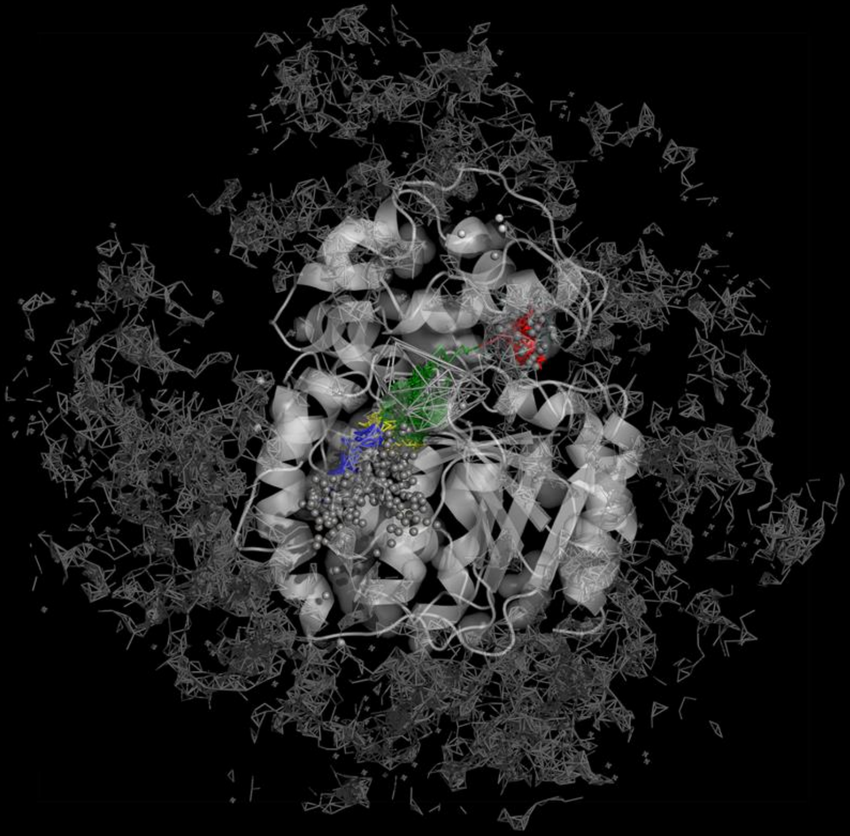
AQUA DUCT by T G



TIPS & TRICKS – trimming paths

TIPS & TRICKS – trimming paths

Trimming paths

This section will give you some insights on how to adjust paths' trimming parameters in order to get optimal results. As an example case study we are using a 10 ns MD simulation of murine soluble epoxide hydrolase structure (PDB ID: 1CQZ).

Lengths of paths

The general rule is that paths are limited to the `scope` area. The bigger `scope` is the longer are the paths, see **Figure 1**. AQUA-DUCT supports definition two types of the `scope` definition: *direct* or *convex hull*. In both cases user can decide on the size of the `scope`. For example, broad *direct* `scope` definition could be: *not protein around 3 protein.* This will select all non-protein residues within 3 Å from the protein. By changing distance value in this definition one can decide on the size of the scope. Another possibility is to change the definition to select residues within different molecular entity: *not protein around 3 backbone*. Similarly, in case of *convex hull* type of definition of the `scope`, its extent can also be adjusted to. For example, to set `scope` to convex hull of protein, protein but hydrogen atoms, protein backbone, and Cα atoms, following `scope` definitions have to be used: *protein*, *not name H and protein*, *backbone*, *name CA*.

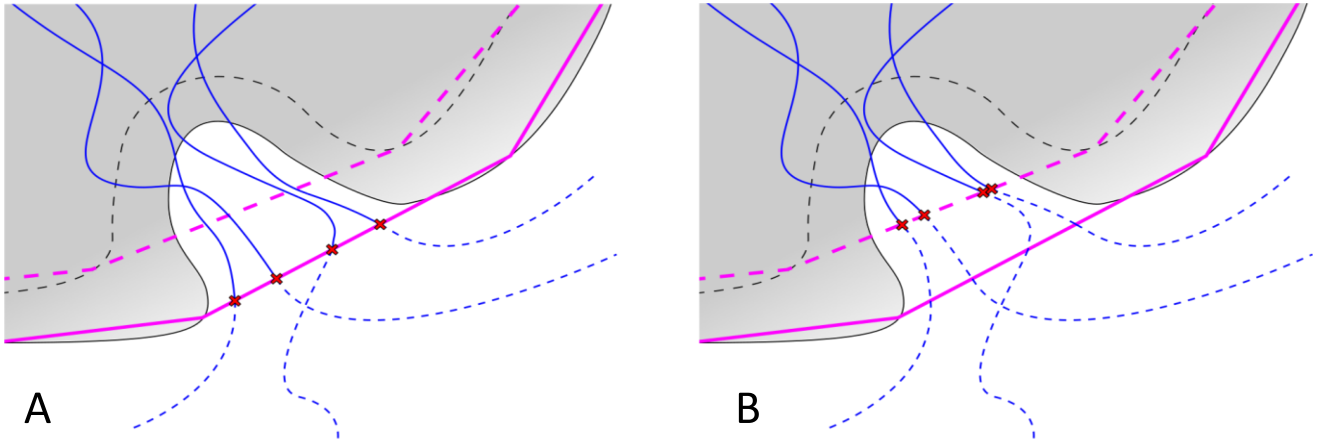


Fig. 1. Lengths of paths depends on the scope definition. Protein represented as gray shape, backbone as dashed gray line. Paths represented as blue lines, dashed if trimmed. Points of trimming indicated with red x signs: a) scope defined as protein – purple line schematically shows the convex hull shell of protein atoms; b) scope defined as backbone – dashed purple line shows the convex hull shell of backbone atoms.

TIPS & TRICKS – trimming paths

Changing the definition of the `scope` is not the only way to influence lengths of paths. Additional trimming of paths can be done with Auto Barber procedure at III stage of AQUA-DUCT calculations. Auto Barber works by creating a set of spheres and removing parts of paths that are inside of any spheres. Centers of spheres used by Auto Barber are set to the ends of paths that are at the edge of the `scope`. Radii are set to distance from the centers to the nearest atom of user defined molecular entity, see **Figure 2**. There are number of options allowing for adjustment of radii, positions of centers depends on the `scope` definition.

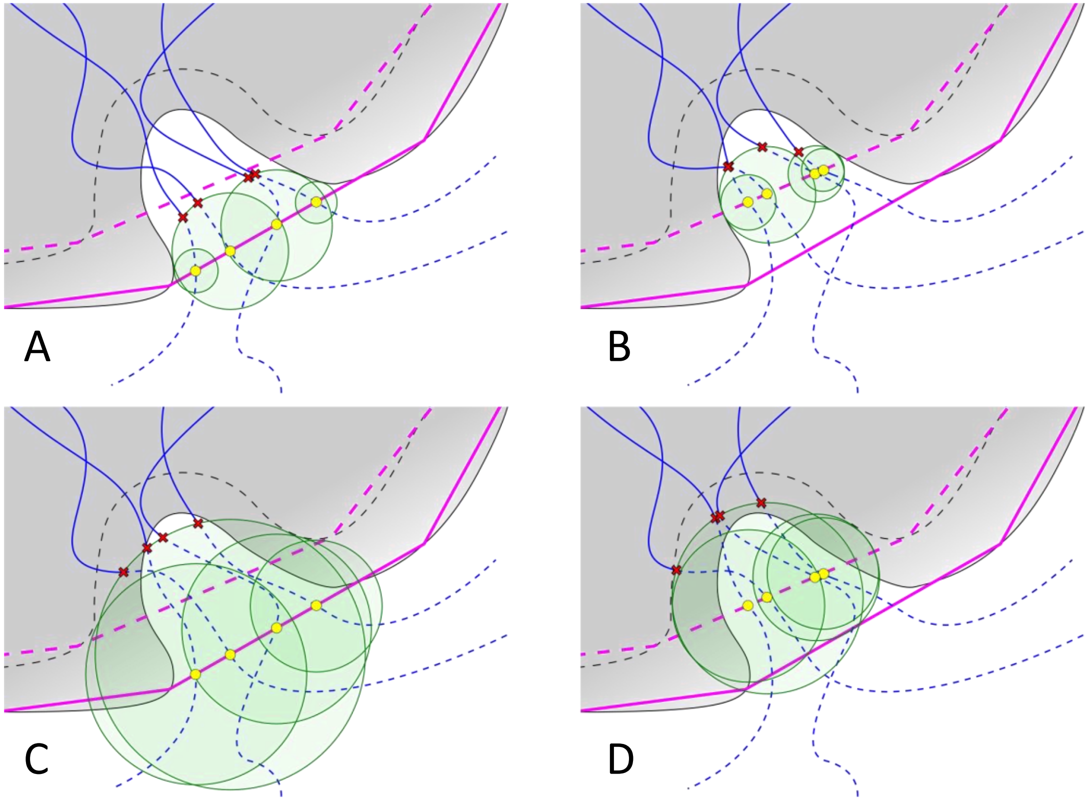


Fig. 2. Lengths of paths depends on Auto Barber settings. Auto Barber spheres are represented as green discs. Yellow spots represents points on the `scope` edge: a) `scope` set to protein, `auto_barber` set to protein; b) `scope` set to backbone, `auto_barber` set to protein; c) `scope` set to protein, `auto_barber` set to backbone; d) `scope` set to backbone, `auto_barber` set to backbone.

TIPS & TRICKS - trimming paths

Case study analysis

The inspection of MD simulations of murine soluble epoxide hydrolase structure and results yielded with `meanshift` clustering method suggests that optimal clustering method should group inlets into one main cluster – C1, located in a huge surface hollow, and two smaller clusters – C2 and C3, located in shallow depressions (**Figures 3a, 3b**). Defined clusters with AQUA-DUCT correspond to the hollows in the protein surface (**Figure 3b**) and groups of entry/exits of smoothed trajectories of water molecules (**Figure 3c**). Other inlets should be classified as outliers, since transport of solvent molecules through the rest of the tunnels is limited in comparison to the main clusters.

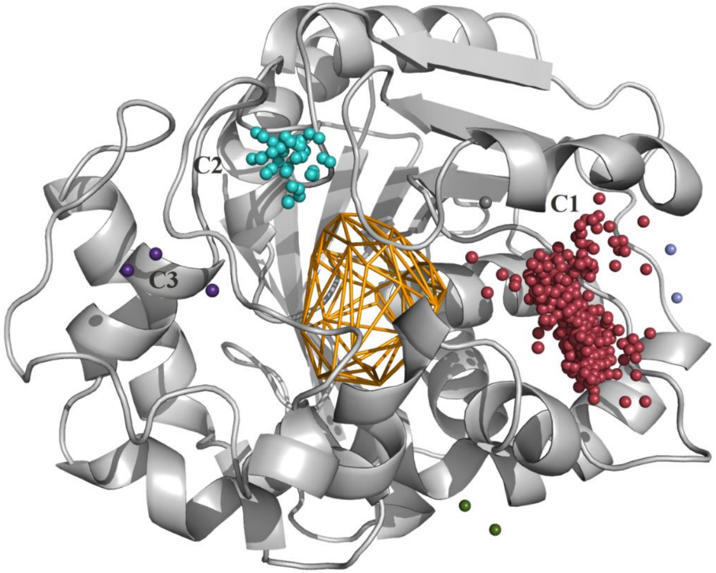


Fig. 3a. Optimal results of clustering of AQUA-DUCT results with `meanshift` method with three main clusters provided: C1 (red), C2 (cyan) and C3 (purple) and other minor 1- or 2-inlet clusters

```
[clustering]
 method = meanshift
 bandwidth = Auto
 cluster_all = True
```

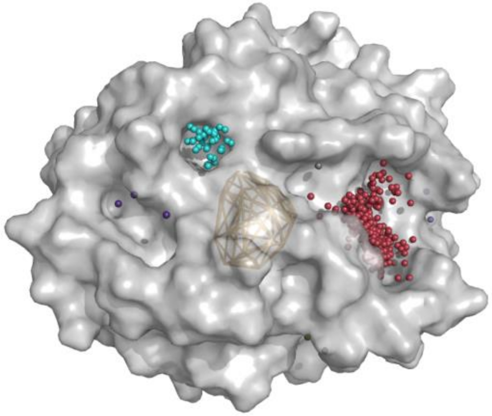


Fig. 3b. Surface of the protein with clusters located in the proximity of the hollows

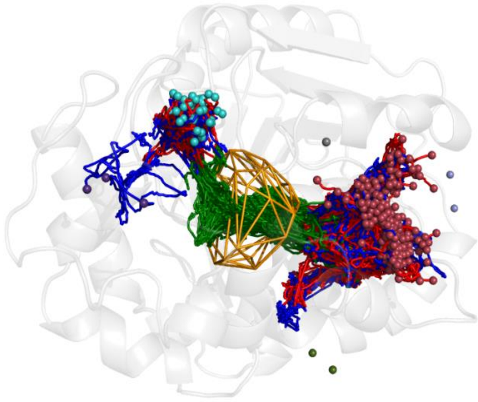


Fig. 3c. Smoothed trajectories of water molecules with entry/exit detected in one of the main clusters

TIPS & TRICKS - trimming paths

Basic information – influence of scope definition for water inlets number

There are several parameters controlling level at which water inlets are trimmed. Setting `scope` at stage `traceable_residues` is the first option to roughly define the level. `Scope` can be also redefined at stage `raw_paths`, which is an additional feature switched off on default settings. A feature that offers several parameters is Auto Barber at stage `separate_paths`. The multitude of Auto Barber options gives the user plurality of possibilities to adjust desired level of trimming and also dispersion of water inlets at the protein surface. Auto Barber is also available at stage `clustering`, provided that method of clustering is set to `barber`, but at this stage **it cannot be used for path trimming**, only for grouping inlets into clusters. We recommend, to use by default Auto Barber at stage `separate_paths` setting switched on, which enables to trim paths in the neighborhood of surface hollows which allows to determine tunnels used for transport of water molecules from stated cluster (**Figures 3d, 3e**). Otherwise paths are trimmed at the altitude of `convexhull` determined by `scope` definition and water inlets are defined too far from the tunnel entries and dispersed at the boundaries of the defined `scope` (**Figures 3f, 3g**).

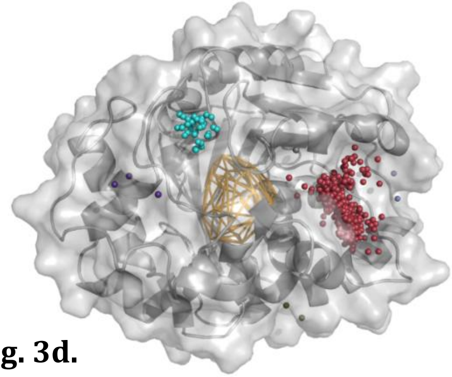


Fig. 3d.

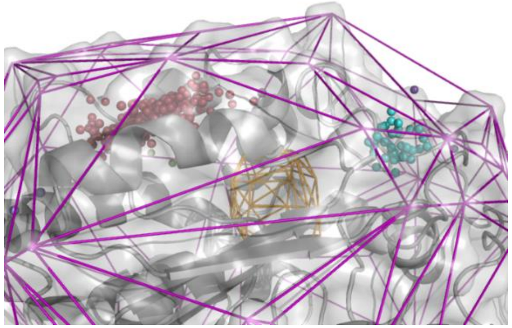


Fig. 3e.

Results of clustering with `meanshift` method and Auto Barber at stage `separate_paths` on default settings.

```
[clusterization]
method = meanshift
bandwidth = Auto
cluster_all = True
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 2.8
auto_barber_mincut = None
auto_barber = protein
```

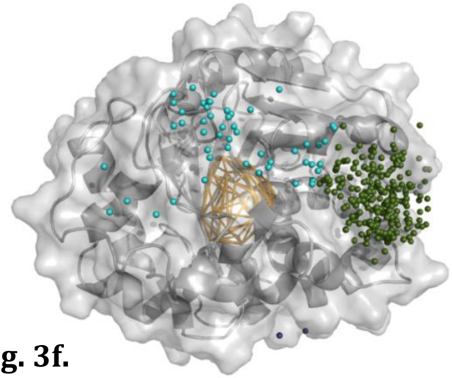


Fig. 3f.

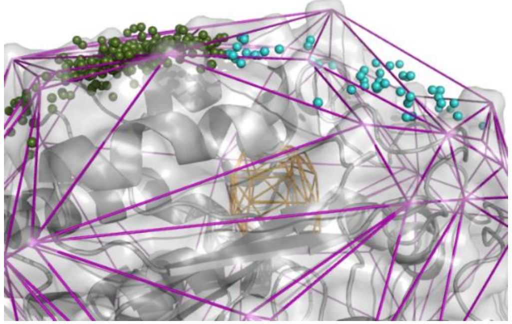


Fig. 3g.

Results of clustering with `meanshift` method and Auto Barber at stage `separate_paths` switched off.

```
[clusterization]
method = meanshift
bandwidth = Auto
cluster_all = True
[separate_paths]
auto_barber = None
```

TIPS & TRICKS - trimming paths

The number of water molecules in each calculation for respective system is constant but the quantity of paths and inlets can vary depending on the set level of trimming, since some water molecules enter the `scope` more than once during a simulation. When the distance between the level and `object` is augmented, the mean lengths of paths increase, as the number of inlets do, because the probability of water molecule entry increases (**Table 1**).

Table 1. Comparison of results (number of inlets and mean lengths of paths) of clustering with `meanshift` method and Auto Barber at stage `separate_paths` switched on/off

Settings	Number of traceable residues	Number of inlets	Path	Size	Mean lengths of paths		
					Input	Object	Output
Meanshift with Auto Barber switched on (Figures 3d, 3e)	151	309	C1:C1	111	351.8 ± 432.5	391.7 ± 613.8	339.8 ± 472.2
			C2:C2	10	282.9 ± 239.0	406.2 ± 650.5	216.8 ± 182.3
Meanshift with Auto Barber switched off (Figures 3f, 3g)	151	264	C1:C1	80	555.5 ± 515.4	501.6 717.6	489.5 ± 580.7
			C2:C2	12	406.6 ± 273.7	391.4 ± 607.6	377.5 ± 218.9

TIPS & TRICKS - trimming paths

Scope definition

The easiest way to roughly define the surface to which paths are trimmed is to change the `scope` definition at stage `traceable_residues`. On default setting, `scope` is set to `protein` but the user can also set it to `backbone`, `name CA` or nonstandard values, such as a group of amino-acids. Changing `scope` definition from default settings (`protein`) to `name CA` immediately changed the depth of water molecules in reference to the protein surface (**Figure 4c**), whereas `scope` defined as `backbone` (**Figure 4b**) gave similar result to default ones (**Figure 4a**). When `scope` is set to `name CA` water molecule trajectories are trimmed closer to the defined `object` (**Figure 4c**) in comparison to when `scope` is set to `protein`. These settings caused location of most of cluster C1 underneath the protein surface and also unification of cluster C3 which was previously divided into two sections.

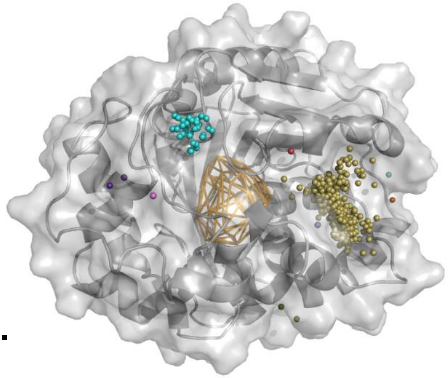


Fig. 4a.

```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 2.8
auto_barber_mincut = None
auto_barber = protein
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = None
auto_barber_mincut = None
auto_barber_tovdw = True
```

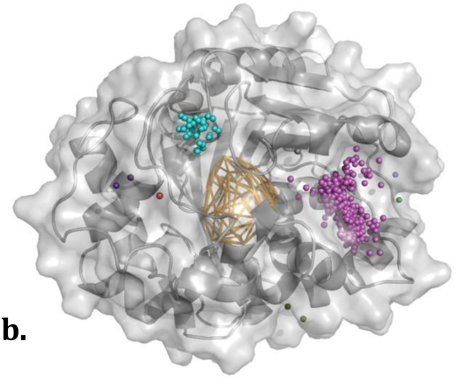


Fig. 4b.

```
[traceable_residues]
scope = backbone
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 2.8
auto_barber_mincut = None
auto_barber = protein
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = None
auto_barber_mincut = None
auto_barber_tovdw = True
```

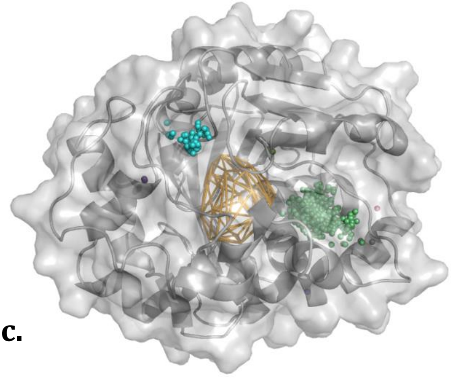


Fig. 4c.

```
[traceable_residues]
scope = name CA
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 2.8
auto_barber_mincut = None
auto_barber = protein
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = None
auto_barber_mincut = None
auto_barber_tovdw = True
```

TIPS & TRICKS - trimming paths

Auto Barber at stage separate_paths

At stage `separate_paths` the level of water inlets can be defined through Auto Barber parameters. There are several options that can be adjusted which gives many possibilities of settings. Similar to as in the `scope` definition, at this stage `auto_barber` can be set to standard values such as `protein` (default), `backbone`, `name CA` or nonstandard ones. Setting this option, however, did not significantly change the depth of water molecules in reference to the protein surface (**Figures 5a, 5b, 5c**) in comparison to the results acquired by changing the `scope` definition (**Figures 4a, 4b, 4c**). The main consequence of changing `auto_barber` option was the location of water inlets of cluster C3. At this stage, also correction of spheres radii can be switched off by setting `auto_barber_tovdw` to `False` (**Figure 5d**). Setting this option to `False` in this case not only changed the location of water inlets of cluster C3 but also caused the unification of cluster C3.

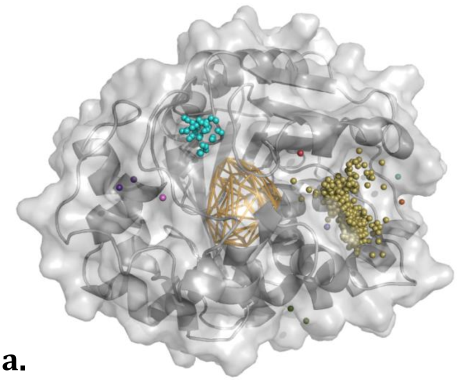


Fig. 5a.

```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 2.8
auto_barber_mincut = None
auto_barber = protein
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = None
auto_barber_mincut = None
auto_barber_tovdw = True
```

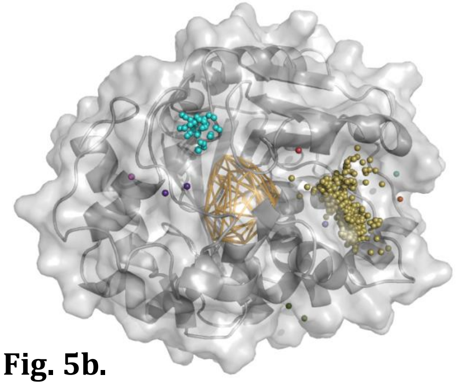


Fig. 5b.

```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 2.8
auto_barber_mincut = None
auto_barber = backbone
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = None
auto_barber_mincut = None
auto_barber_tovdw = True
```

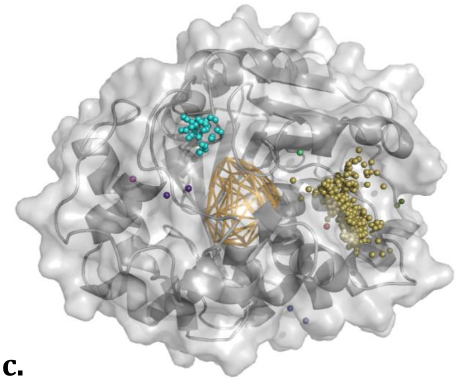


Fig. 5c.

```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 2.8
auto_barber_mincut = None
auto_barber = name CA
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = None
auto_barber_mincut = None
auto_barber_tovdw = True
```

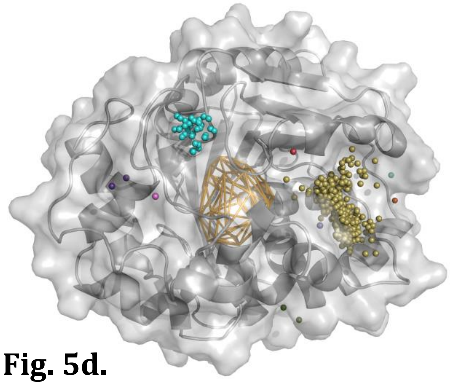
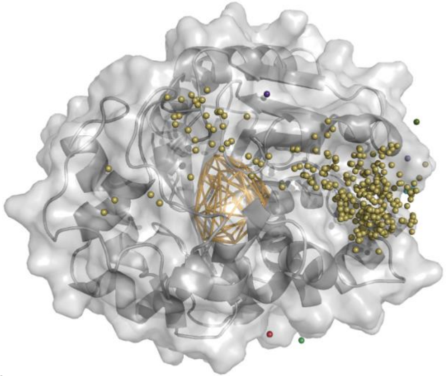


Fig. 5d.

```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = False
auto_barber_maxcut = 2.8
auto_barber_mincut = None
auto_barber = protein
 [clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = None
auto_barber_mincut = None
auto_barber_tovdw = True
```

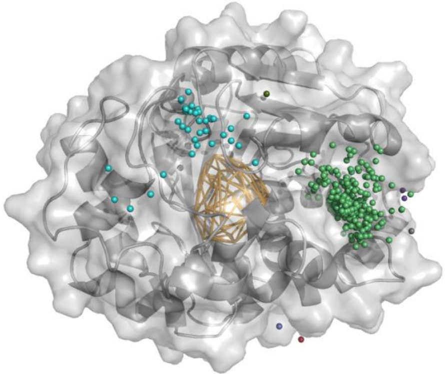
TIPS & TRICKS - trimming paths

In Auto Barber, the `maxcut` and `mincut` options can be set by the user to float values. Through these options at stage `separate_paths` dispersion of water inlets at the protein surface can be set along with the depth of water molecules location in reference to the protein surface. Setting `auto_barber_maxcut` to `0.1` caused dissipation of water inlets and formation of one large major cluster and several 1-inlet clusters (**Figure 5e**). Increasing this value to `1.0` resulted in division of dispersed water molecules into two main clusters with C2 and C3 merged together (**Figure 5f**). Setting `auto_barber_maxcut` to `2.0` caused defining of clusters C1 and C3 and segmentation of dissipated cluster C3 into three clusters (**Figure 5g**). When `auto_barber_maxcut` was increased to `4.0`, the results provided correctly defined three major clusters C1, C2 and C3 (**Figure 5h**). Augmentation of `auto_barber_maxcut` value, apart from focusing water inlets into more dense clusters, also causes diminution of distance between water inlets and defined `object`.



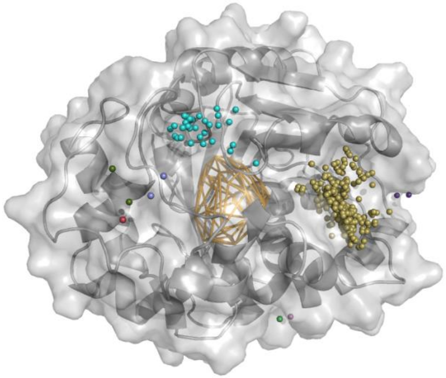
```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 0.1
auto_barber_mincut = None
auto_barber = protein
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = None
auto_barber_mincut = None
auto_barber_tovdw = True
```

Fig. 5e.



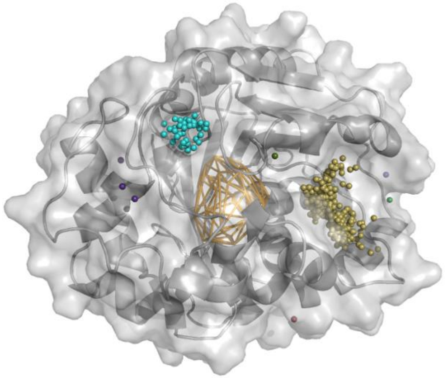
```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 1.0
auto_barber_mincut = None
auto_barber = protein
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = None
auto_barber_mincut = None
auto_barber_tovdw = True
```

Fig. 5f.



```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 2.0
auto_barber_mincut = None
auto_barber = protein
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = None
auto_barber_mincut = None
auto_barber_tovdw = True
```

Fig. 5g.



```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 4.0
auto_barber_mincut = None
auto_barber = protein
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = None
auto_barber_mincut = None
auto_barber_tovdw = True
```

Fig. 5h.

TIPS & TRICKS - trimming paths

Compared to `auto_barber_maxcut`, changing `auto_barber_mincut` value did not lead to such great dissimilarity between the results. Augmentation of this parameter causes concentration of water inlets and diminution of distances between water inlets and defined `object` (**Figures 5i, 5j**).

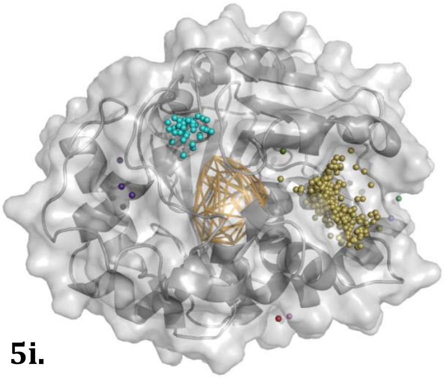


Fig. 5i.

```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = None
auto_barber_mincut = 0.1
auto_barber = protein
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = None
auto_barber_mincut = None
auto_barber_tovdw = True
```

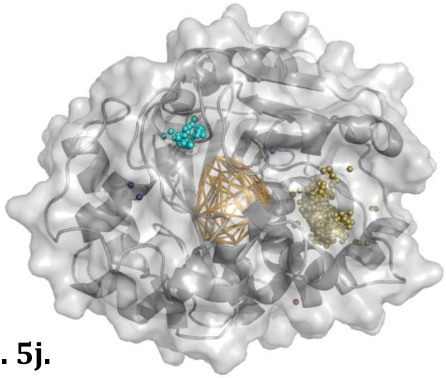


Fig. 5j.

```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = None
auto_barber_mincut = 5.0
auto_barber = protein
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = None
auto_barber_mincut = None
auto_barber_tovdw = True
```

TIPS & TRICKS - trimming paths

Stage raw_paths

In `raw_paths` stage `scope` and `object` can be redefined. It is an additional feature which is switched off on default settings. In this document we only changed the `scope` definition in order to show how results can vary by changing settings at this stage of calculations.

In comparison to changing `scope` definition to `backbone` at stage `traceable_residues` (**Figure 4b**), redefining `scope` at stage `raw_paths` gives sharper differences in results, whereas `scope` redefined as `protein` and `name CA` gave similar results. Redefining `scope` as `backbone` (**Figure 6b**) or `name CA` (**Figure 6c**) led to trimming of water paths at greater depth and boost of density of the clusters in comparison to `scope` defined as `protein` (**Figure 6a**).

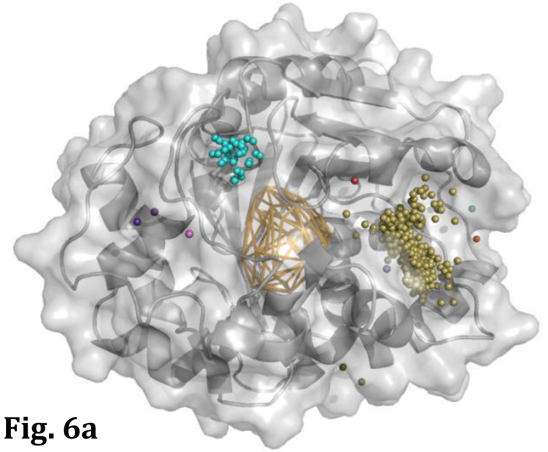


Fig. 6a

```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 2.8
auto_barber_mincut = None
auto_barber = protein
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = None
auto_barber_mincut = None
auto_barber_tovdw = True
[raw_paths]
scope = protein
```

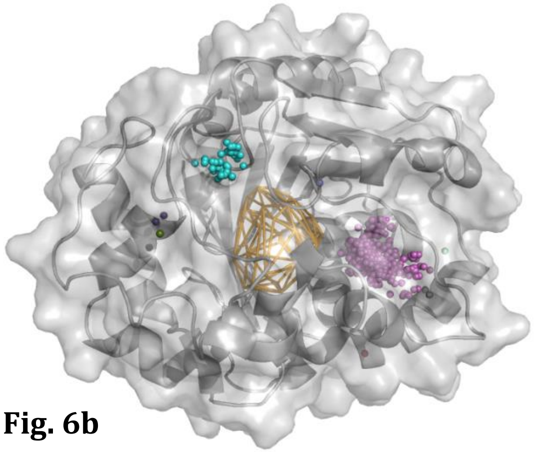


Fig. 6b

```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 2.8
auto_barber_mincut = None
auto_barber = protein
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = None
auto_barber_mincut = None
auto_barber_tovdw = True
[raw_paths]
scope = backbone
```

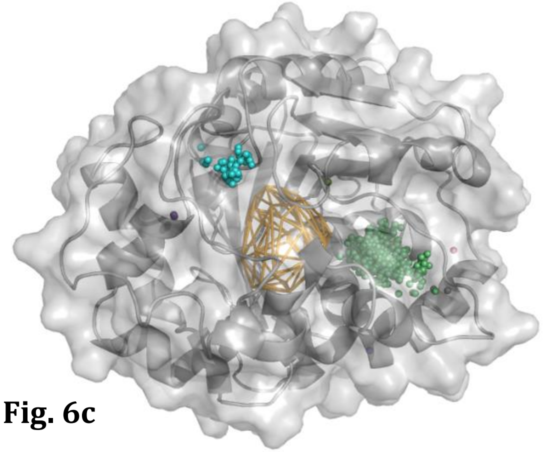


Fig. 6c

```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 2.8
auto_barber_mincut = None
auto_barber = protein
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = None
auto_barber_mincut = None
auto_barber_tovdw = True
[raw_paths]
scope = name CA
```

TIPS & TRICKS - trimming paths

Important - Auto Barber at stage clustering

At stage `clustering` Auto Barber parameters can also be set like at stage `separate_paths`, provided that the clustering method is set to `barber`, **however, at this stage the trimming of the paths cannot be defined**. The possible options at this stage are the same as in `separate_paths` but the results may differ. At this stage `auto_barber` option can be set to standard values such as `protein` (default), `backbone`, `name CA` or nonstandard ones. Changing this value in this case only caused segmentation or unification of cluster C3 (**Figures 7a, b, c**). When `auto_barber = name CA`, cluster C3 was not divided into two sections but formed one separate cluster (**Figure 7c**), unlike when this option was set to `protein` or `backbone` (**Figures 7a, 4b**). Setting `auto_barber_tovdw` to `False` also caused unification of C3 (**Figure 7d**).

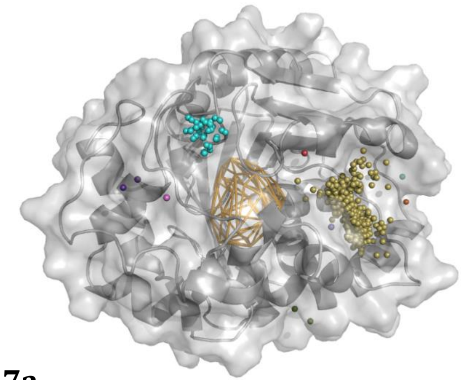


Fig. 7a.

```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 2.8
auto_barber_mincut = None
auto_barber = protein
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = None
auto_barber_mincut = None
auto_barber_tovdw = True
```

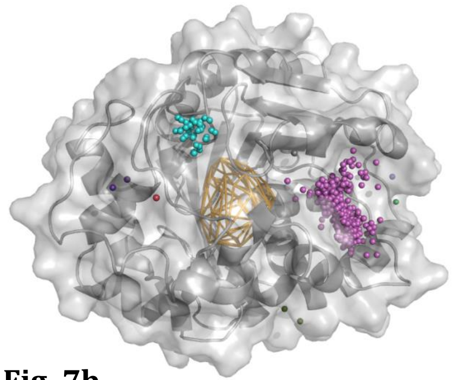


Fig. 7b

```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 2.8
auto_barber_mincut = None
auto_barber = protein
[clustering]
method = barber
auto_barber = backbone
auto_barber_maxcut = None
auto_barber_mincut = None
auto_barber_tovdw = True
```

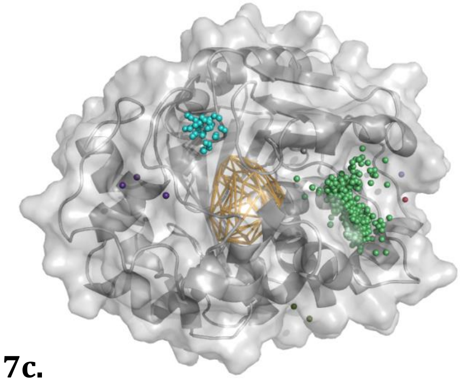


Fig. 7c.

```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 2.8
auto_barber_mincut = None
auto_barber = protein
[clustering]
method = barber
auto_barber = name CA
auto_barber_maxcut = None
auto_barber_mincut = None
auto_barber_tovdw = True
```

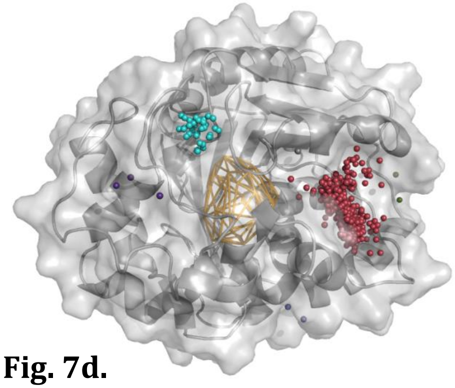


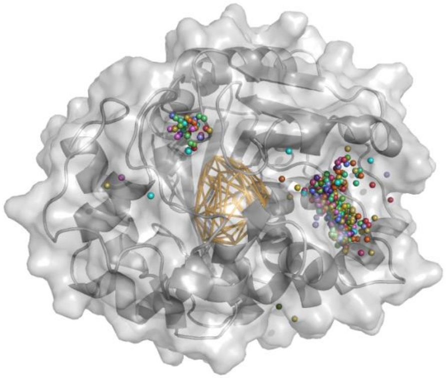
Fig. 7d.

```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 2.8
auto_barber_mincut = None
auto_barber = protein
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = None
auto_barber_mincut = None
auto_barber_tovdw = False
```

TIPS & TRICKS - trimming paths

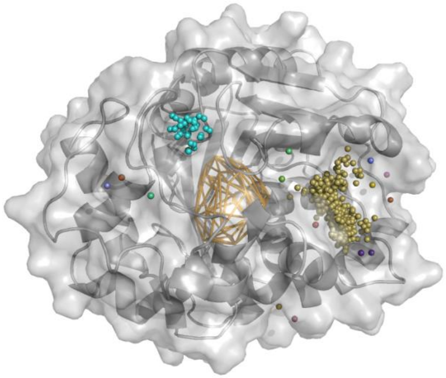
The `maxcut` and `mincut` options can be set by the user to float values. Through these options at stage `clustering` grouping of water inlets into clusters can be set.

Option `auto_barber_maxcut` can be used to divide large clusters into smaller ones. When this parameter was set to `0.1`, each water inlet was organized into one separate cluster (**Figure 7e**). Increasing the value of `auto_barber_maxcut` caused demarking of large separate clusters C1 and C2 (**Figures 7f, 7g, 7h**). Augmentation of this value from `2.0` to `6.0` did not, however, lead to unification of cluster C3, but it remained divided into two sections (**Figures 7g, 7h**).



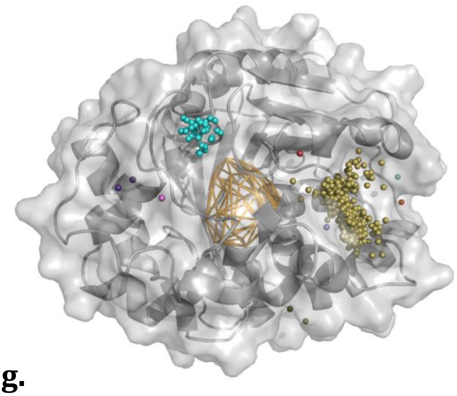
```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 2.8
auto_barber_mincut = None
auto_barber = protein
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = 0.1
auto_barber_mincut = None
auto_barber_tovdw = True
```

Fig. 7e.



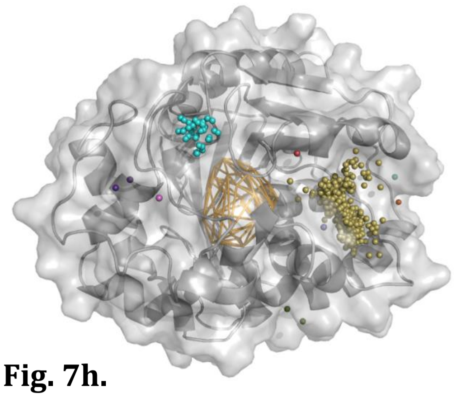
```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 2.8
auto_barber_mincut = None
auto_barber = protein
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = 1.0
auto_barber_mincut = None
auto_barber_tovdw = True
```

Fig. 7f.



```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 2.8
auto_barber_mincut = None
auto_barber = protein
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = 2.0
auto_barber_mincut = None
auto_barber_tovdw = True
```

Fig. 7g.



```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 2.8
auto_barber_mincut = None
auto_barber = protein
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = 6.0
auto_barber_mincut = None
auto_barber_tovdw = True
```

Fig. 7h.

TIPS & TRICKS - trimming paths

A parameter that can be used to merge separate clusters is `auto_barber_mincut`. Augmentation of this option’s value leads to fusion of major clusters. Setting `auto_barber_mincut` to `1.0` caused unification of cluster C3. (**Figure 7j**). When this option was set to `4.0` clusters C2 and C3 were merged into one cluster (**Figure 7k**). Further increase of this value to `6.0` resulted in unification of all major clusters C1, C2, C3 along with the minor ones into one cluster (**Figure 7l**).

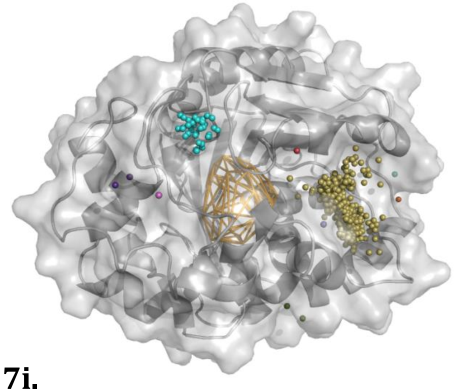


Fig. 7i.

```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 2.8
auto_barber_mincut = None
auto_barber = protein
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = None
auto_barber_mincut = 0.1
auto_barber_tovdw = True
```

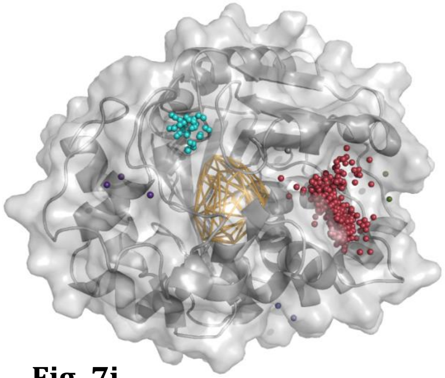


Fig. 7j.

```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 2.8
auto_barber_mincut = None
auto_barber = protein
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = None
auto_barber_mincut = 3.0
auto_barber_tovdw = True
```

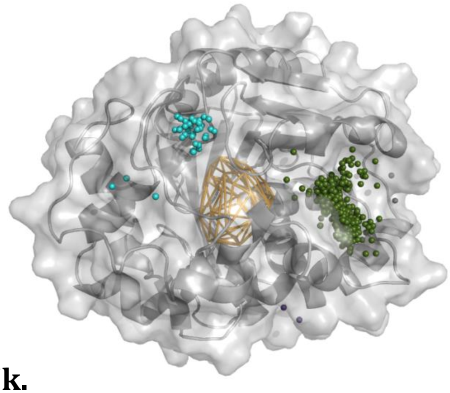


Fig. 7k.

```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 2.8
auto_barber_mincut = None
auto_barber = protein
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = None
auto_barber_mincut = 4.0
auto_barber_tovdw = True
```

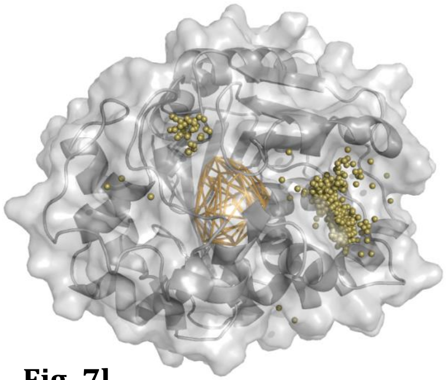


Fig. 7l.

```
[traceable_residues]
scope = protein
[separate_paths]
auto_barber_tovdw = True
auto_barber_maxcut = 2.8
auto_barber_mincut = None
auto_barber = protein
[clustering]
method = barber
auto_barber = protein
auto_barber_maxcut = None
auto_barber_mincut = 6.0
auto_barber_tovdw = True
```

TIPS & TRICKS - trimming paths

Our tips

- Defying or redefying `scope` leads to adjustment of level at which water paths are trimmed what can be used to obtain quick changes in acquired results,

- The user should set `scope` to `protein`, if they want the inlets to be located at a great distance from the defined `object`, to `name CA`, to have them grouped in the hollows and near entries of tunnels, or to `backbone`, to balance the distance between water inlets and `object`,

- Auto Barber parameters work differently at respective stages of calculations, at stage `clustering` these options are only responsible for organization of water inlets into separate clusters, whereas at stage `separate_paths` they additionally change the level at which water paths are trimmed and therefore the density of the clusters can be set through them,

- At stage `separate_paths` decrease of `auto_barber_maxcut` value, apart from dispersing water inlets at the protein surface, also causes augmentation of distance between water inlets and defined `object`, whereas increase of `auto_barber_mincut` leads to concentration of water inlets and diminution of the level at which water paths are trimmed

Auto Barber at stage clustering tip

- At stage `clustering` decrease of `auto_barber_maxcut` value leads to division of large clusters into smaller ones, whereas augmentation of `auto_barber_mincut` can be used to merge clusters.

TIPS & TRICKS – trimming paths

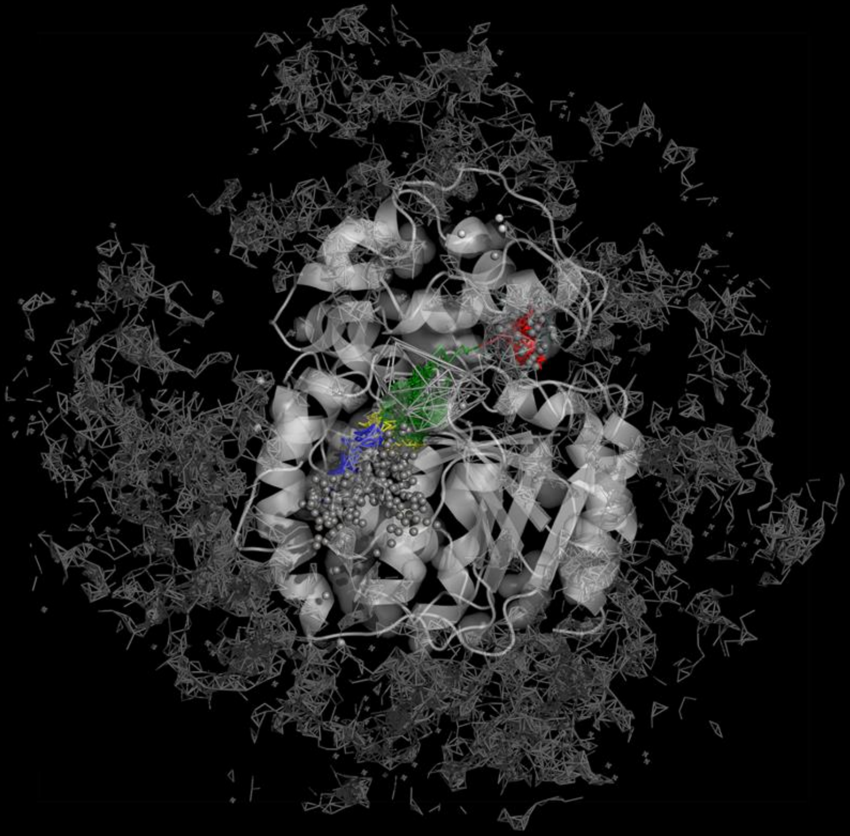
AQUA DUCT

Instalation and guide:

http://www.aquaduct.pl

More info:

info@aquaduct.pl



TG

Tunneling Group
Biotechnology Centre
Silesian University of Technology
ul. Krzywoustego 8
44-100 Gliwice
POLAND