Modifying a Pharmacophore Query in CSD-CrossMiner
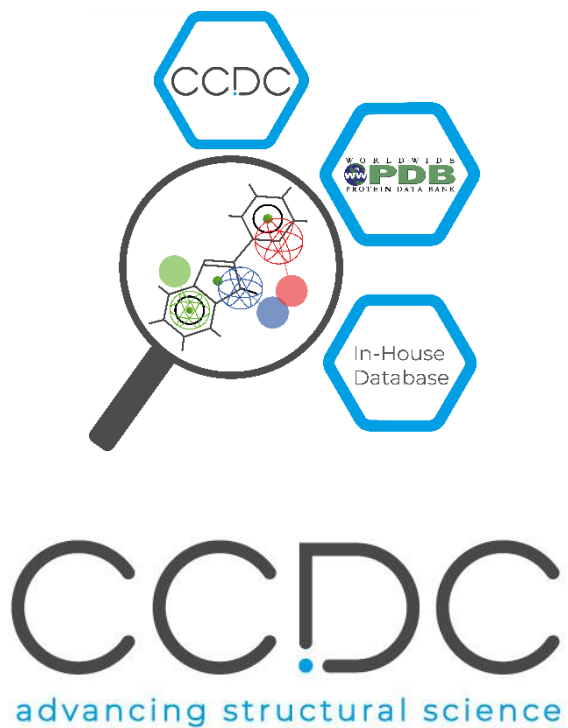
2021.2 CSD Release

Table of Contents

Introduction

CSD-CrossMiner can be thought of as a pharmacophore-based query tool. However, it is much more powerful than traditional pharmacophore query tools as it allows you to query not only databases of ligands, but also proteins and protein-ligand interactions. CSD-CrossMiner includes a preconfigured database of biologically relevant subsets of the Cambridge Structural Database (CSD) and the Protein Data Bank (PDB). The pharmacophore used in the query is interactive, allowing you to easily edit it and in a number of ways through a simple user interface. This delivers an overall interactive search experience with application areas in interaction searching, scaffold hopping or the identification of novel fragments for specific protein environments.

We provide a feature database that can be downloaded from the CSD-CrossMiner application or by accessing to the 'Data & Software Updates' section of our Downloads page.

The supplied feature database contains a subset of the CSD and PDB. The CSD subset consists of structures which are organic plus a small list of transition metals i.e., Mn, Fe, Co, Ni, Cu, Zn, have an R-factor of at maximum 10%, have 3D coordinates, have no disorder, and are not polymeric (more than 400 000 structures total). The supplied PDB database is divided in two subsets, one composed by protein-ligand complexes and another subset composed by protein-ligand-nucleic acids complexes. For the PDB subsets only the protein-ligand binding site and protein-ligand-nucleic acid binding site is provided, where the binding site is defined as all molecules with an atom within a 6Å radius around the ligand (> 300 000 binding sites). For further discussion, please refer to the CSD-CrossMiner User Guide or the original paper: Korb O *et al.,* "Interactive and Versatile Navigation of Structural Databases" *J Med Chem*, **2016**, 59(9):4257, DOI: 10.1021/acs.jmedchem.5b01756.

 This tutorial is geared towards the novice CSD-CrossMiner user who has Life Science experience. It covers the primary features, in searching across ligands, proteins, and ligand-protein interactions. Some of the results may vary depending on your version of CSD-CrossMiner. When you have completed this tutorial, you should be able to modify and save a pharmacophore query and choose a hit as a reference molecule for later usage.

Overview of CSD-CrossMiner

CSD-CrossMiner Terminology

CSD-CrossMiner uses several terms, some common to the field of drug discovery, and some not. For reference, these terms are defined as below:
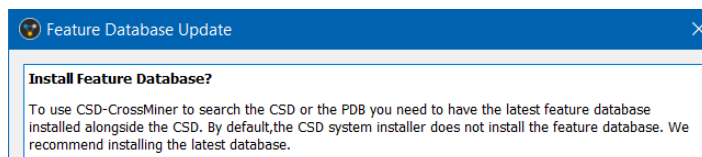
Features: can be defined as an ensemble of steric and electronic features that characterise a protein and/or a small molecule. In CSD-CrossMiner a feature is defined as point(s), centroid or vector which represent a SMARTS query and, in the case of a vector, this includes geometric rules.

Pharmacophore point: is a feature that has been selected to be part of a pharmacophore because its presence is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger or block its biological response.

Structure database: is a database containing the 3D coordinates of small molecule structures and/or protein-ligand binding sites. This database is used to create a feature database.

Feature database: is a database containing the structures from the structure database, indexed with a set of feature definitions provided by CSD-CrossMiner and any additional features defined by the user. This is the database that CSD-CrossMiner uses to perform the actual 3D search against a pharmacophore query.

Exit vector: is a two-point feature that represents a single, non-ring bond between two heavy atoms features, and it will be represented as two mesh spheres. In the case of CSD-CrossMiner, directionality in an exit vector does not matter.

Feature Database Update ✕

Install Feature Database?

To use CSD-CrossMiner to search the CSD or the PDB you need to have the latest feature database installed alongside the CSD. By default,the CSD system installer does not install the feature database. We recommend installing the latest database.

CSD-CrossMiner is a powerful tool with a simple user interface. This quick section will familiarise you with the basic functions and underlying data components before moving on to exploring some scientific questions.

1. Launch CSD-CrossMiner clicking on the CSD-CrossMiner icon: .

2. If it is the first time you opened the application, *Feature Database Update* pop-up window will guide you through the process of downloading the feature database. Clicking on **Install** will start the downloading process and when completed the database will be automatically loaded in CSD-CrossMiner.

Note that the speed of the download depends on the quality of your network.
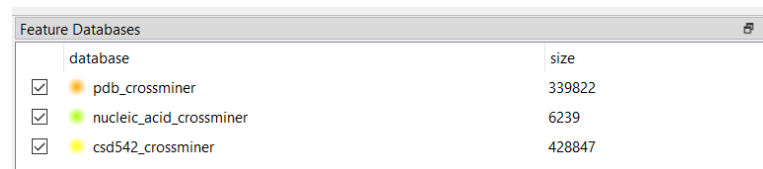
3. The *csd_pdb_crossminer.feat* feature database, will be saved in the `crossminer_data` folder in `CSD_2021` directory. The location of the loaded feature database will be automatically remembered between separate CSD-CrossMiner sessions.

    If you have alredy downloaded the supplied feature database, it will be automatically loaded when you launch CSD-CrossMiner.

Loading will take a few minutes, but even once the bar hits 100%, it will need a moment to initialise the structures.

4. Once loaded, you will see the CSD and PDB (*pdb_crossminer* and *nucleic_acid_crossminer)* databases listed in the *Feature Databases* window. You can load multiple databases and use the tick boxes to indicate which database should be searched.

5. You will also see a list of features in the bottom right *Pharmacophore Features* window. These are the features used to generate these databases. The features with the *show in reference* tick-box toggled are displayed in the 3D view.
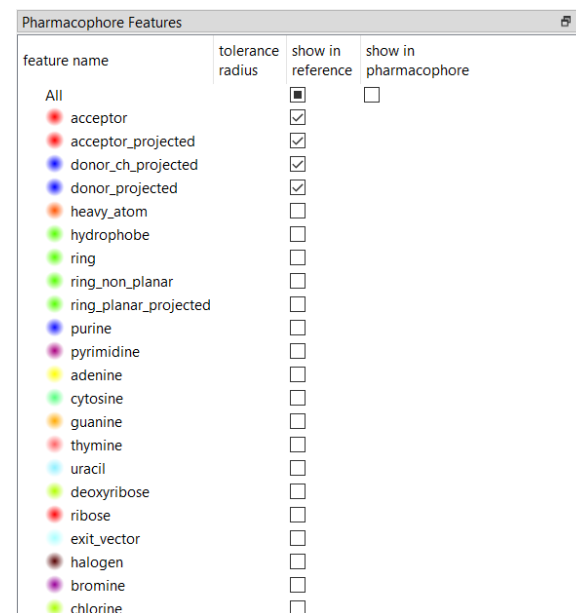
4

Feature Databases	
database	size
☑ ● pdb_crossminer	339822
☑ ● nucleic_acid_crossminer	6239
☑ ● csd542_crossminer	428847

5

Pharmacophore Features			
feature name	tolerance radius	show in reference	show in pharmacophore
All		◼	☐
● acceptor		☑	
● acceptor_projected		☑	
● donor_ch_projected		☑	
● donor_projected		☑	
● heavy_atom		☐	
● hydrophobe		☐	
● ring		☐	
● ring_non_planar		☐	
● ring_planar_projected		☐	
● purine		☐	
● pyrimidine		☐	
● adenine		☐	
● cytosine		☐	
● guanine		☐	
● thymine		☐	
● uracil		☐	
● deoxyribose		☐	
● ribose		☐	
● exit_vector		☐	
● halogen		☐	
● bromine		☐	
● chlorine		☐	

Features and Pharmacophore Representation

In the CSD-CrossMiner 3D view, features are represented as small translucent spheres coloured as defined in the *Pharmacophore Features* window. A pharmacophore point is represented as a mesh sphere which reflects the uncertainty in the position of the pharmacophore point. In the 3D view:
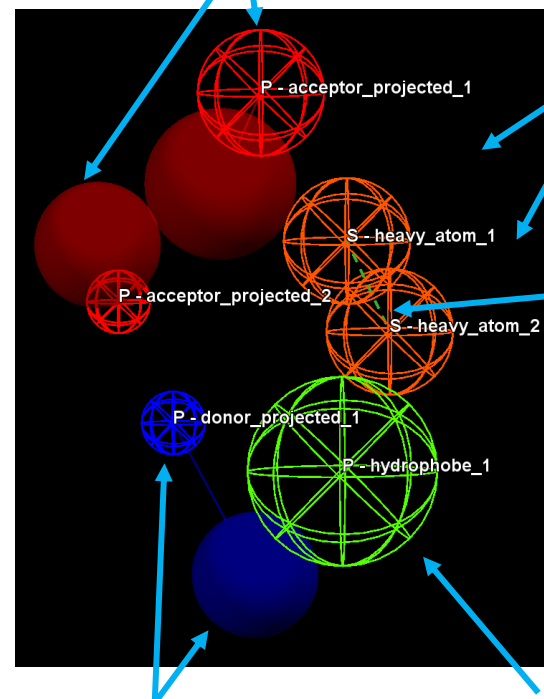
- **P**: Protein feature
- **S**: Small molecule feature
- **Dashed line**: intra and intermolecular constraints. Constrained features must belong to either the same molecule as each other (*intra*, dashed green line) or different molecules (*inter*, dashed red line).
- **Mesh sphere**: the actual feature itself, where the sphere size represents the radius of tolerance.
- **Solid sphere**: the projected virtual point to represent the directionality of a hydrogen bond acceptor/donor. A feature can have more than one projected point. For example, a H bond acceptor can have multiple potential lone pair preferred projections.

Note that the colour coding of the pharmacophore points is defined in the *Pharmacophore Features* browser; e.g. hydrophobe features are green, hydrogen bond acceptors are red, and so on.

In the directional features, the mesh sphere (the actual feature itself) is defined as *B* in the *Pharmacophore Features* window (**B**ase feature), and the projected virtual point representing the directionality of the feature is defined as *V* (**V**irtual point).

Protein (P): H bond acceptor feature (mesh) with projected directionality (solid)

Small molecule (S): heavy atoms

Intramolecular constraint

Protein (P): H bond donor feature (mesh) with projected directionality (solid)

Protein (P): hydrophobe feature

feature name	tolerance radius	show in reference	show in pharmacophore
All		☐	☑
acceptor		☐	
acceptor_projected		☐	
acceptor_projected_1			☑
B	1.00		
V	1.00		
acceptor_projected_2			☑
B	0.50		
V	1.00		

Modifying a Pharmacophore Query

One of the powers of CSD-CrossMiner is the ability to manually interact with the pharmacophore query and edit a pharmacophore point at any time (even while a search is running). For this example, you will be editing the cathepsin L pharmacophore that is provided in the CSD-CrossMiner installation folder.

If you already have work in CSD-CrossMiner, close it by clicking **File** > **Close Pharmacophore** and/or **File** > **Close Reference**.

1. Load the cathepsin L pharmacophore by clicking on **File** > **Load Pharmacophore**… and select:

```
<CCDC installation
folder>\Discovery_2021\CSD_CrossMiner\example_pharmacophore
s\catl_s3.cm
```

2. Right click on the hydrophobe feature (green in the 3D view) to bring up the pharmacophore context menu. Click **Morph Into** and then **water**. You will notice that the molecule type also changed from **Protein (P)** to **Small Molecule (S),** as water is not considered part of the protein.

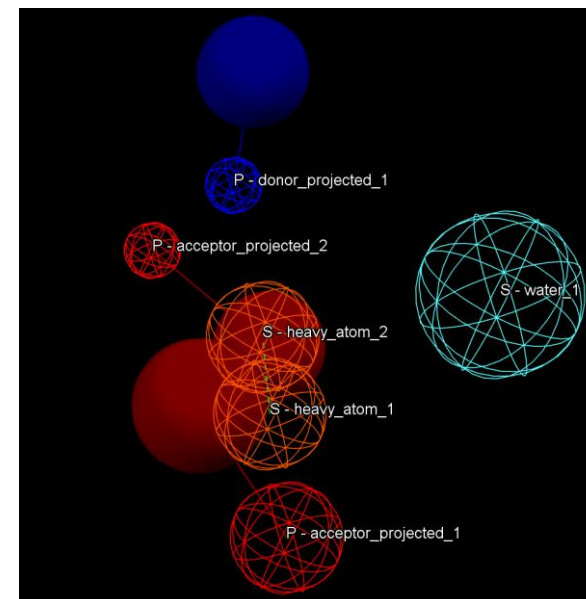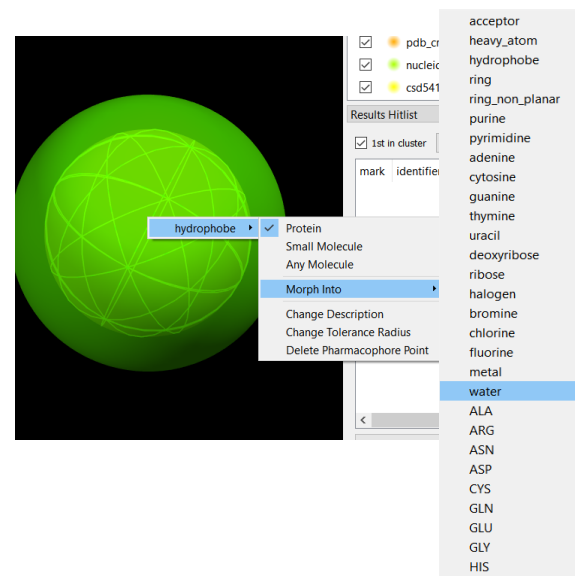Through the pharmacophore context menu for a feature it is possible to:

- Define where the feature belongs: a **Protein**, a **Small Molecule**, or **Any**.
- Change the feature type (**Morph Into**).
- Change the label of the feature (**Change Description**).
- Change the tolerance radius of the pharmacophore point (**Change Tolerance Radius).**
- Delete the pharmacophore point (**Delete Pharmacophore Point**).

If a reference molecule is loaded, you have the additional option to:

- Move the pharmacophore point to the nearest atom (**Snap To Atom**).

Note that the context menu of small molecule (S) pharmacophore points, there is the additional option to:

- Add constraints (**Constrain To**).

3. Double-click on the radius size of virtual *acceptor_projected_1* (*V*) in the *Pharmacophore Features* window or use **Change Tolerance Radius** from the pharmacophore context menu to change the radius of virtual *acceptor_projected_1 (V)* from *1.20* to *1.00*. Then change the *water* radius from *1.50* to *1.00*. This will reduce the uncertainty in the position of the water pharmacophore point.

4. Be sure that the **1st in cluster** tick-box is checked, this will show only the cluster representatives is in the *Results Hitlist* window. Then start the search by clicking on the **Play** button: ▶ in the *Search:* toolbar, the matched hits will populate the *Results Hitlist* window, as well as the 3D view. The progress bar at the bottom of the *Results Hitlist* window show the total number of hits.

The search may require several minutes; however, let it go to completion. It is possible to change the number of processors dedicated to the pharmacophore search by changing the **Number of threads** in **Edit** > **Options.**

Note that the pharmacophore search options (**Options**) are not available to be changed (greyed-out) when the pharmacophore search is running or paused.

5. Locate and visualise the result with the lowest RMSD by clicking on the *rmsd* column in the *Results Hitlist* window, to show ascending order. You may wish to show the 2D diagram of the matched hits by right-clicking in the *Results Hitlist* window and then click on **+ diagram**.

CSD-CrossMiner allows you to visually interact with the pharmacophore:

- You can edit the pharmacophore points in the *Pharmacophore Features* window or in the 3D view (changing the pharmacophore tolerance, molecule type, pharmacophore type, etc.).

- You can move and resize pharmacophore points from the 3D view. All changes are in real time and have immediate feedback on your query and on the hits found.
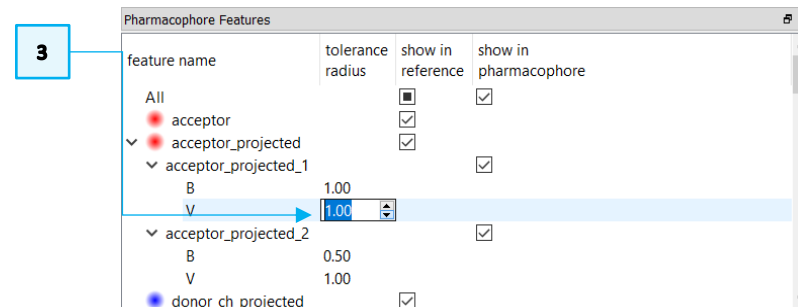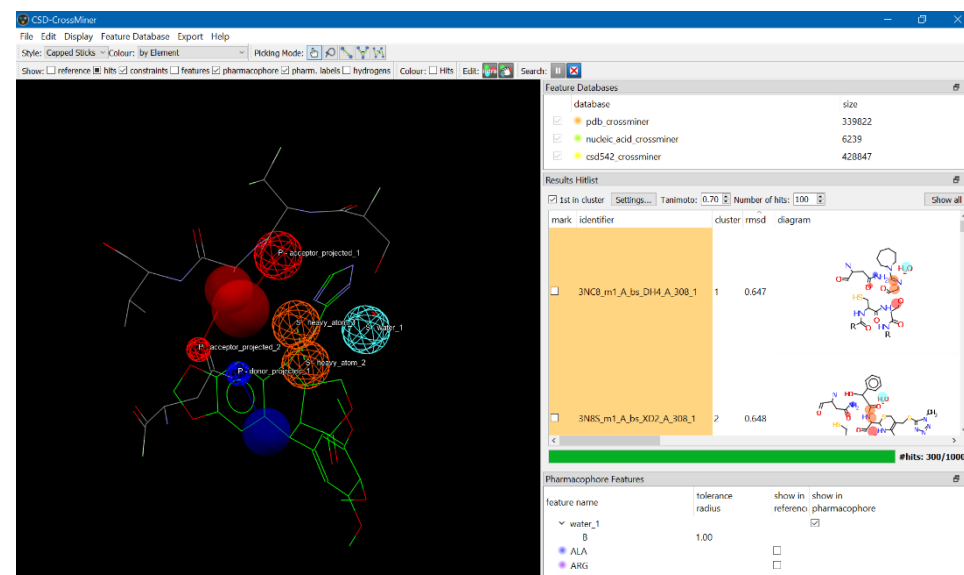
To do so:

6. Switch to interactive editing by click on the **Pharmacophore editing** button
    in the *Edit:* toolbar (an open hand means editing is off, and a closed hand
    means editing is on). This will turn the mouse cursor in a small hand.

7. If you have a three-button mouse: hover over the water feature, then click
    and drag using the middle mouse button (MMB). This will change the radius
    of the water sphere and automatically restart the search.

    If you do not have a three-button mouse or you have trouble getting this to
    work, you can use the *Pharmacophore Features* window (See point 3) to
    change the radius of the water pharmacophore sphere.

8. Change the water radius back to *1.00* using **Pharmacophore editing** button
    or from the *Pharmacophore Features* window.

Whenever you edit the pharmacophore, the current results in the *Results Hitlist*
window and in the 3D view will disappear (as a new search will start). If you want
to edit a pharmacophore overlaid with a molecule from the hits, you will need to
first set the molecule as a reference.

9. Right-click on one of the hits in the *Results Hitlist* window and click **Use as
    reference**. This will load the molecule into the 3D view and will show the
    donor and acceptor features associated with the molecule.

Note that if a different choice of displayed features was made during the CSD-
CrossMiner session, those features (if present in the reference molecule) will be
displayed instead.

10. Display all the features associated with the reference structure by clicking on the *feature* ■ tick-box in the CSD-CrossMiner *Show:* toolbar (which will turn to ☑) or alternatively by ticking the tick-box for *All* features in the *show in reference* column in the *Pharmacophore Features* window.

11. Hide all features by clicking on features ☑ (which will turn to ☐) or alternatively by unticking the *All* tick-box in the *Pharmacophore Features* window.

12. Change the style of the small molecule in the 3D view by left-clicking on one of the small molecule atoms while pressing **Shift** key. This will select all the atoms of the same molecule (the selected atoms are highlighted with small yellow spheres). Right-click on one of the selected atoms and pick **Styles**, from the pull-down menu and select the desired style.

13. With **Pharmacophore editing** mode on ( ), drag one of the heavy atom pharmacophore points (in orange in the 3D view) to a nearby carbon atom using the left mouse button (LMB).

Note that the search starts as soon as you let it go. The reference molecule stays visible even while the search runs.

14. To undisplay the reference molecule, untick the **reference** tick-box in the *Show:* toolbar. This *Show:* toolbar controls the visibility of the reference molecule, hits, constraints, features, pharmacophore, pharmacophore labels and hydrogens.

Every CSD-CrossMiner user has accidently grabbed a pharmacophore point and dragged it across the screen when they intended to rotate the molecule viewer. In such cases **Ctrl + z** (or **Edit** > **Undo**) is very helpful. It will undo the last change made to the pharmacophore. However, the search will start over again. Therefore, be sure to turn off the **Pharmacophore editing** mode when you are

not using it (open hand ![button] button).

15. Press ![X] to stop and erase the pharmacophore search.

16. You can also add new features from the *Pharmacophore Features* window. Scroll to fluorine in the feature browser and right click, then click **Create fluorine**. This will drop the new feature into the 3D view.

17. Align the new fluorine near to the closest atom by right-clicking on the fluorine feature and select **Snap To Atom**.

18. Start the search by clicking on ![play]. This new search probably won't find any results. This step was just to demonstrate how to add and manoeuvre new features.

19. Save your edited pharmacophore by clicking **File** > **Save Pharmacophore**. Name the file *pharmacophore_edited.cm.*

There are two options for saving a pharmacophore underneath the **File** menu, as defined below:


Save Pharmacophore: Will save the current pharmacophore in a CSD-CrossMiner file format (*cm*). This is a text file with all feature definitions and the (x, y, z) of the instances used.

Save PyMOL Pharmacophore: Will save the current pharmacophore as a python script to be run in PyMOL. This will create graphic objects in PyMOL in the same coordinate frame as the saved hits.
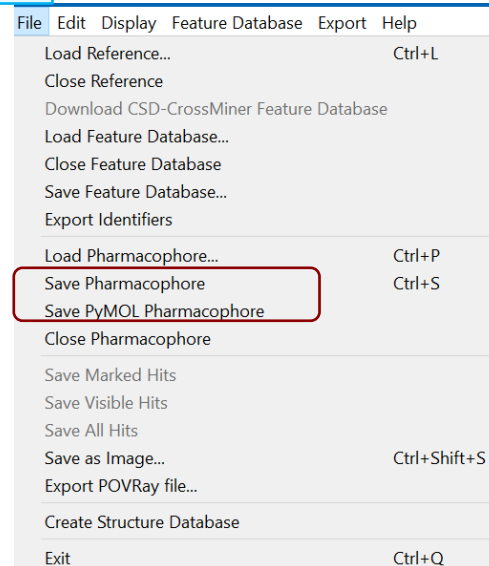
This ends the tutorial.
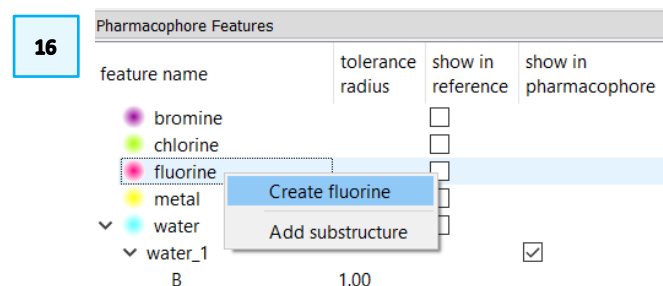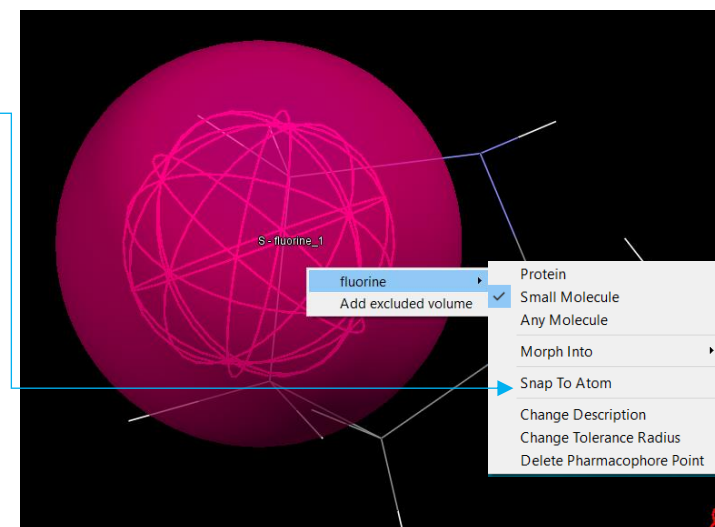
Feedback

We hope this workshop improved your understanding of CSD-CrossMiner and you found it useful for your work. As we aim to continuously improve our training materials, we would love to get your feedback. Click on [this link](#) to a survey (link also available from workshops webpage), it will take less than 5 minutes to complete. The feedback is anonymous. You will be asked to insert the workshop code, which for this self-guided workshop is CROSS-003. Thank you!