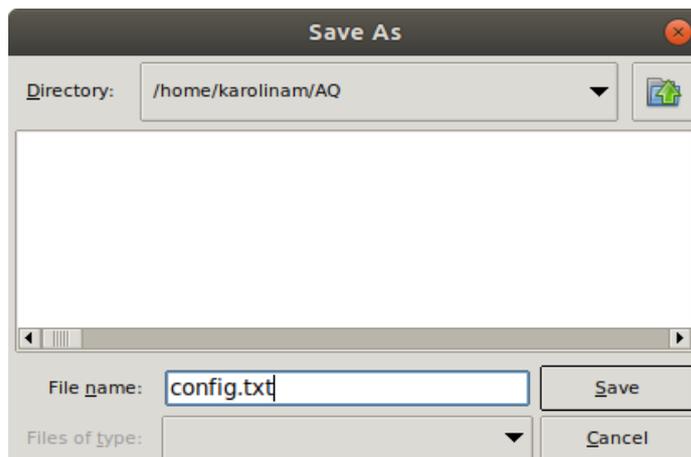
Figure 8. Saving config file.

You can run AQUA-DUCT now! Select *Run → Valve*. At the end of the window, you will find *Run* button. Click on it and begin the calculations!
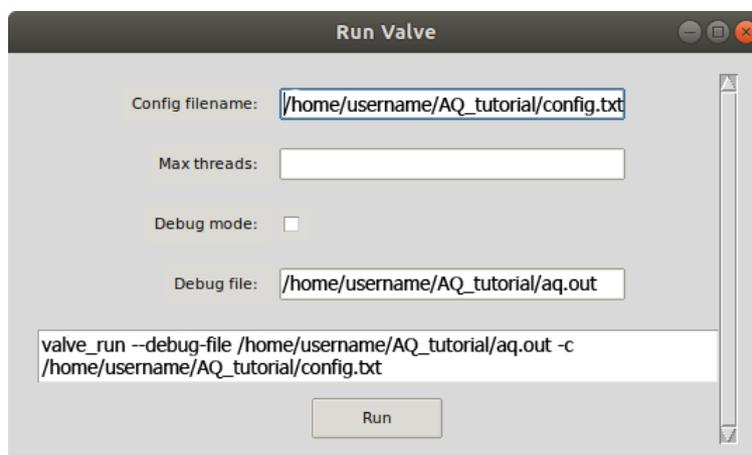
Figure 9. Running AQUA-DUCT by GUI. Please note, that the names of the config and log files should consist of the direct path to the file.

After your AQUA-DUCT calculations will finish (that could take a while) you'll get some files which stores results from each stage of analysis:

- dump files (`1_traceable_residues_data.dump`, `2_raw_paths_data.dump`, `3_separate_paths_data.dump` and `4_inlets_clustering.dump`) containing the essential data for further AQUA-DUCT analysis steps,
- `5_analysis_results.txt` and `5_analysis_results.txt.csv` files that contains all important data from the analysis,
- `6_visualize_results.py` and `6_analysis_results.tar.gz` files to prepare your visualization.

The most important for us is now the file for visualization. AQUA-DUCT uses PyMOL (so make sure you have it installed!) for visualizing the results. The easiest way to preview your results is to use the `6_visualize_results.py` script.

AQUA-DUCT by TTG

The script for visualization is quite smart and allows you to decide which items you want to see in your PyMOL session, change colours of some objects or automatically save your generated visualization as a PyMOL session.

TIP: You need to know that sometimes, if your system is big, the number of identified paths, inlets and clusters are great and you decide to visualize all the available items in PyMOL, PyMOL could not handle it. Then you can try with the `all_paths_amount` *and* `ctypes_amount` *options available in the Expert mode.*

Back to the visualizing script now! If you want to learn more about its functions, simply type in your terminal:

```
python 6_visualize_results.py --help
```

And we should type:

```
python 6_visualize_results.py --keep 'molecule cluster scope object
raw' --discard 'master io all paths smooth'
```

For now we are omitting the smooth paths and many more items AQUA-DUCT is capable of visualizing, since we only want to look at the clusters and the raw data.
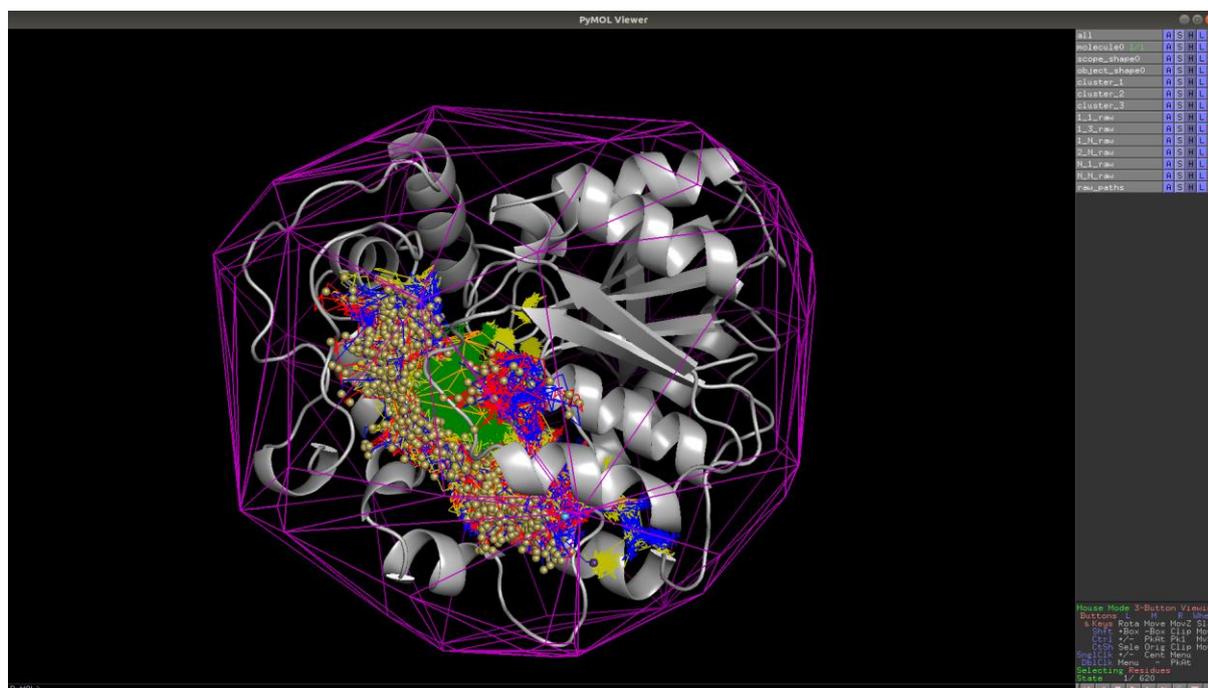
Figure 10. Visualization in PyMOL.

Let's take a look on the visualization. The purple shape represents the *Scope*, the orange stands for the *Object*. We can turn `scope_shape0` off. By default the protein is shown in cartoon representation. Let's now focus once again on paths. Paths are colour-coded: they are blue when coming into the *Object*, green inside the *Object*, blue when coming out from the *Object*. Yellow paths are special case when molecule wander from the *Object* and came back. Now let's look at the protein's surface. The sand, cyan and purple balls are inlets marking the

molecule entries/exits. Let's turn off the inlets representations: `cluster_1`, `cluster_2`, `cluster_3` and `cluster_4`. There are four visible entries to the active site pocket. If we look at all paths, we can see four regions of red and blue paths, what indicates that there are four possible entries/exits. To see them better we have changed in PyMOL protein representation into surface, and background for white, and we pointed four entries by arrows (we have also rotated the protein).
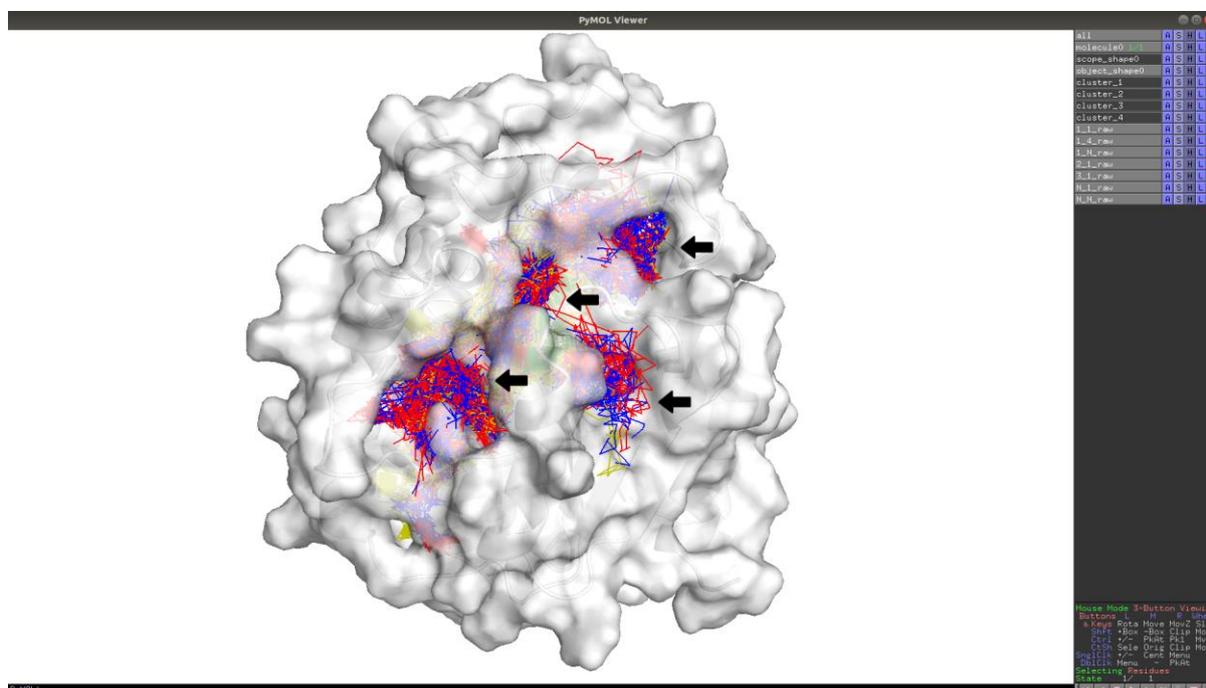
Figure 11. Can you see four different entries/exits?

Unfortunately, almost all inlets (Figure 10) are coloured sand, and it does not correspond to the picture we see on Figure 11. It means, that we need to repeat clustering procedure once again.

NORMAL MODE ANALYSIS

So now we need to tell AQUA-DUCT about how we would like to divide the clusters. Let's run Valve Configurator once again, **Open** your previously generated config file and this time let's use the **Normal mode**. The Normal mode consists of five tabs:
- General options, similar to Initial from the **Easy mode**: where you set calculation options and directories of the trajectory and topology files,
- Tracking: where you define what kind of molecules you want to track (the *Object*) and in which region of interests you want to see the paths (the *Scope*),
- Inlets clustering: where you have more options related with clustering and post-clustering actions,
- Clustering: where you have definitely more available options to run your preferred clustering method(s),
- Visualization, as it is in the **Easy mode**, but much more complex: where you can choose which items you want to visualize.
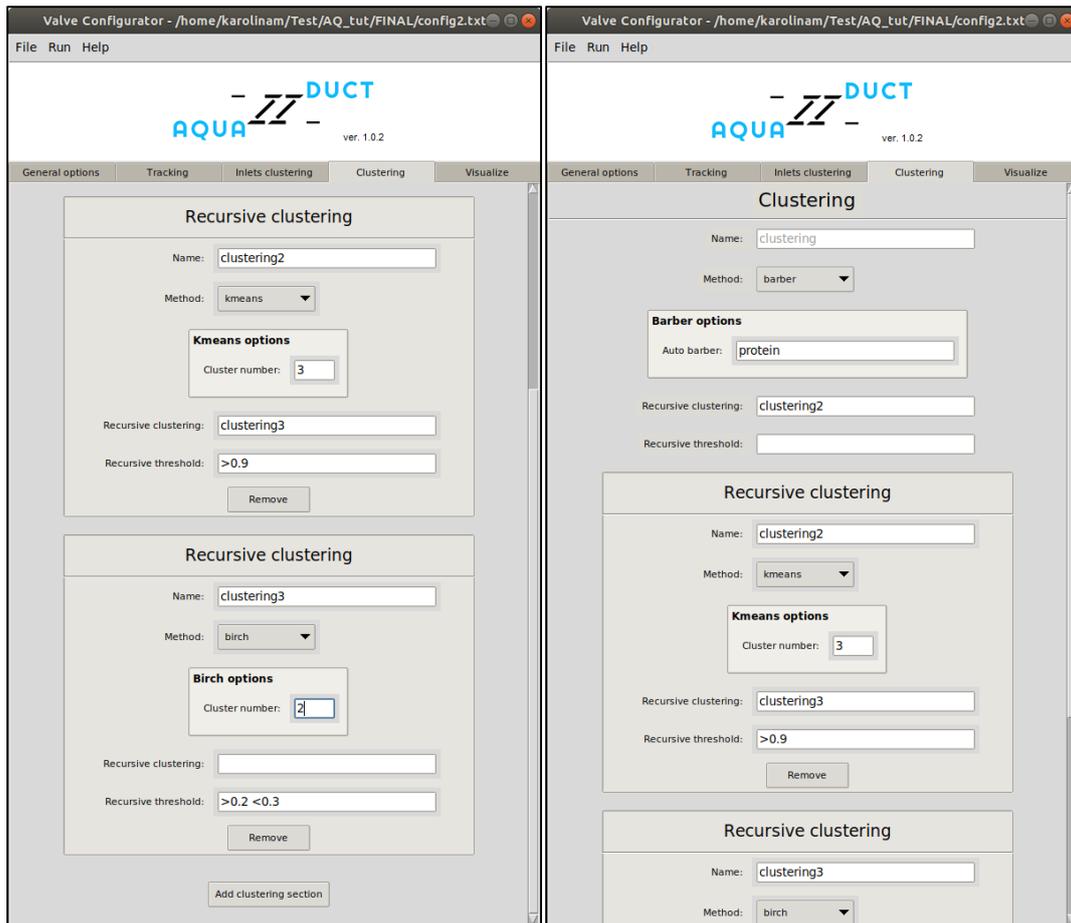
Figure 12. Normal mode - Clustering.

The proposed clustering procedure to obtain four clusters might seem complicated at a very beginning, but you'll see it's really simple! Since none of our clustering methods haven't succeeded in dividing `cluster_1` into four smaller parts corresponding to our identified entries/exits, we need to go a little bit deeper. First we will divide the big `cluster_1` into three smaller clusters and in the next clustering round we will divide the smaller `cluster_7` covering two different entries/exits into smaller chunks. And this is exactly why we sometimes need the **Recursive clustering** method.

To add another clustering method you need to hit the **Add clustering section**, set the **Name** of your clustering step and choose the clustering method (e.g., barber, birch, kmeans, dbscan, affprop and meanshift - you can check them all in AQUA-DUCT Documentation). If you want to add more clustering sections, you need to hit the **Add clustering section** button as long as you need. Please make sure that the **Name** of your clustering step matches the **Recursive clustering** of the previous step.

So here the drill is:
- `clustering`: using the **barber** method what gave us four clusters: big `cluster_1` and smaller clusters: `cluster_2, cluster_3` and `cluster_4` consisting of only 1 inlet,
- `clustering2`: using **kmeans** method to divide big `cluster_1` (**Recursive threshold** set to >0.9) into three parts,

- `clustering3`: using **birch** method to divide smaller `cluster_7` (**Recursive threshold** set to >0.2 <0.3) into two parts.

You can set your target cluster (the cluster that you want to divide) by setting **Recursive threshold** to a specific value or range according to the size of selected cluster. (For more information about clustering methods, please visit our Tips&Tricks - Clustering methods).
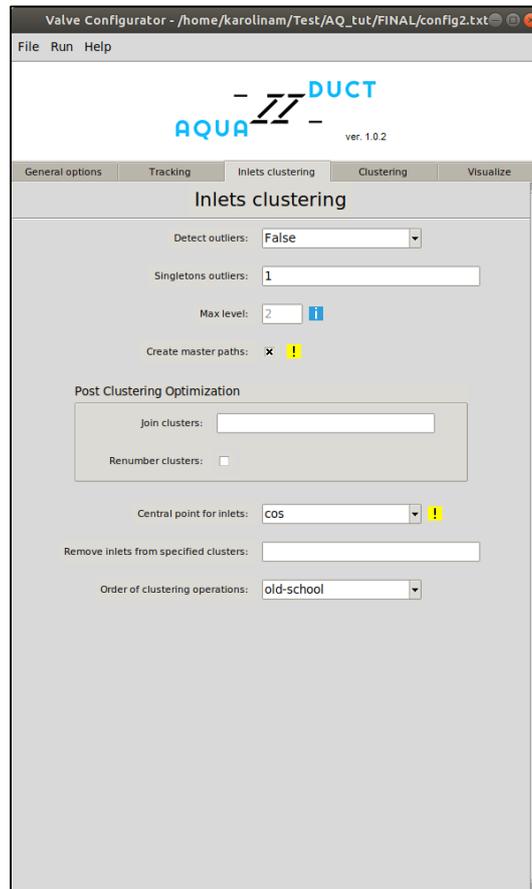
Figure 13. Normal mode - Inlets clustering.

And have you spotted the **Max level** option yet? This is a very important option for the recursive clustering steps. When it is set to 0, only the basic, first clustering step will be conducted. In our case it is set to 2, because we have added two more clustering steps.

Before you start, just check if the other GUI windows looks like in our example below:

Valve Configurator - /home/karolinam/Test/AQ_tut/FINAL/config2.txt

File Run Help

AQUA-DUCT ver. 1.0.2

| General options | Tracking | Inlets clustering | Clustering | Visualize |

General options

Topology file: `topology.prmtop` [Load]
 `[        ]` [Load]

Trajectory file: `trajectory.nc` [Load]
 `[        ]` [Load]

Sandwich: ☐

Waterfall: ☐

Minimal frame: `0`

Maximal frame: `[        ]`

Frame step: `1`

Cache directory: `cache` [Load]

Memory cache: ☐

Valve Configurator - /home/karolinam/Test/AQ_tut/FINAL/config2.txt

File Run Help

AQUA-DUCT ver. 1.0.2

| General options | Tracking | Inlets clustering | Clustering | Visualize |

Tracking

Traceable residues

Scope: `protein`

Convex hull scope: ☒

Object: `(resname WAT) and (sphzone 6.0 (re`

Separate paths

Auto Barber: `protein`

Calculate scope object size

Calculate scope and object size: ☒

Scope hull definition: `protein`

Object convex hull: `(resname *) and (sphzone 6.0 (resnu`

Cluster area: ☒

Valve Configurator - /home/karolinam/Test/AQ_tut/FINAL/config2.txt

File Run Help

AQUA-DUCT ver. 1.0.2

| General options | Tracking | Inlets clustering | Clustering | Visualize |

Inlets clustering

Detect outliers: `False ▼`

Singletons outliers: `1`

Max level: `2` ⓘ

Create master paths: ☒ ⚠

Post Clustering Optimization

Join clusters: `[        ]`

Renumber clusters: ☐

Central point for inlets: `cos ▼` ⚠

Remove inlets from specified clusters: `[        ]`

Order of clustering operations: `old-school ▼`

Valve Configurator - /home/karolinam/Test/AQ_tut/FINAL/config2.txt

File Run Help

AQUA-DUCT ver. 1.0.2

| General options | Tracking | Inlets clustering | Clustering | Visualize |

Clustering

Name: `clustering`

Method: `barber ▼`

Barber options

Auto barber: `protein`

Recursive clustering: `clustering2`

Recursive threshold: `[        ]`

Recursive clustering

Name: `clustering2`

Method: `kmeans ▼`

Kmeans options

Cluster number: `3`

Recursive clustering: `clustering3`

Recursive threshold: `>0.9`

[Remove]

Recursive clustering

Name: `clustering3`

AQUA-DUCT by

Figure 14. Example of Normal mode tabs: General options, Tracking, Inlets clustering, Clustering and Visualization (in two sections).

OK, now we can save the config file.
And… wait!

Did you know that you do not have to repeat the whole AQUA-DUCT calculations over and over after such small changes in your config file? At each stage you can decide if you want to run it only once and next time simply load the particular *.dump file (`execute = runonce`), run the stage every time (`execute = run`) or skip the stage calculations (`execute = skip`). So let's open the config file in your favourite text editor and add the `execute = run` line in the `[inlets_clustering]` section and run the calculations again.

```
[inlets_clustering]
execute = run
dump = 4_inlets_clustering_data.dump
detect_outliers = False
singletons_outliers = 1
max_level = 2
create_master_paths = True
separate_master_all = True
exclude_passing_in_clustering = True
inlets_center = cos
clustering_order = old-school
```

This time we can use the terminal and simply type:

```
valve_run -c config.txt --debug-file aq.out
```

After a while we should see the clustering information in the terminal:

Figure 15. AQUA-DUCT clustering history.

On the **Clustering history** tree you can see how all the clustering steps were conducted and what were their results. We have divided big `cluster_1` into three parts using **kmeans** and then we have selected the small `cluster_7` to be divided into two other chunks using **birch**. This is how we have clusters covering the corresponding entries/exits. We also added smallest clusters `cluster_2`, `cluster_3` and `cluster_4` into the outliers clusters (`cluster_out`, marked on the tree as 0).

When AQUA-DUCT is finished, we can once again use the `6_visualize_results.py` script.

```
python 6_visualize_results.py
```
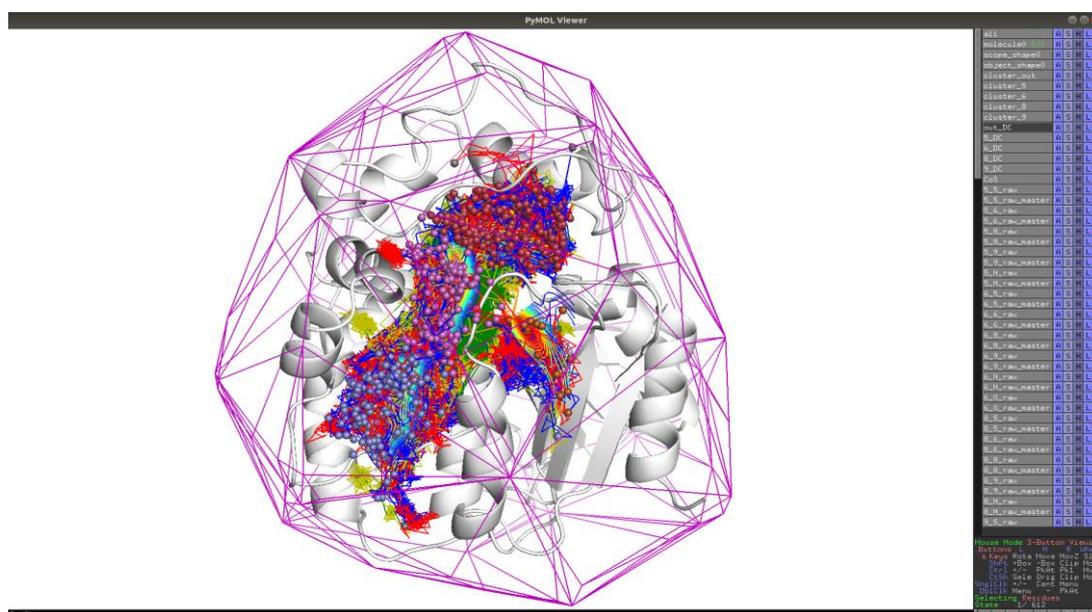

Figure 16. And now we have four clusters!

All right! Now everything looks just fine! As you can see we have four different clusters, just like we wanted. Our clusters have been renumbered according to their appearance.

You're probably surprised about the number of data generated by AQUA-DUCT. In your visualization file, the PyMOL session you can see the basics: the molecule, *Scope*, *Object* and clusters, but also:

- cluster areas,
- CoS point (centre of Scope, the reference point to place the cluster areas),
- raw paths,
- smooth paths,
- raw master paths,
- smooth master paths,
- smoothed raw master paths,
- all raw paths (as separate states in PyMOL session),
- all smooth paths,
- raw paths IO (marking the entrance and exit points for each path),
- smooth paths IO,
- all 'in' parts of raw paths,
- all 'out' parts of raw paths,
- all 'object' parts of raw paths,
- all 'in' parts of smooth paths,
- all 'out' parts of smooth paths,
- all 'object' parts of smooth paths.

Let's stay a moment with the visualization results. You can visually check if your *Scope* and *Object* definitions are defined as you want them to be. Sometimes moving the *Object* into some other place would increase the readability of your results.

Look at all the information that AQUA-DUCT gives you: you can see all the paths grouped into a single cluster and the transfers between the same or different entries/exits. You can have representation of raw and smoothed paths, master paths, or a single path. And that's only the beginning! Please, save your PyMOL session at this point, you'll be needing it later. Below, you will find examples, of what you can learn looking at visualisation.
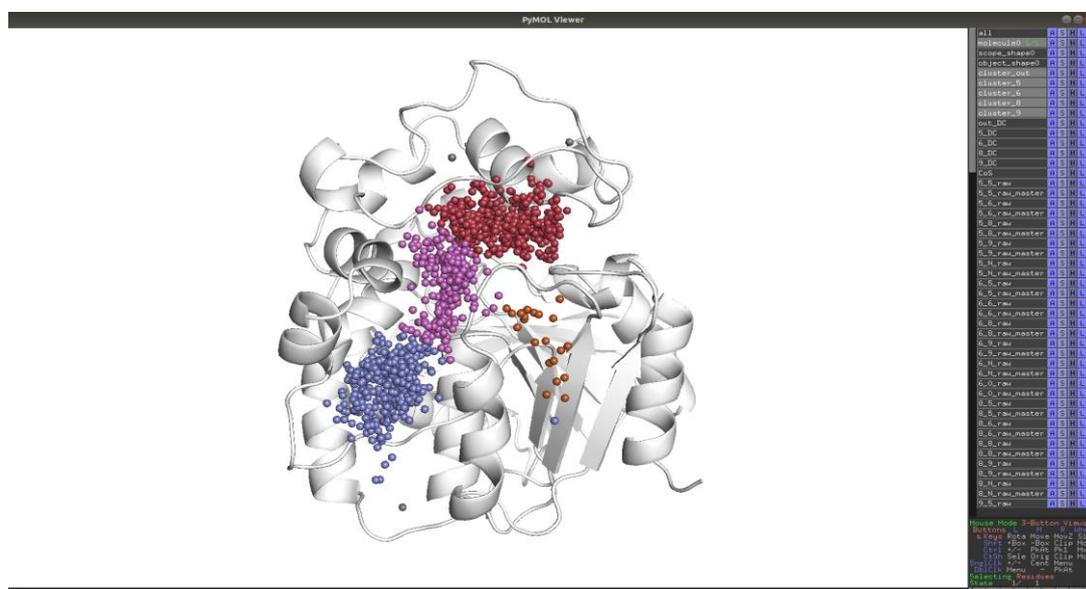
Figure 17. Different representations of AQUA-DUCT results: inlets. The inlets mark the shape and location of the tunnel entrance. They supposed to be sitting on the protein's surface.
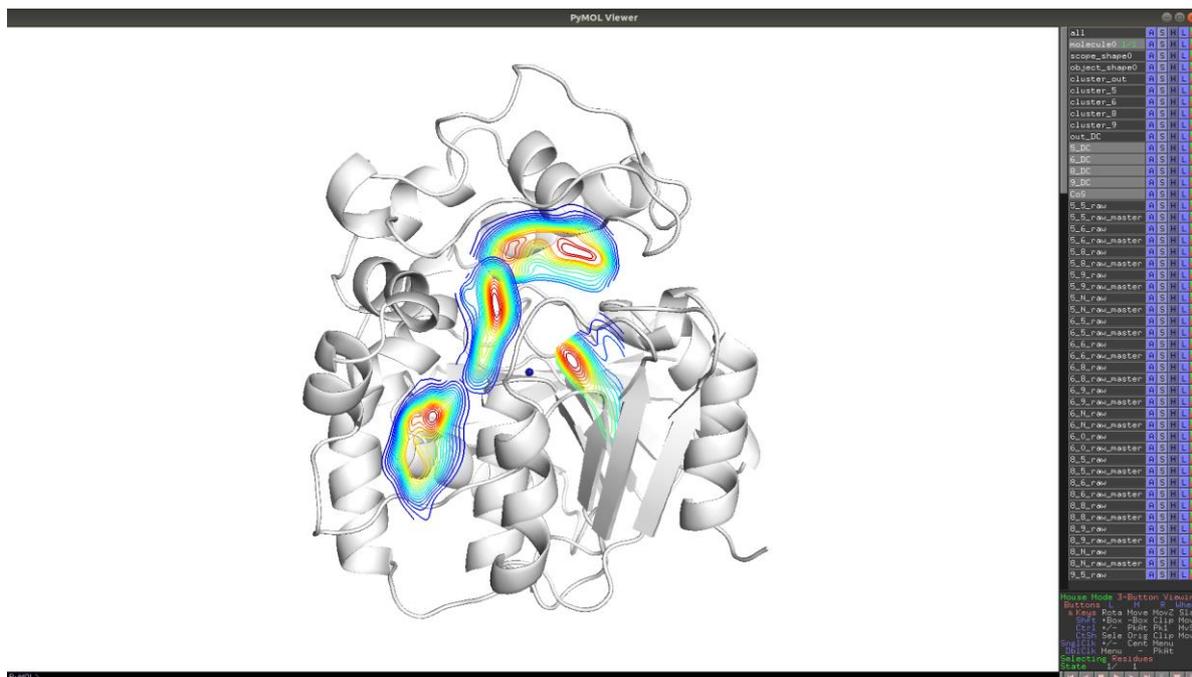
Figure 18. Different representations of AQUA-DUCT results: cluster areas. Cluster areas mark the shape, location and the density of each inlets cluster. They are calculated for clusters bigger than 3 inlets. The redish color mark region of high inlets density, the blueish - low inlets density.
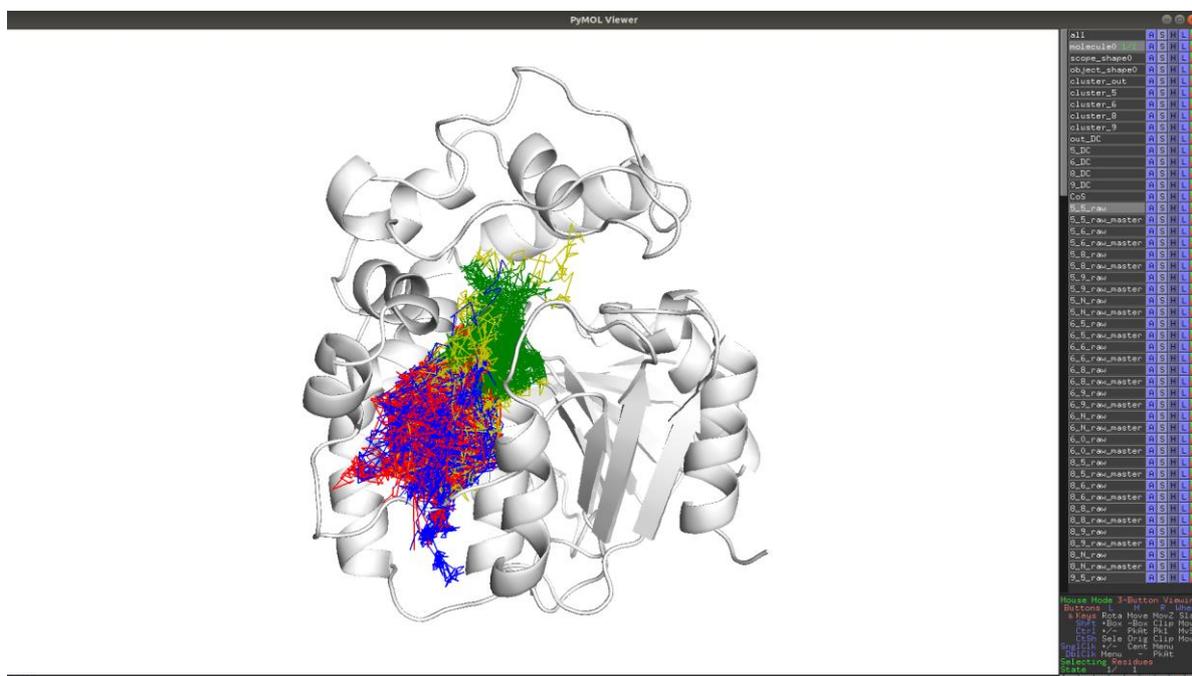
Figure 19. Different representations of AQUA-DUCT results: raw paths. Raw paths represent the data from the trajectory file - the path of tracked small molecule. The red part of the path stands for the moment when the molecule enters the *Scope* and head to the *Object*. The green part mark the molecule movements inside the *Object* and the blue part stands for molecule leaving the *Object* and the *Scope*. Sometimes the molecule visits the *Object* several times and then it is marked by yellow. Here the raw paths are grouped into corresponding inlets cluster.
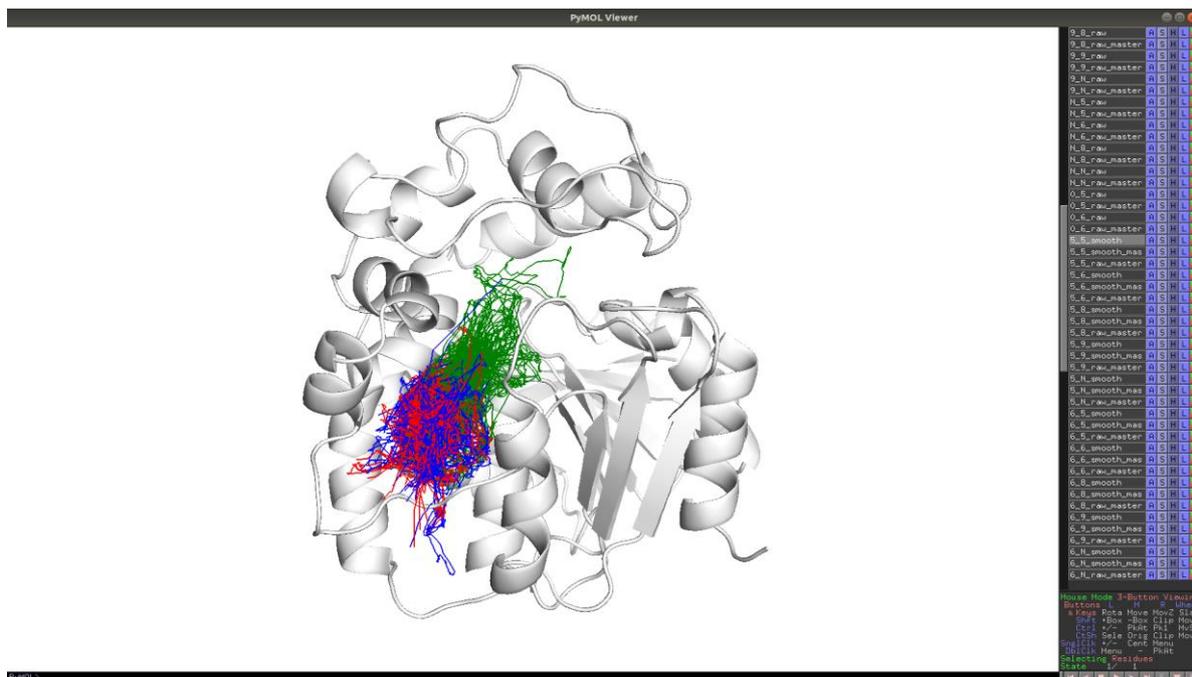
Figure 20. Different representations of AQUA-DUCT results: smooth paths. Smooth paths are raw paths pre-processed by the smoothing algorithm and should be used in case of energy profile calculations. They are color-coded in similar manner as the raw paths, but do not consists of yellow parts. Here the smooth paths are grouped into corresponding inlets cluster.
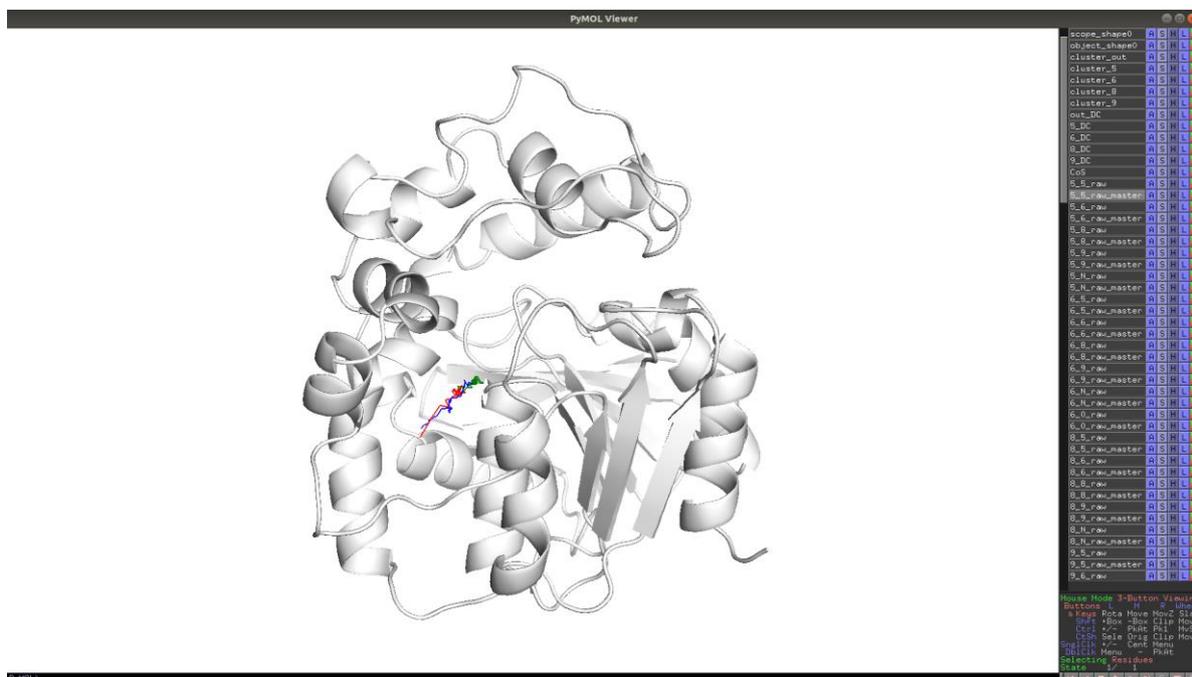
Figure 21. Different representations of AQUA-DUCT results: master path. The master path is the general view of all raw paths grouped as one inlets cluster.
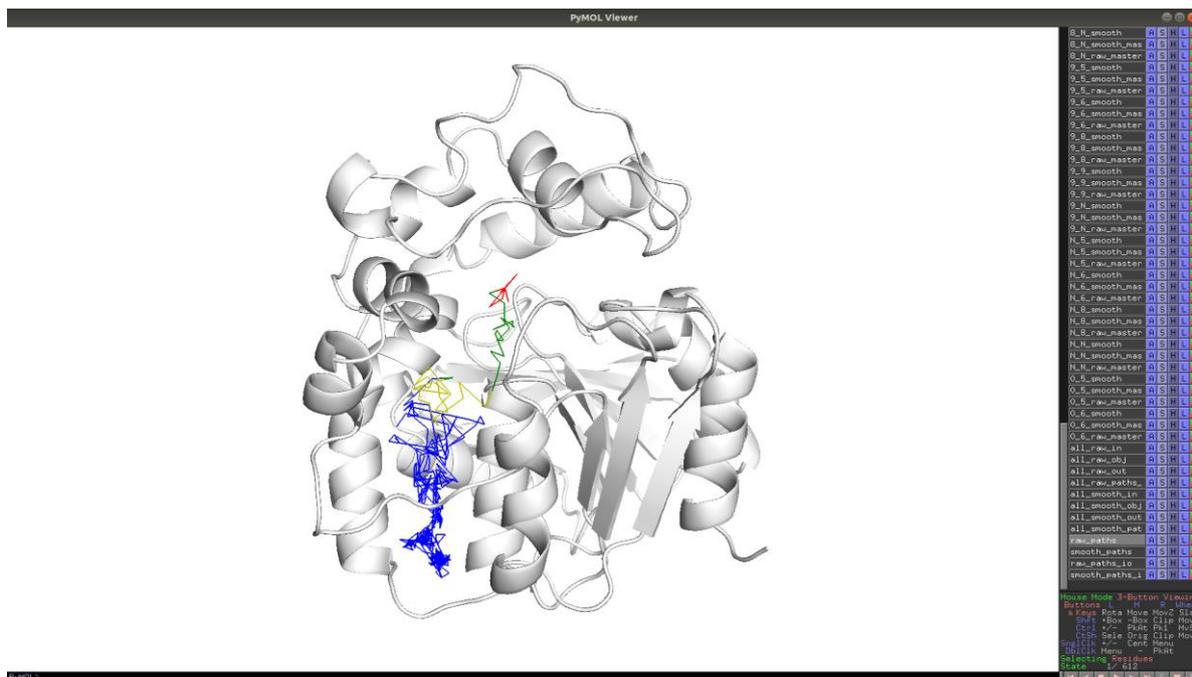
Figure 22. Different representations of AQUA-DUCT results: single path. The single path stands for a separate path of the small molecule during MD simulation. Single paths are color-coded similarly as raw paths.

KRAKEN ANALYSIS

If you liked our Valve Configurator, you'll definitely like the Kraken! Kraken is a small and smart GUI that will help you gather all quantitative results AQUA-DUCT have provided. Please type in the terminal:

`kraken_run`
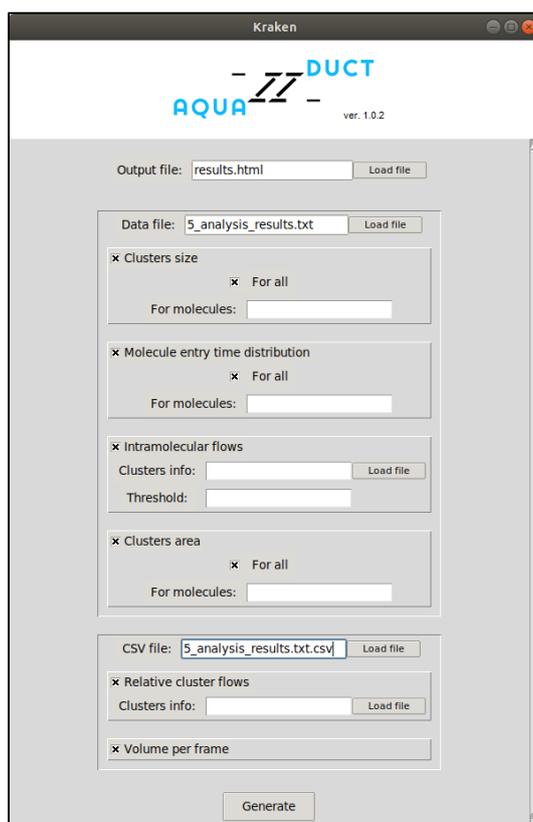
Figure 23. Kraken main window.

Then you need to provide direct localization of your `5_analysis_results.txt` and `5_analysis_results.txt.csv` files as **Data file** and **CSV file**, respectively, and enter the name of your **Output file**, here we have `results.html`. You should use the *.html format to have a good preview of your results. Don't worry, you can always save your Kraken results as separate pictures!

Let's run through the Kraken results now. Since we are only tracking water molecules, we will stay with the **For all** option. If you would like to track more than one molecule type (e.g., water and cosolvent) you could provide the names of your molecules to have separate results for each molecule type. Most of the graphs are quite self-explanatory.
Clusters size
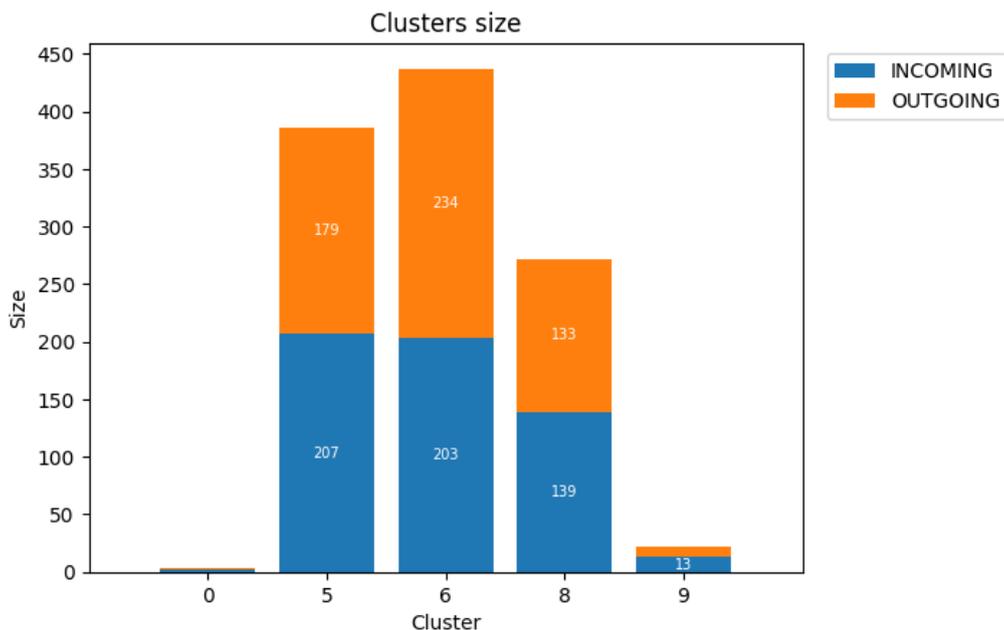
Figure 24. Kraken: Clusters size.

The first picture, **Clusters size** shows you the size of all identified clusters [in inlets] divided into INCOMING and OUTGOING inlets. Please remember, that clusters in our case simple means particular tunnels entry/exits.

Relative clusters flow
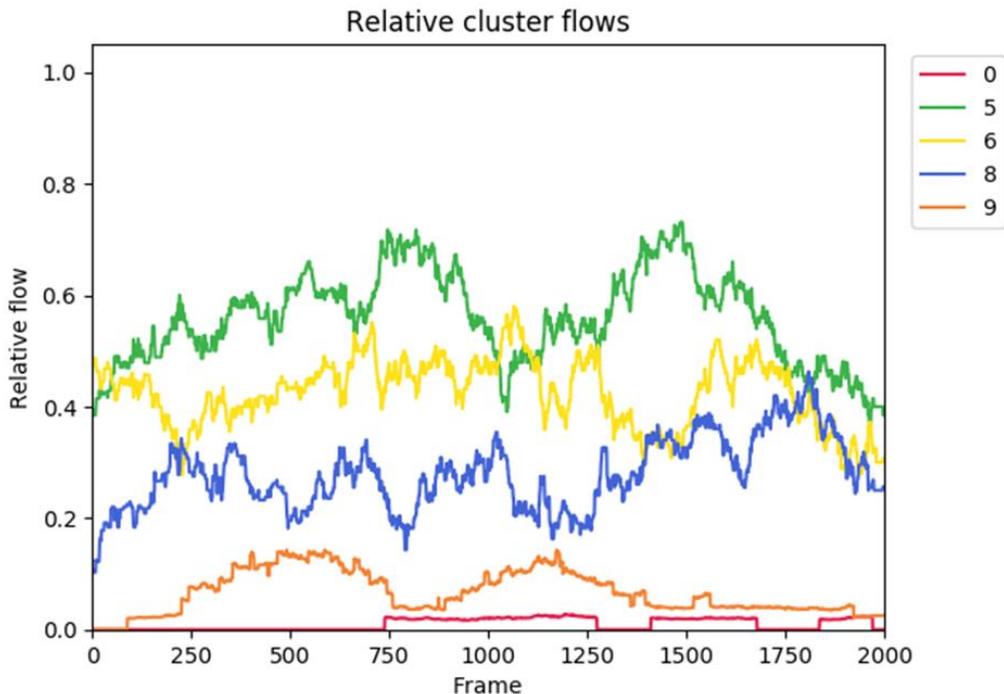
Figure 25. Kraken: Relative clusters flow.

The Relative clusters flow shows you the contribution of particular exits in total flow of the ligands in time.

Molecule entry time distribution
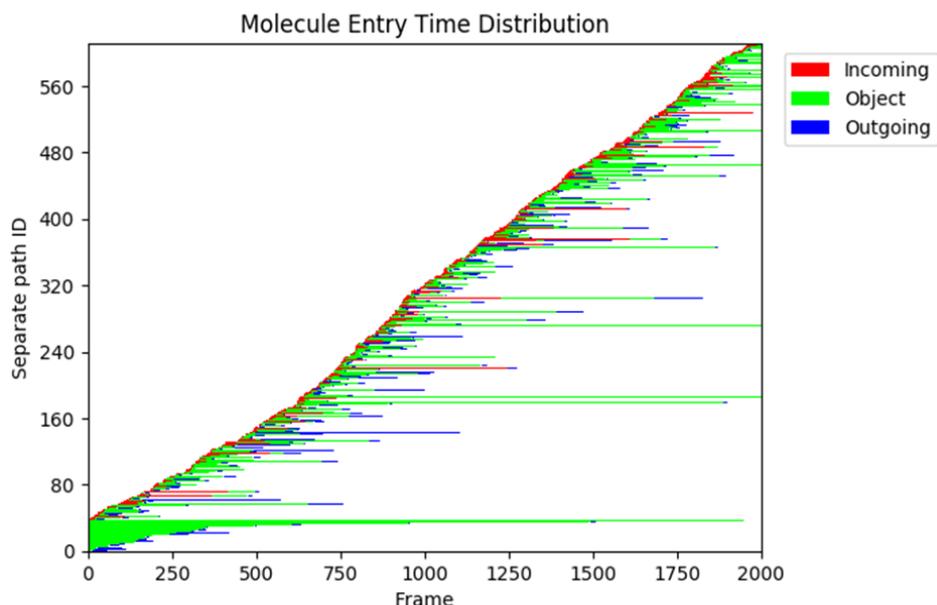
Figure 26. Kraken: Molecule Entry Time Distribution plot.

Molecule entry time distribution plot shows you the color-coded paths sorted by the moment they have entered the *Scope*. As you can see there are few water molecules that have entered the *Scope*, went to the *Object* and stayed there till the end of the simulation, they are mostly green. But some water molecules have problems with entering the *Scope* (mostly red lines) and some other were troubled exiting (mostly blue lines). You can easily find them by searching the `List of separate paths and properties` table in `5_analysis_results.txt` file.

Cluster size and Intramolecular flows
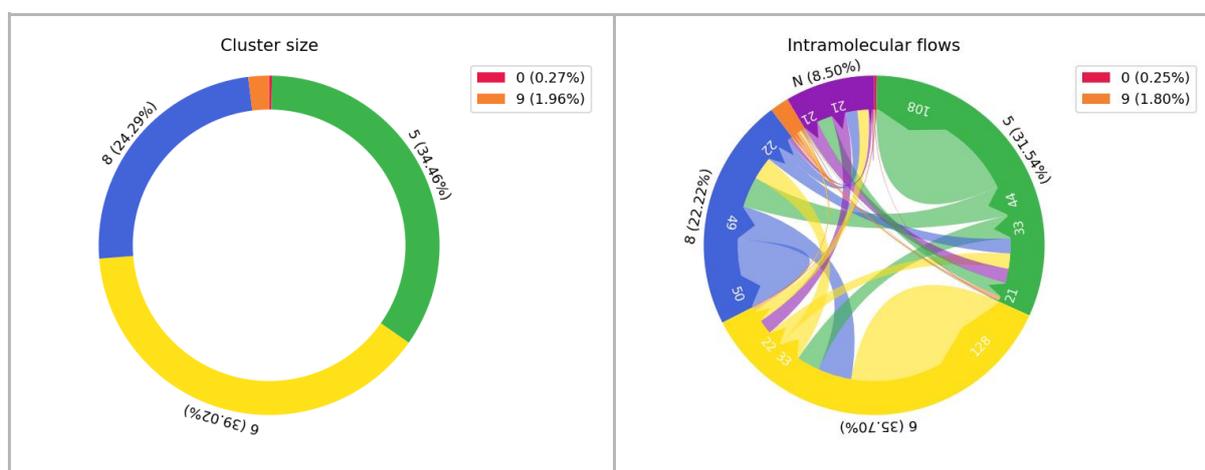
Figure 27. Kraken: The Cluster size and Intramolecular flows graphs.

The Cluster size and Intramolecular flows graphs are giving you the information on the global molecules flow during the simulation. Cluster size - as you would expect - gives you the information on the percentage value of the size of each cluster. The Intramolecular flows graph is much more interesting. It shows you how much water molecules have been exchanged between clusters and in what direction. The biggest inlets cluster, `cluster_6`, exchanges

water molecules within itself, but also with `cluster_5`, `cluster_8` and `N`. Noticed the N cluster already? Well, technically it is not a cluster. The N paths are paths that have their beginning or end within the *Scope*.

TIP: The size of the N 'cluster' depends on size of the system and the length of the MD simulations. For longer simulations the relative size of the N 'cluster' should decrease. If such situation does not occur, something could be wrong with your simulation or with Scope *or* Object *definition. The N:N pathways, simply show molecules which are 'permanently' trapped in the protein core. Again, their number should be rather small and should inform you about molecules, which are essential e.g., for catalysis (like catalytic water in* Solanum tuberosum *epoxide hydrolase.*

TIP: If you observe some weird N paths, that shows in the Object *in the middle of your simulation, that's a sign that something is wrong with your trajectory file of with your config file. In some cases when the* Object *is really close to the* Scope *boundaries, AutoBarber could cut too much of the paths. You can try setting the* `auto_barber_maxcut` *in the* `[separate_paths]` *section (using your favourite text editor to change the config file or using the GUI's Expert mode) to a lower value and see if that will fix your problem.*

Scope and *Object* area and volume
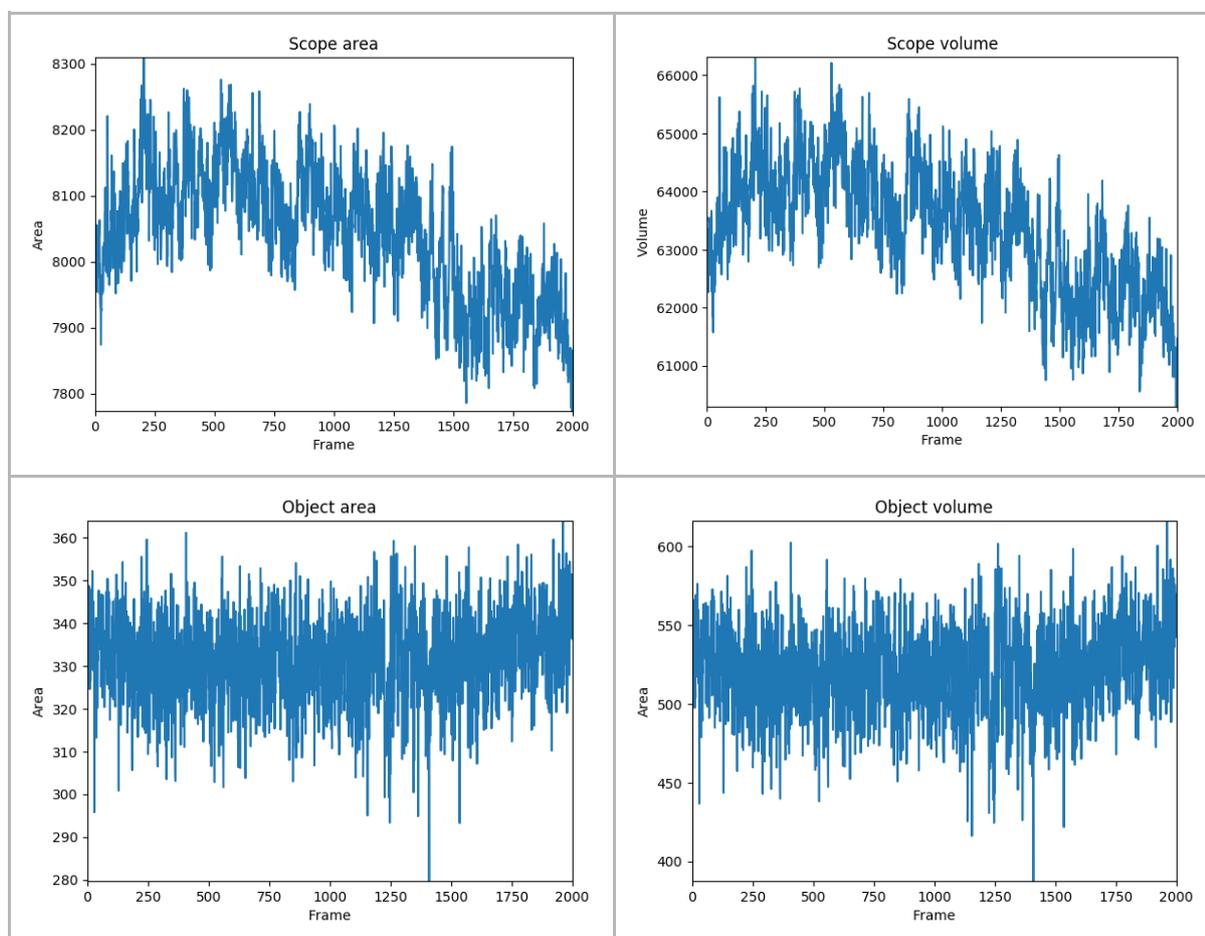
Figure 28. Kraken: The plots of A) *Scope* area, B) *Scope* volume, C) *Object* area and D) *Object* volume.

Next we have plots of *Scope* and *Object* area and volume. These results are based on the ConvexHull approximation. For the sanity of your calculations the plots of *Object* area and *Object* volume should be pretty stable. The shifts on the *Scope* are and *Scope* volume plots could probably tell you something about some conformational changes of the protein structure.

TIP: If you notice large shift of the Scope *or* Object *volume, you can use time-window mode and analyse both states separately. Such analysis will give you the answer about differences of solvent/ligands behaviour in different conformations.*

Clusters area
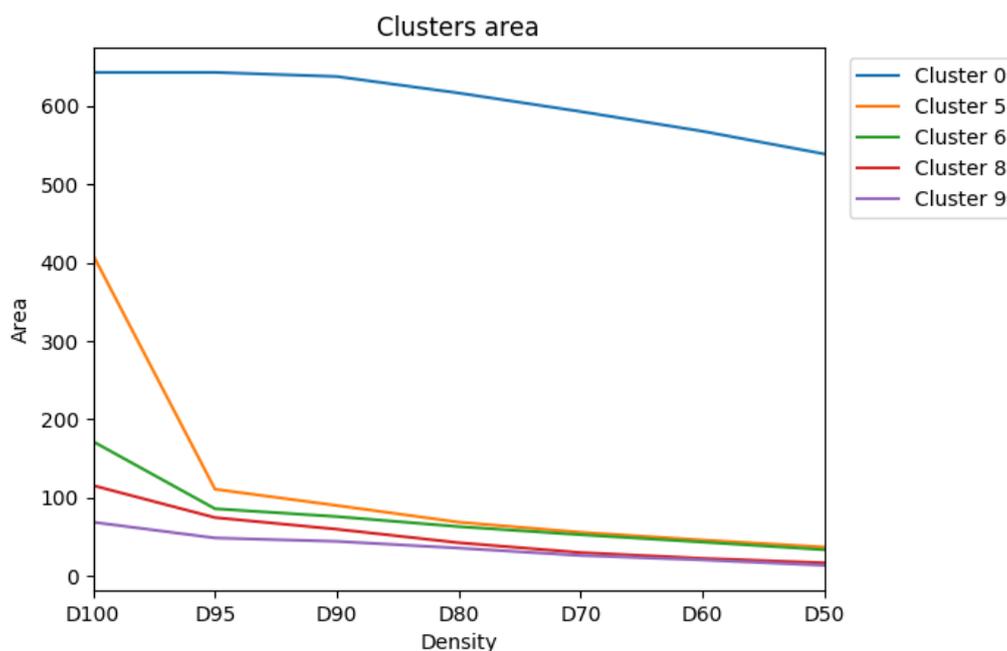
Figure 29. Kraken: Clusters area.

Last but not least comes the Clusters area plots. Remember that rainbow-like representations of inlets clusters? They are based on the inlets clusters density where the red-ish lines indicate the regions of highest density, while the blue-ish lines indicate regions of lowest density. The Cluster area plot shows the area of each region of defined density.

OK, now you've got some basics of AQUA-DUCT analysis. Want to learn something more?

POND: HOT- SPOTS AND ENERGY PROFILES ANALYSIS

Let's see what you can learn from solvent distribution analysis - we're going to generate the hot-spots and energy profiles! For this task we are going to use other module of the AQUA-DUCT software - Pond. You can run Pond directly in the terminal, simply type to get help:

```
pond_run -h
```

Let's use `--pockets`, `--hotspots` and `--raw` option to analyse all the raw data for proper hot-spots calculations. Let's calculate this for the whole simulation. Pond can also divide your simulation into smaller parts and calculate all of previously mentioned aspects for those parts only, we call them windows. At the beginning we're using the `--window-full` option for clarity. Energy profile calculations are based on the density of paths inside the *Scope* and the Boltzmann distribution. To run such calculations you need to have a reference value of density, therefore we will use the `--reference-calc` option to make Pond calculate the reference value for you. if you already know the reference density of your system you can use the `--reference-value` to provide it for Pond in [kJ/mol/K]. Let's save the results of our calculations as *Full_window* directory using the `-r` option.

So for now we need to type:

```
pond_run -c config.txt --debug-file pond.hs.out --pockets --hotspots
--window-full --raw --reference-calc -r Hotspots/
```

Now just hit **Enter** and start Pond calculations!

TIP: For more precise hot-spots and energy profile calculations, please consider using the `add_passing` *option (stage* `[traceable_residues];` *check the* AQUA-DUCT Documentation *for more info!),* `add_passing` *could be set to any molecular entity you want to track during your calculations, e.g.,* `resname WAT`, *when you want to track water molecules that have visited the* Scope *but not the* Object *region. More molecules gives more precise sampling and therefore could enhance identification of hot-spots and give more adequate energy profile.*

After everything is done, once again we're going to need PyMOL. Let's load `inner_full.mol2`, `outer_full.mol2` and `hotspots_full.mol2` into the previously generated PyMOL session. The word 'full' in the file names marks that they were calculated for the whole (full) simulation. If you used the `--windows` and `--wsize` options, your files end with an 'WX', where the X stands for the window number.

Pockets represent the space that was visited during simulation time by analysed molecules, here: by water molecules. The *Inner pocket* represents places that were very often visited, while the *Outer pocket* stands for the regions that were rarely penetrated. You can modify the description of the Inner pocket by setting the `io-threshold` to some other percentage value. By default the median density value is used.
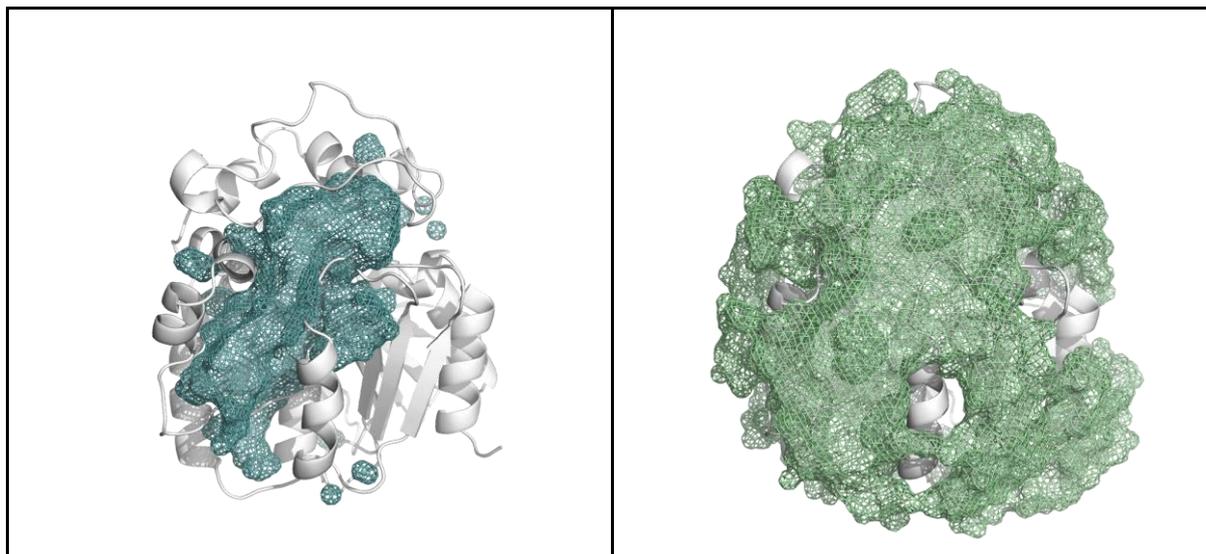
Figure 30. Pond results: *Inner* (light blue mesh) and *Outer pocket* (light green mesh).

Hot-spots mark those regions of the *Scope* interior where the density of paths is really high. This means that they indicate those places and/or amino acids where water molecules spent most of the simulation time. And yes, they have something in common with the *Outer pocket*. Among all hot spots some of them are just hotter than others. And here we provide a script to properly resize the hot-spots according to their real 'temperature'.

First you will need to tell PyMOL that you would like to use a script for hot-spots resize. In the PyMOL window you need type the direct path to the script, and it should be located probably in /usr/local/bin (you can check it by simply typing `which hs_resize.py` in your console). Next you need to use the script on your loaded hot-spots. Pond will need also the `pond_meta.json` file which should be located in your Pond results directory. Try something like:

```
run /path/to/hs_resize.py
hs_resize /path/to/pond_meta.json, hotspots_full
```
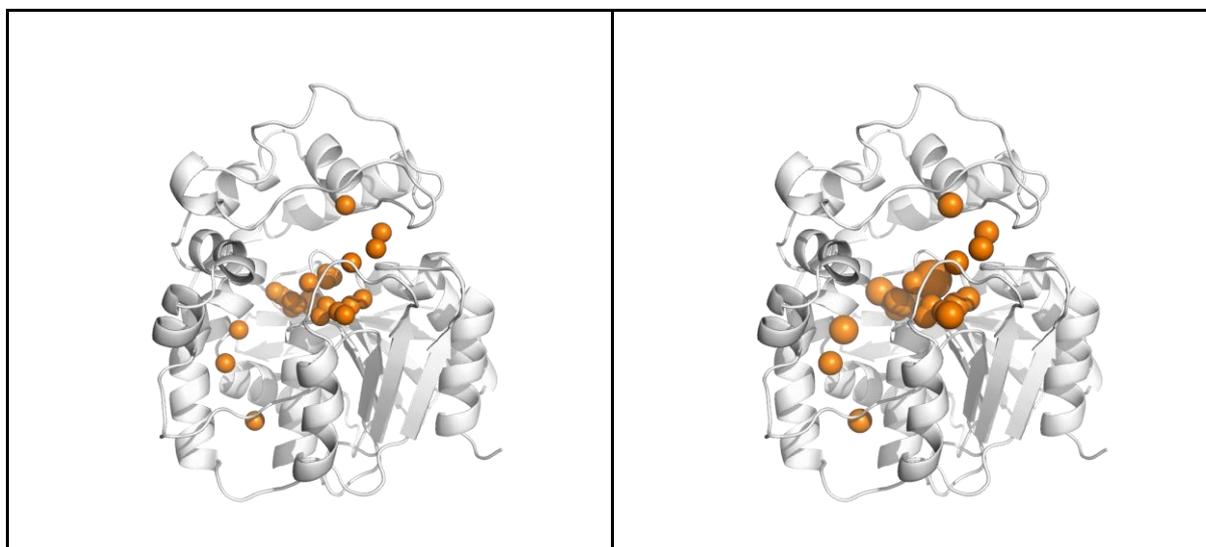

Figure 31. Resized hot-spots.

Besides the resize and simplifier scripts, you can use some of PyMOL built-in features like the `spectrum` command which allows you to colour your hot-spots according to their density. We like when the 'coldest' hot spots are blue and the 'hottest' - red. PyMOL has a palette for this, *blue_white_red*. So we simply type:

```
spectrum pc, blue_white_red, hotspots_full
```

And enjoy the view!

Can you tell which hot-spots are the hottest of all?
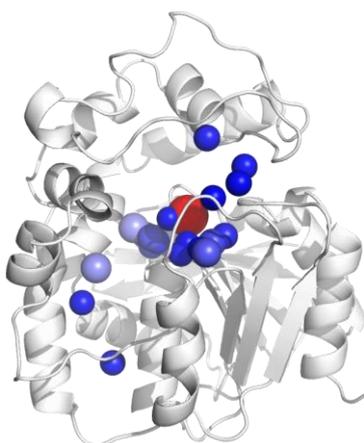
Figure 32. Resized and coloured hot-spots.

And the last but not least - the energy profile. During analysis you will probably spot some more and less interesting paths. If the path is smooth enough (we encourage you to try some of the smoothing paths options - check the AQUA-DUCT Documentation) and do not have any knots, you can calculate the energy profile along the selected path using Pond and the --energy-profile option. Here we used 105 recursive runs of window method by adding:

```
[smooth]
method = window
recursive = 105
```

to the config file and re-run AQUA-DUCT calculations at least from the 3rd stage, `[separate_paths]`. Now you need to define the **Path ID** of your selected path and you can find it in `5_analysis_results.txt` file in the `List of separate paths and properties` table.

For example: the path 31 in PyMOL session will have the Path ID of 0:718:0, and path 109 will have ID like: 0:2124:0. The number of the path in the PyMOL session is equal to the first column (*Nr*) in that table. The **Path ID** is in the second column (*ID*). Please keep in mind that you want to use the smooth path and always use this option together with --path-smooth option. So let's run the calculations in the terminal by typing:

```
pond_run -c config.txt --debug-file pond.energy.out --energy-profile
--path-id 0:2124:0 --path-smooth --reference-calc -r Energy_profile/
```

In the results directory you'll get two files: `path0:2124:0__radius.dat` and `path0:2124:0__radius.mol2`. The *.mol2 is now more interesting, let's open it in the PyMOL session that you have generated. Here you can also use the `spectrum` command to colour your energy profile as you like. Here we will use:

```
spectrum pc, green_red, path0_2124_0__radius
```
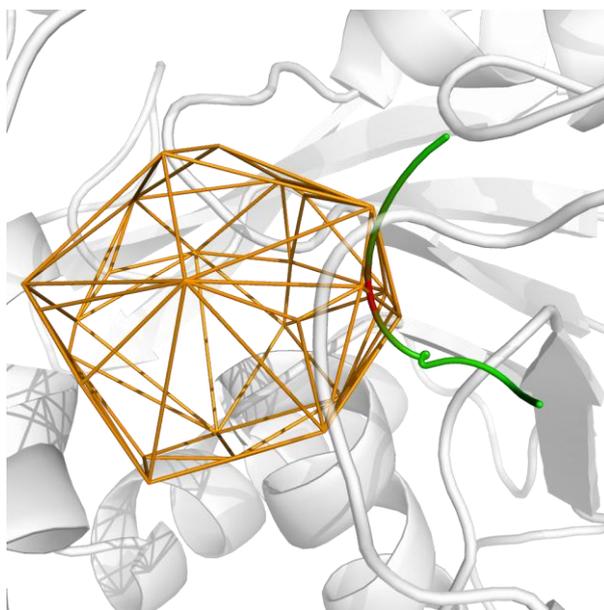

Figure 33. Coloured energy profile.

And the last thing - the *.dat files. You can use them in your favourite plotting software to present the 2D plot of the energy profile. As an example see here:
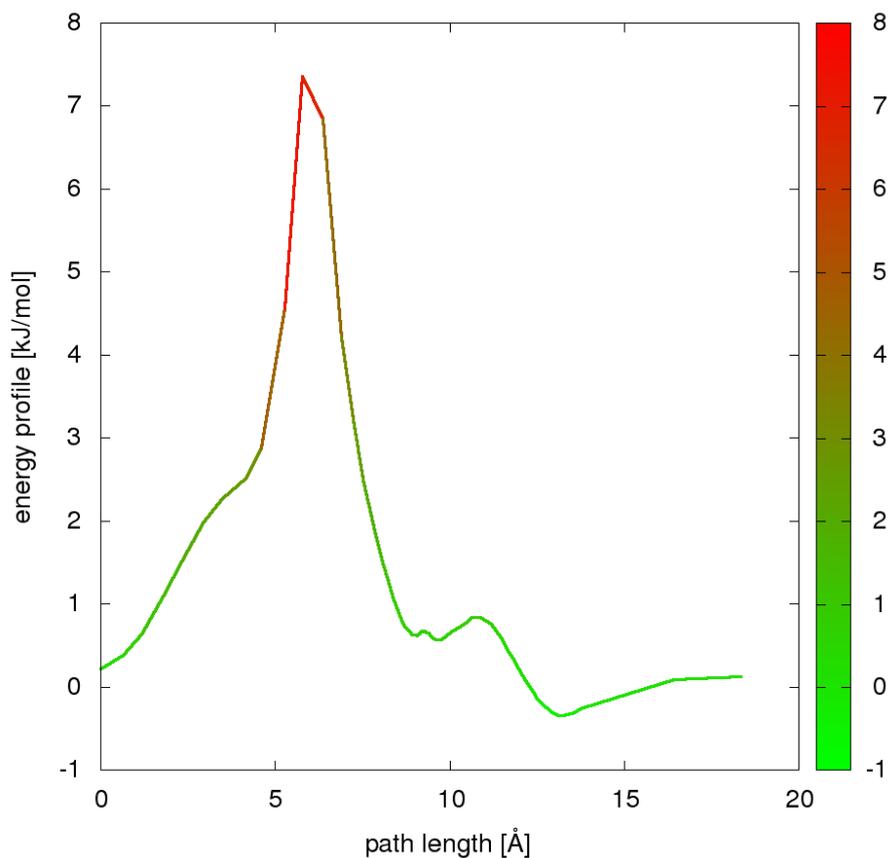
Figure 34. 2D plot of energy profile.

TIME-WINDOW ANALYSIS

Sometimes it's better to use small chunks of simulation to better understand the time evolution of the system. Pond allows to such analysis when you use two options: `--wsize` and `--windows,` where you respectively can determine the number of frames in each analysed chunk and the number of analysed chunks. And they can overlap as well (please check the AQUA-DUCT Documentation to see how to conduct analysis in overlapping windows). Here we show what happens with the hotspots and the energy profile if you would like to use the time window analysis.

Our sample trajectory file consists of 2000 frames (10 ns saved each 5 ps). For demonstration purpose we divided the simulation into 5 windows, 400 frames each.

You can try such analysis on our sample data by typing:

```
pond_run -c config.txt --debug-file pond.energy.out --energy-profile
--path-id 0:2124:0 --path-smooth --hotspots --pockets --windows 5 --
wsize 400 --raw --reference-calc -r Analysis_in_windows/
```
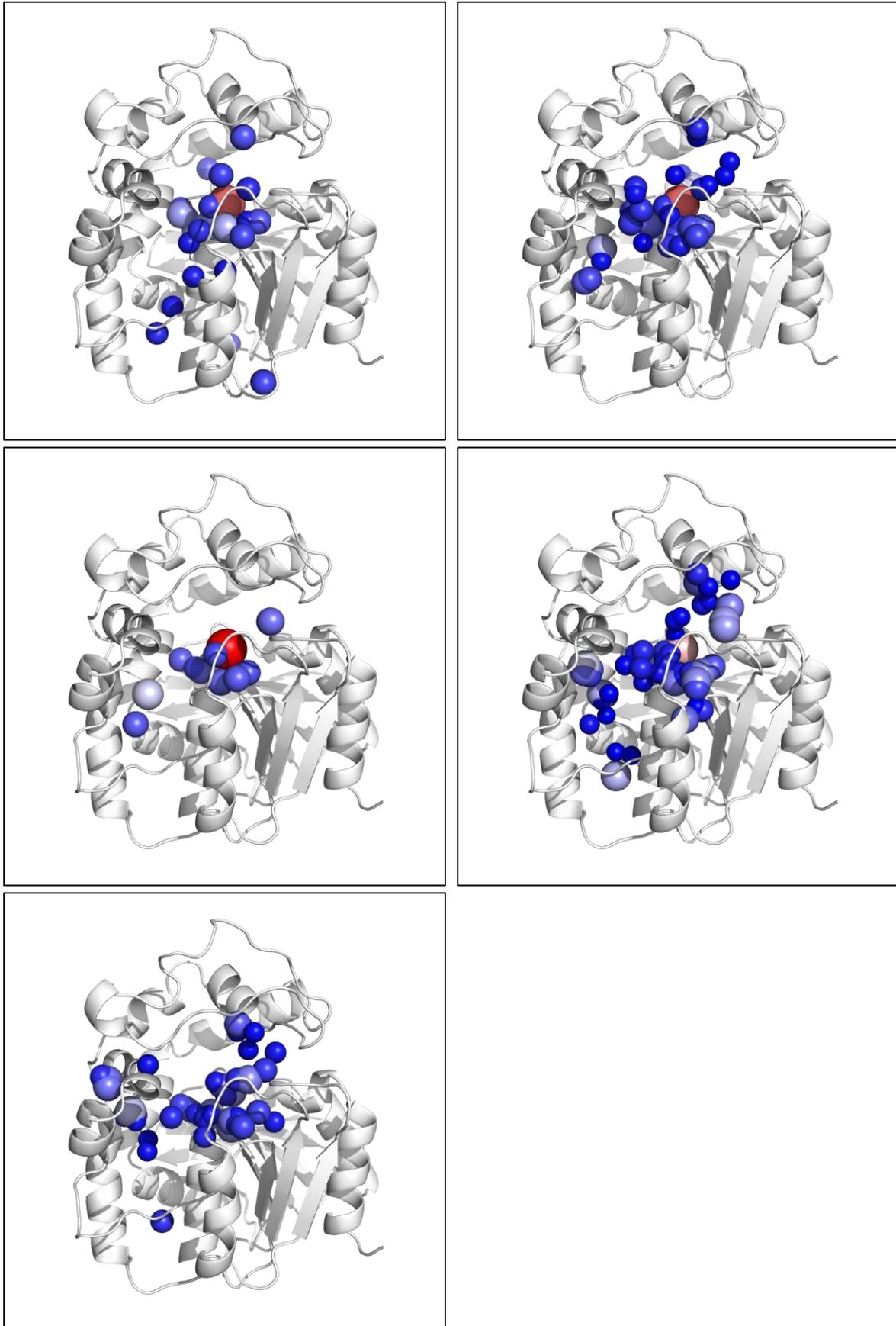
Figure 35. Hot-spots calculated in 5 windows.
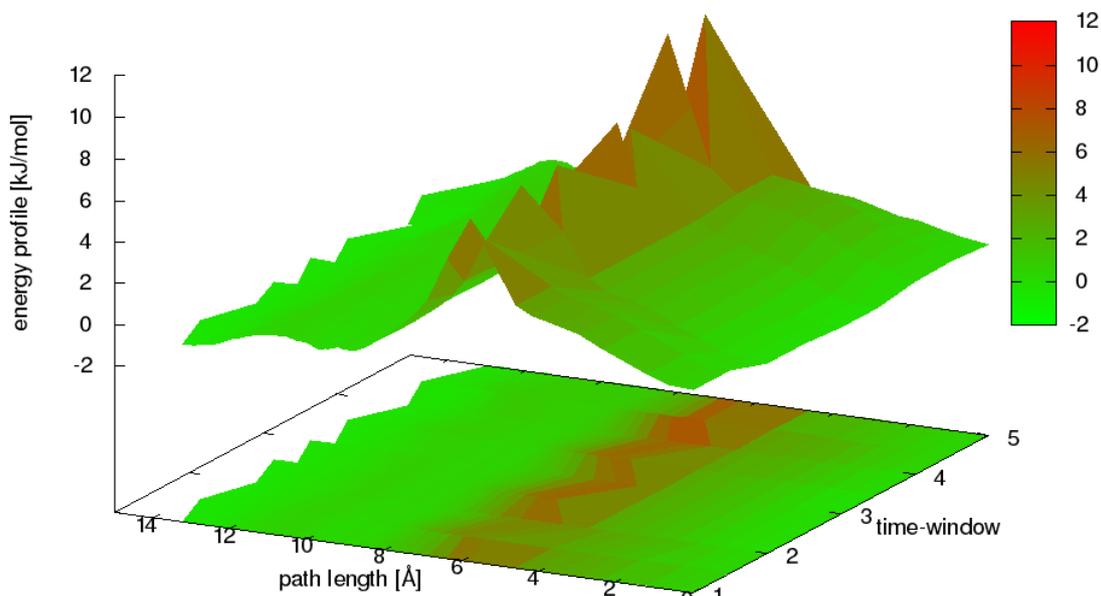
Figure 36. Energy profile plot for 5 windows.

TIP: You can analyse the Inner and Outer Pocket in windows as well, especially when you suspect that your macromolecule of interests changes conformation and that could affect the flow of small molecules. Go on, try it!

Well, that's the end! Thank you for reading this user guide! Hope you have learn how to deal with AQUA-DUCT and its little helpers. Don't forget to check our Tips&Tricks for more AQUA-DUCT tutorials! If you still have some questions or you think AQUA-DUCT is missing some key features, please contact us at info@aquaduct.pl!