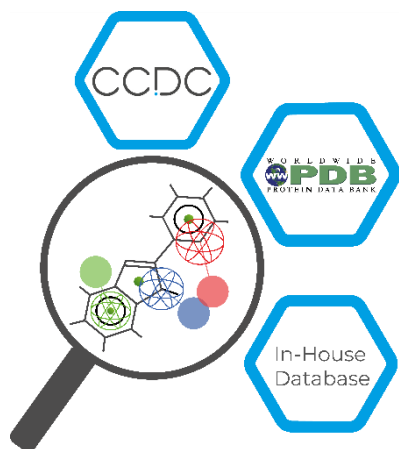


Performing a Pharmacophore Search using CSD-CrossMiner

2020.3 CSD Release



CCDC
advancing structural science

Table of Contents

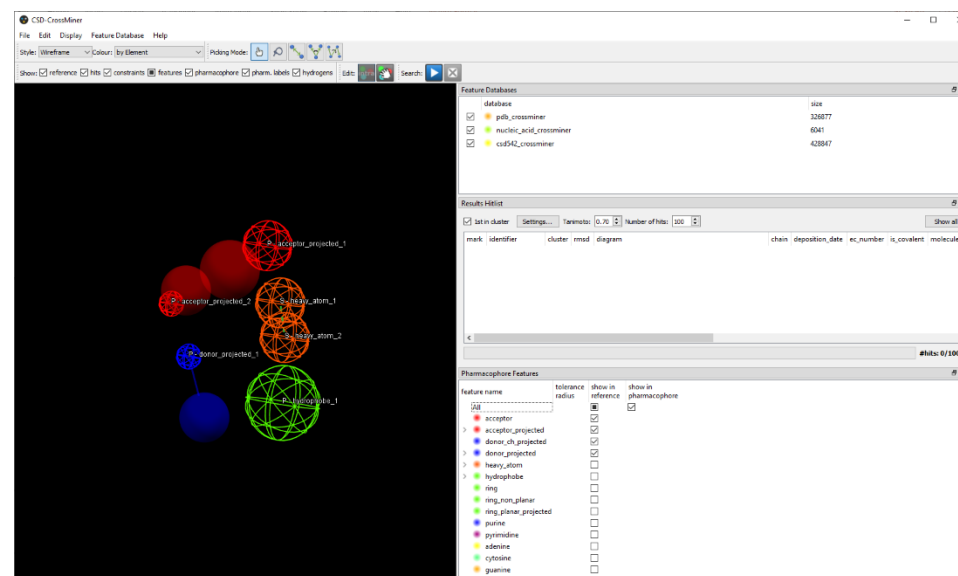
Introduction.....	2
Objectives	3
Pre-required skills.....	3
CSD-CrossMiner Terminology	3
Overview of CSD-CrossMiner	4
Exercise 1: Searching with a Pharmacophore	5
Exercise 2: Modifying a pharmacophore query	10
Bonus exercise: Adding new pharmacophore features	15
Conclusions.....	16
Next steps.....	16
Feedback	16

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.

The subset of the CSD included with CSD-CrossMiner 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 included PDB database is divided in two subsets, one composed by of all 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 (> 320 000 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.

Before beginning this workshop, ensure that you have installed and activated CSD-CrossMiner. Please contact your site administrator or workshop host for further information.



Objectives

In this workshop we will learn to build, edit, and run a pharmacophore search using CSD-CrossMiner. In particular, we will:

- Familiarise with the interface.
- Learn how to build queries.
- Learn how to execute pharmacophore searches.
- Learn how to save hits.

This workshop will take approximately **30 minutes** to be completed, excluding the bonus questions and exercises.

Some of the results may vary depending on your version of CSD-CrossMiner.

Note: The [CSD-CrossMiner Terminology box](#) on the righthand side contains useful terminology for the exercises and for the demo. Make sure you are familiar with these terms before starting the exercises.

Pre-required skills

This tutorial is geared towards the novice CSD-CrossMiner user who has Life Science experience.

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.

Overview of CSD-CrossMiner




CSD-CrossMiner is a powerful tool with a simple user interface. This brief overview will introduce you to 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: . The *csd_pdb_crossminer.feet* feature database, provided in the CSD package (\$CSD_2021/crossminer_data), is loaded as default feature database.

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

2. 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.























2

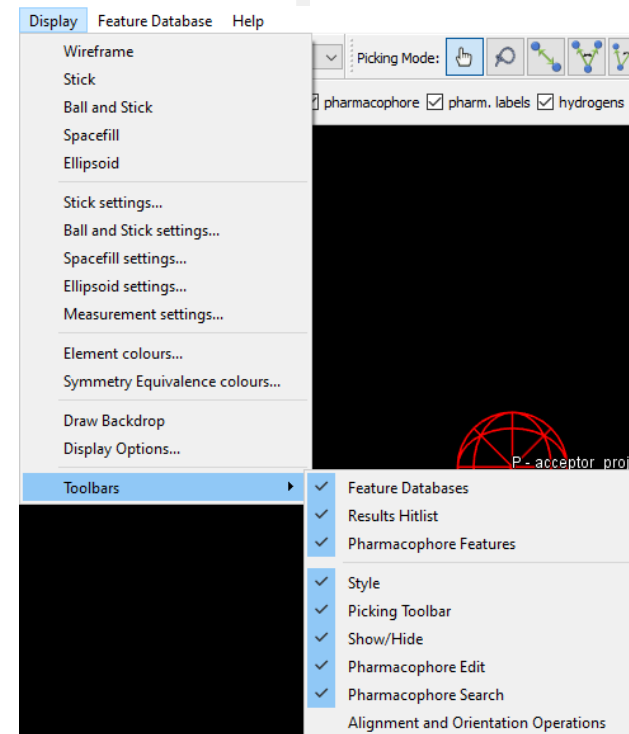
database	size
<input checked="" type="checkbox"/>  pdb_crossminer	326877
<input checked="" type="checkbox"/>  nucleic_acid_crossminer	6041
<input checked="" type="checkbox"/>  csd542_crossminer	428847

3. 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. If these windows are missing, go to **Display>Toolbars** and make sure **Feature Databases** and **Pharmacophore Features** are ticked.

3

feature name	tolerance radius	show in reference	show in pharmacophore
All		<input checked="" type="checkbox"/>	<input type="checkbox"/>
 acceptor		<input checked="" type="checkbox"/>	
 acceptor_projected		<input checked="" type="checkbox"/>	
 donor_ch_projected		<input checked="" type="checkbox"/>	
 donor_projected		<input checked="" type="checkbox"/>	
 heavy_atom		<input type="checkbox"/>	
 hydrophobe		<input type="checkbox"/>	
 ring		<input type="checkbox"/>	
 ring_non_planar		<input type="checkbox"/>	
 ring_planar_projected		<input type="checkbox"/>	
 purine		<input type="checkbox"/>	
 pyrimidine		<input type="checkbox"/>	
 adenine		<input type="checkbox"/>	
 cytosine		<input type="checkbox"/>	
 guanine		<input type="checkbox"/>	
 thymine		<input type="checkbox"/>	
 uracil		<input type="checkbox"/>	
 deoxyribose		<input type="checkbox"/>	
 ribose		<input type="checkbox"/>	
 exit_vector		<input type="checkbox"/>	
 halogen		<input type="checkbox"/>	
 bromine		<input type="checkbox"/>	
 chlorine		<input type="checkbox"/>	



Exercise 1: Searching with a Pharmacophore

In this tutorial we will learn how to run a pharmacophore search in CSD-CrossMiner using one of the pharmacophore examples provided in the CSD-CrossMiner installation folder.

Human cathepsin L plays a major role in protein catabolism and is implicated in several pathological processes. As such, it is a common research target and serves as a good example for novel drug discovery. For this example, we will be using a pharmacophore included with CSD-CrossMiner to explore the CSD and PDB for possible hits.

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

[CSD-CrossMiner installation]\example_pharmacophores\catl_s3.cm

2. This will load the pharmacophore into the viewing area. Take a moment to rotate the pharmacophore and understand the different pharmacophore feature points:

P: Protein feature

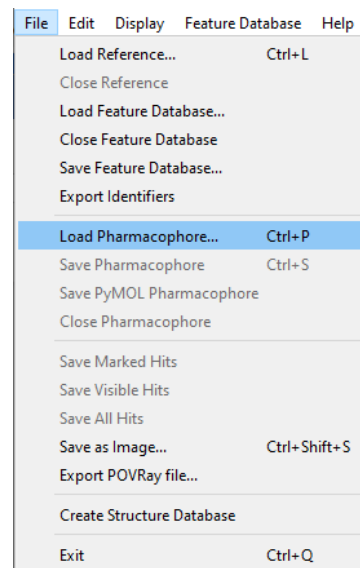
S: Small molecule feature

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.

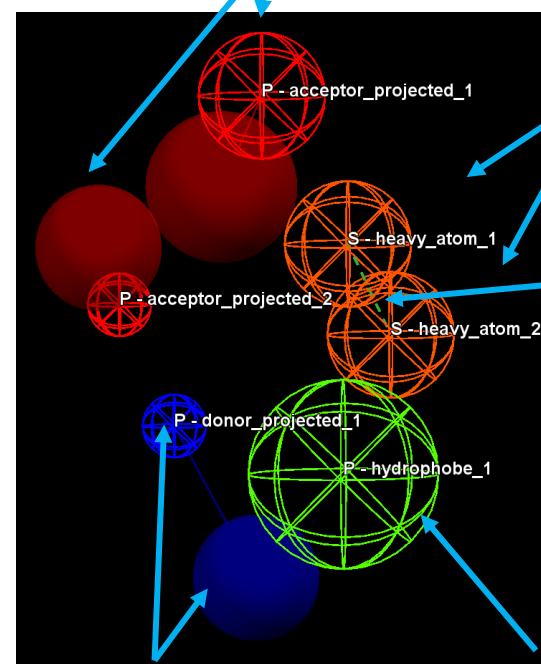
Solid sphere: the projected virtual point to represent the directionality of a hydrogen bond acceptor/donor. A feature can have more than one projected point. For example, a H bond acceptor can have multiple potential lone pair preferred projections.

1



1a9u_lig_no_surface.cm
11b_hsd1_csd.cm
aromatic_box.cm
catl_s3.cm
intramolecular_2hzi_csd.cm
intramolecular_2hzi_pdb.cm
p38_back_pocket.cm

2



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

Note that the colour coding is defined in the *Pharmacophore Features* browser. E.g., hydrophobe features are green, hydrogen bond acceptors are red, and so on. The pharmacophore in your viewer has only one projected H bond donor. These points correlate to the feature browser: B indicates the **B**ase feature, and V indicates the accompanying **V**irtual point.

- Sphere size is directly correlated to tolerance: the smaller the sphere the lower the tolerance for geometric matching. The radius of each feature is expressed in Å is shown in the *Pharmacophore Features* window.
- Note that the pharmacophore features can be hidden by unticking the *All* tick-box in the *show in pharmacophore* column. For this example, we will keep it ticked.
- Click the **Play** button  to begin searching across the databases for matches. As the search runs, you will see results populating the *Results Hitlist* window, as well as the 3D view and the progress bar with total number of hits listed at the bottom of the *Results Hitlist* window.

This search returns a high number of hits and may require several minutes.

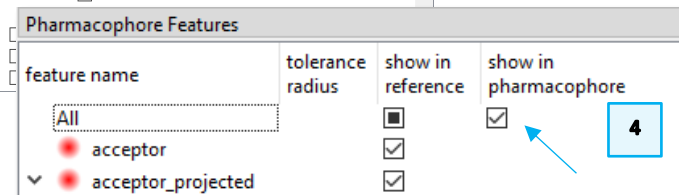
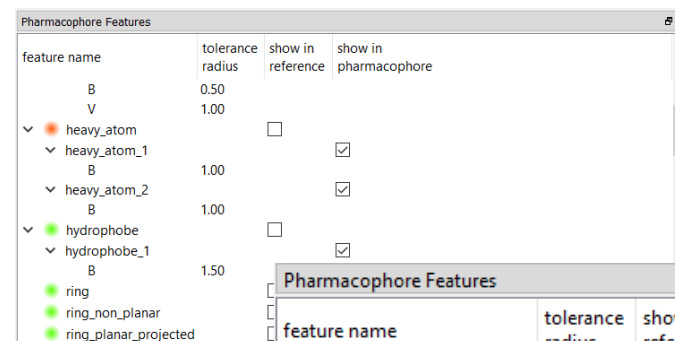
- You can use the **Pause**  button to pause the search when you have a few hundreds of hits. Do not stop the search (pressing ) , as that stops the entire process and removes all hits.

By default, all results are overlaid in the 3D view, which provides an easy appreciation of the common motifs matching the pharmacophores. However, you can view one hit at the time by clicking on each result in the *Results Hitlist* window. Or, hold down the *Shift* or *Ctrl* keys to select multiple.

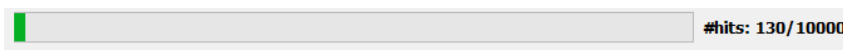
You can go back to displaying all results by clicking **Show All** in the *Results Hitlist* window.

- For clarity, disable the **hydrogens** tick-box in the *Show:* toolbar. This will hide the hydrogen atoms of the matched hits in the 3D view.

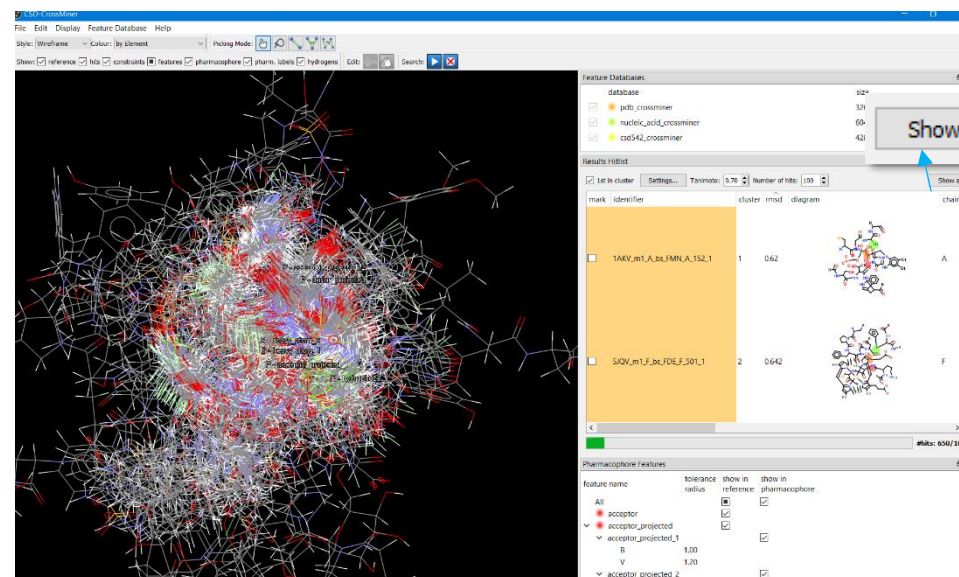
3



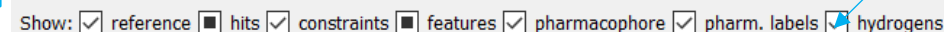
5



6



7




8. Locate the result with the lowest RMSD by clicking on the *rmsd* column in the *Results Hitlist* window, to show ascending order. The lowest RMSD result in this example is 1AKV_m1_A_bs_FMN_A_152_1 – **note that yours may be different, depending on when you stopped your search**. The results come back with a 2D diagram, with pharmacophore matches indicated. Select your smallest RMSD result by clicking on that row in the *Results Hitlist* window.

Take a few moments to explore how the returned result matches the pharmacophore query. In particular, note that:


- Feature matching is based on size of sphere (and thus the tolerance level). Tiny spheres with tight tolerance will result in hits with very close alignment to the centre of the sphere, while larger spheres have a wider area for alignment.

The pharmacophore search performed has provided a high number of hits, likely due to the large radius of the features. Lowering the tolerance of the features will get a smaller result set, with more precise alignment to the pharmacophore centres.

To do this, you will need to edit the pharmacophore query, which cannot be done in pause mode.

9. Click the **Stop**  button to clear the results and enable editing.

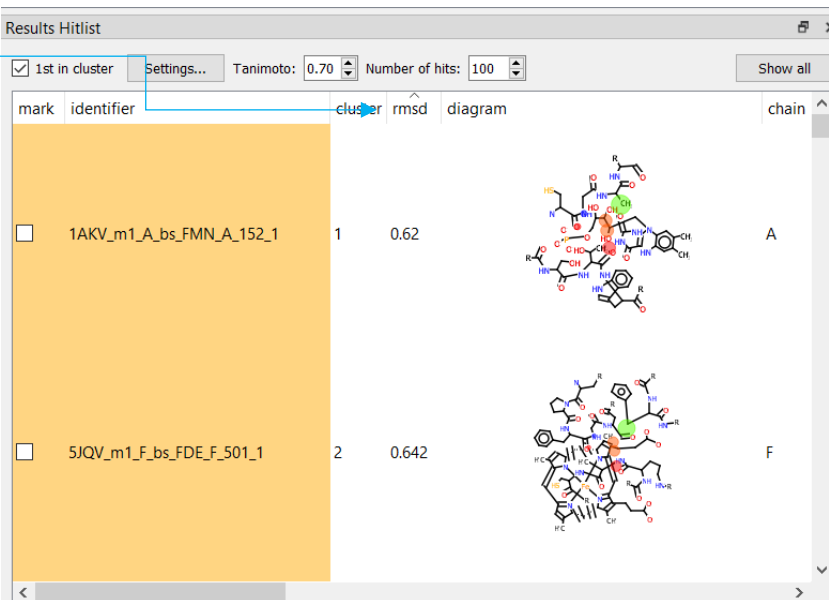
10. Double click on the radius sizes in the *Pharmacophore Features* window to change them. Change the radius of every feature to 0.50 to lower the tolerance. This will result in fewer hits, which are all more closely aligned with the centre of the pharmacophore features.

Then click on  to start the search on the new, tighter, pharmacophore.

It will take a bit more time to pick up hits, as there are far fewer. Let the search go to completion this time, resulting in 24 hits.

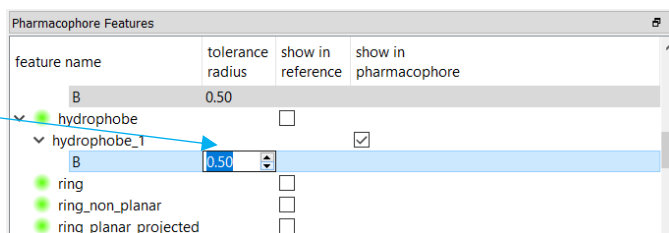
11. If **1st in cluster** is selected in the *Results Hitlist*, then the results are clustered based on the adjacent Tanimoto value. If clustered, you will see representatives of those similar groups in the results viewer (and in this case only seven hits are returned). Uncheck the **1st in cluster** tick-box to see all

8



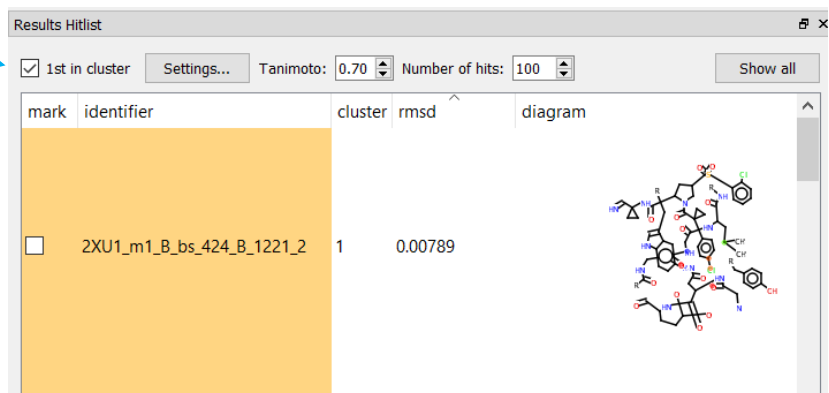
mark	identifier	cluster	rmsd	diagram	chain
<input type="checkbox"/>	1AKV_m1_A_bs_FMN_A_152_1	1	0.62		A
<input type="checkbox"/>	5JQV_m1_F_bs_FDE_F_501_1	2	0.642		F

10



feature name	tolerance radius	show in reference	show in pharmacophore
B	0.50	<input type="checkbox"/>	<input type="checkbox"/>
hydrophobe		<input type="checkbox"/>	<input type="checkbox"/>
hydrophobe_1		<input type="checkbox"/>	<input checked="" type="checkbox"/>
B	0.50	<input type="checkbox"/>	<input type="checkbox"/>
ring		<input type="checkbox"/>	<input type="checkbox"/>
ring_non_planar		<input type="checkbox"/>	<input type="checkbox"/>
ring_planar_projected		<input type="checkbox"/>	<input type="checkbox"/>

11



mark	identifier	cluster	rmsd	diagram
<input type="checkbox"/>	2XU1_m1_B_bs_424_B_1221_2	1	0.00789	

hits, including those which are very similar to each other.

12

12. Hide the 2D diagram of the matched hits by right-clicking in the *Results Hitlist* window and then click on **- diagram**.

13. Click and scroll the horizontal scrollbar in the *Results Hitlist* to have access to the annotations of the matched hits. Stop at the *molecule* annotation.

You will see that nearly all matched structures correspond to the **cathepsin L** PDB entries. This is not surprising because of the low radius of tolerance.

If you explore the top result (2XU1_m1_B_bs_424_B_1221_2), you'll see that the hit matches the pharmacophore exactly, it was indeed the original structure used to derive this pharmacophore.

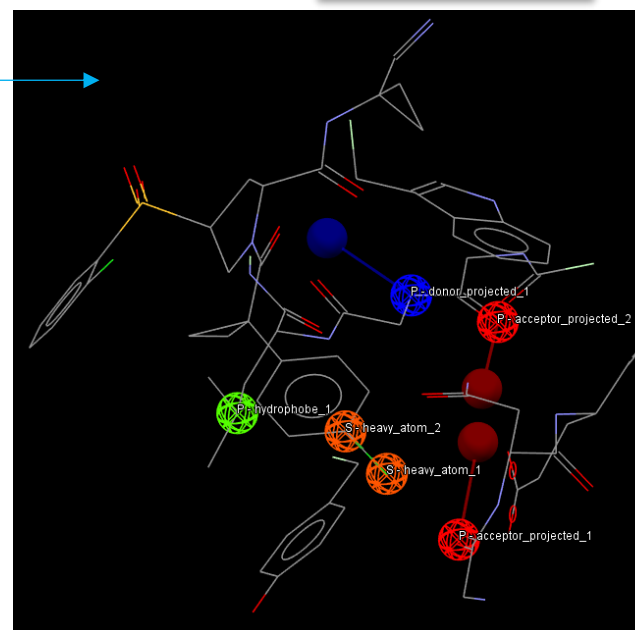
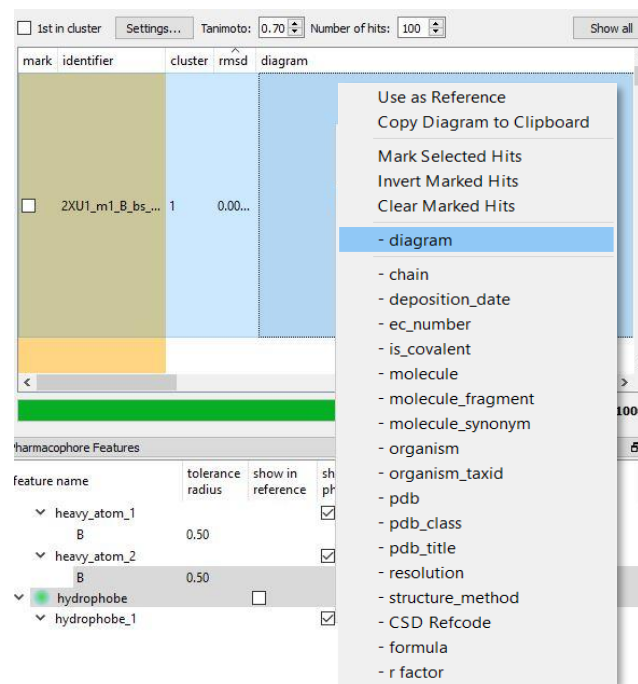
13

Results Hitlist

☐ 1st in cluster Settings... Tanimoto: 0.70 Number of hits: 100 Show all

identifier	cluster	rmsd	chain	deposition_date	ec_number	is_covalent	molecule	molecule_f
2XU1_m1_B_bs_424_B_1221_2	1	0.00...	B	2010-10-14	3.4.22.15	Yes	cathepsin I1	catalytic do
2XU1_m2_B_bs_424_B_1221_2	1	0.00...	B	2010-10-14	3.4.22.15	Yes	cathepsin I1	catalytic do
2XU1_m1_D_bs_424_D_1221	1	0.101	D	2010-10-14	3.4.22.15	Yes	cathepsin I1	catalytic do
2YJ8_m1_A_bs_YJ8_A_1221_2	1	0.163	A	2011-05-19	3.4.22.15	Yes	cathepsin I1	catalytic do
2YJ8_m2_A_bs_YJ8_A_1221_2	1	0.163	A	2011-05-19	3.4.22.15	Yes	cathepsin I1	catalytic do
2YJ2_m1_A_bs_YJ2_A_1221	1	0.172	A	2011-05-18	3.4.22.15	Yes	cathepsin I1	catalytic do
2XU1_m1_A_bs_424_A_1221	1	0.185	A	2010-10-14	3.4.22.15	Yes	cathepsin I1	catalytic do
2XU3_m1_A_bs_XU3_A_1221_2...	1	0.197	A	2010-10-14	3.4.22.15	Yes	cathepsin I1	catalytic do
2YJC_m1_A_bs_424_A_1221_2	1	0.201	A	2011-05-19	3.4.22.15	Yes	cathepsin I1	catalytic do
2YJC_m2_A_bs_424_A_1221_2	1	0.201	A	2011-05-19	3.4.22.15	Yes	cathepsin I1	catalytic do
2XU1_m1_C_bs_424_C_1221	1	0.217	C	2010-10-14	3.4.22.15	Yes	cathepsin I1	catalytic do
2YJB_m1_A_bs_YJ9_A_1221	1	0.253	A	2011-05-19	3.4.22.15	Yes	cathepsin I1	catalytic do
5F02_m1_A_bs_5T9_A_302	1	0.253	A	2015-11-27	3.4.22.15	Yes	cathepsin I1	catalytic do
3OF9_m1_A_bs_I0X_A_501	2	0.31	A	2010-08-14	3.4.22.15	Yes	cathepsin I1	cathepsin I
2YJ9_m1_A_bs_YJ9_A_1221	1	0.323	A	2011-05-19	3.4.22.15	Yes	cathepsin I1	catalytic do
3H8C_m1_A-B_bs_NSZ_B_400	3	0.328	B	2009-04-29	3.4.22.15	No	cathepsin I1	cathepsin I
2XU3_m1_A_bs_XU3_A_1221_2...	1	0.333	A	2010-10-14	3.4.22.15	Yes	cathepsin I1	catalytic do
1MHV_m1_B-D_bs_BP4_H_41_1	4	0.347	H	2002-08-21		No	4-biphenylacetyl-cys-(...	residues 11:
3OF8_m1_A_bs_I0Y_A_501	2	0.367	A	2010-08-14	3.4.22.15	Yes	cathepsin I1	cathepsin I
3H8B_m1_B-C-F_bs_NSY_F_800	5	0.387	F	2009-04-29	3.4.22.15	No	cathepsin I1	cathepsin I
3H8B_m1_B_bs_NSY_B_400	5	0.401	B	2009-04-29	3.4.22.15	No	cathepsin I1	cathepsin I
2HX2_m1_A-B_bs_H7J_A_811	6	0.42	A	2006-08-04	3.4.22.27	Yes	cathepsin s	
1MHV_m1_B-D_bs_BP4_H_41_2	4	0.42	H	2002-08-21		No	4-biphenylacetyl-cys-(...	
2F1G_m1_A-B_bs_GNF_A_1002	7	0.423	A	2005-11-14	3.4.22.27	No	cathepsin s	cathepsin s

#hits: 24/10000



14. To save interesting hits for further work outside CSD-CrossMiner, you can mark them by ticking the respective tick-boxes in the *mark* column.

Note that marking them does not display them in the 3D view, and similarly, displaying them in the 3D view (multiple selection available via SHIFT or CTRL) does not mark them.

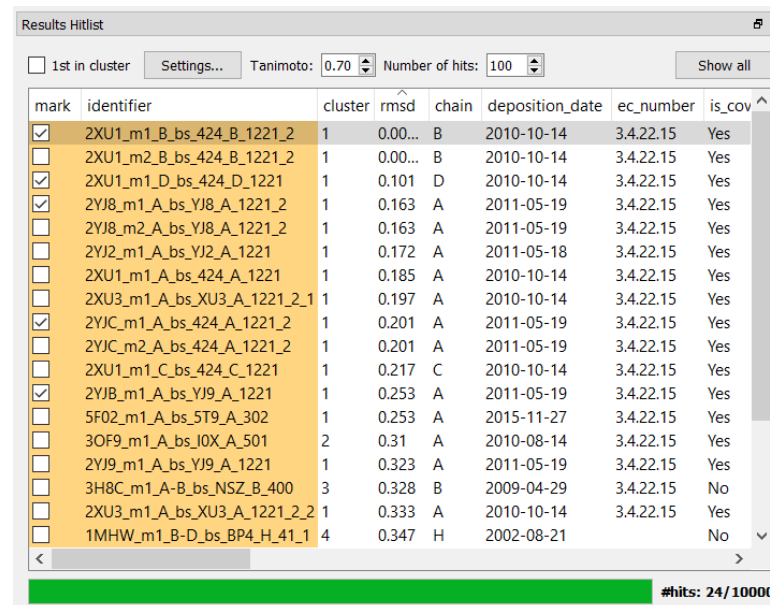
15. The 3D coordinates of the marked hits can be then saved to disk by clicking on **File** in the CSD-CrossMiner top-level menu and then **Save Marked Hits**. Save your hits as *cathepsin_hits.mol2*.

Alternatively, in the **File** menu other two different save options are available: **Save Visible Hits**, which will save the visible hits (in the 3D view), **Save All Hits** that will save the 3D coordinates of all matched hits.

Note that the hits in the *Results Hitlist* browser (visible, marked and all hits) can be additionally saved as a table by choosing *csv* format in the *Save Hits* window.

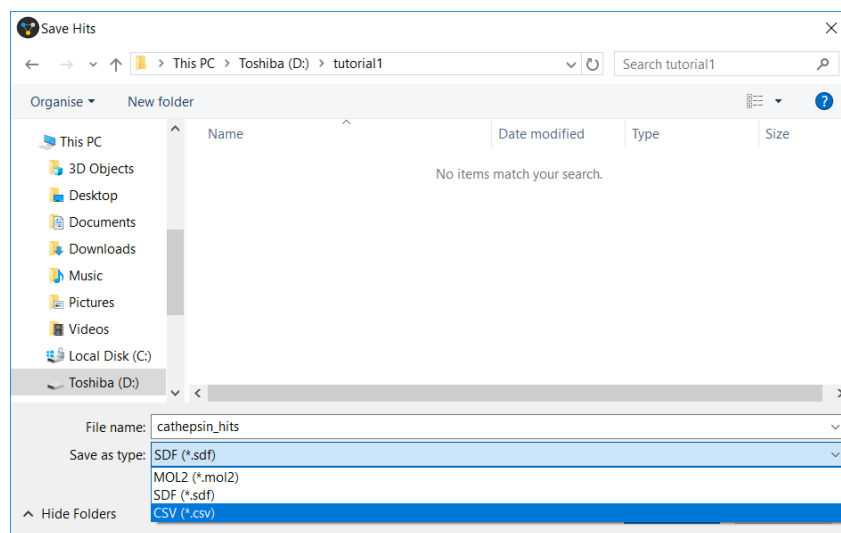
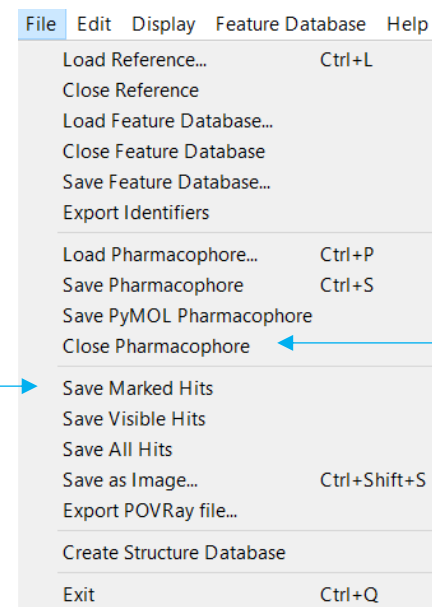
16. Now click on **File > Close Pharmacophore** to clear the 3D view without saving the edited pharmacophore.

This ends the Exercise 1.



mark	identifier	cluster	rmsd	chain	deposition_date	ec_number	is_cov
<input checked="" type="checkbox"/>	2XU1_m1_B_bs_424_B_1221_2	1	0.00...	B	2010-10-14	3.4.22.15	Yes
<input type="checkbox"/>	2XU1_m2_B_bs_424_B_1221_2	1	0.00...	B	2010-10-14	3.4.22.15	Yes
<input type="checkbox"/>	2XU1_m1_D_bs_424_D_1221	1	0.101	D	2010-10-14	3.4.22.15	Yes
<input checked="" type="checkbox"/>	2YJ8_m1_A_bs_YJ8_A_1221_2	1	0.163	A	2011-05-19	3.4.22.15	Yes
<input type="checkbox"/>	2YJ8_m2_A_bs_YJ8_A_1221_2	1	0.163	A	2011-05-19	3.4.22.15	Yes
<input type="checkbox"/>	2YJ2_m1_A_bs_YJ2_A_1221	1	0.172	A	2011-05-18	3.4.22.15	Yes
<input type="checkbox"/>	2XU1_m1_A_bs_424_A_1221	1	0.185	A	2010-10-14	3.4.22.15	Yes
<input type="checkbox"/>	2XU3_m1_A_bs_XU3_A_1221_2_1	1	0.197	A	2010-10-14	3.4.22.15	Yes
<input checked="" type="checkbox"/>	2YJC_m1_A_bs_424_A_1221_2	1	0.201	A	2011-05-19	3.4.22.15	Yes
<input type="checkbox"/>	2YJC_m2_A_bs_424_A_1221_2	1	0.201	A	2011-05-19	3.4.22.15	Yes
<input type="checkbox"/>	2XU1_m1_C_bs_424_C_1221	1	0.217	C	2010-10-14	3.4.22.15	Yes
<input checked="" type="checkbox"/>	2YJB_m1_A_bs_YJ9_A_1221	1	0.253	A	2011-05-19	3.4.22.15	Yes
<input type="checkbox"/>	5F02_m1_A_bs_5T9_A_302	1	0.253	A	2015-11-27	3.4.22.15	Yes
<input type="checkbox"/>	3OF9_m1_A_bs_I0X_A_501	2	0.31	A	2010-08-14	3.4.22.15	Yes
<input type="checkbox"/>	2YJ9_m1_A_bs_YJ9_A_1221	1	0.323	A	2011-05-19	3.4.22.15	Yes
<input type="checkbox"/>	3H8C_m1_A-B_bs_NSZ_B_400	3	0.328	B	2009-04-29	3.4.22.15	No
<input type="checkbox"/>	2XU3_m1_A_bs_XU3_A_1221_2_2	1	0.333	A	2010-10-14	3.4.22.15	Yes
<input type="checkbox"/>	1MHW_m1_B-D_bs_BP4_H_41_1	4	0.347	H	2002-08-21		No

15



Exercise 2: 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). In this example we focus on editing a pharmacophore query.

1. If you already have work in CSD-CrossMiner, be sure that you have cleaned the 3D view. You can use **File > Close Pharmacophore** and/or **File > Close Reference** to do so.
2. Load the cathepsin L pharmacophore by clicking on **File > Load Pharmacophore...** and select:
[CSD-CrossMiner installation]\example_pharmacophores\catl_s3.cm
3. 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:

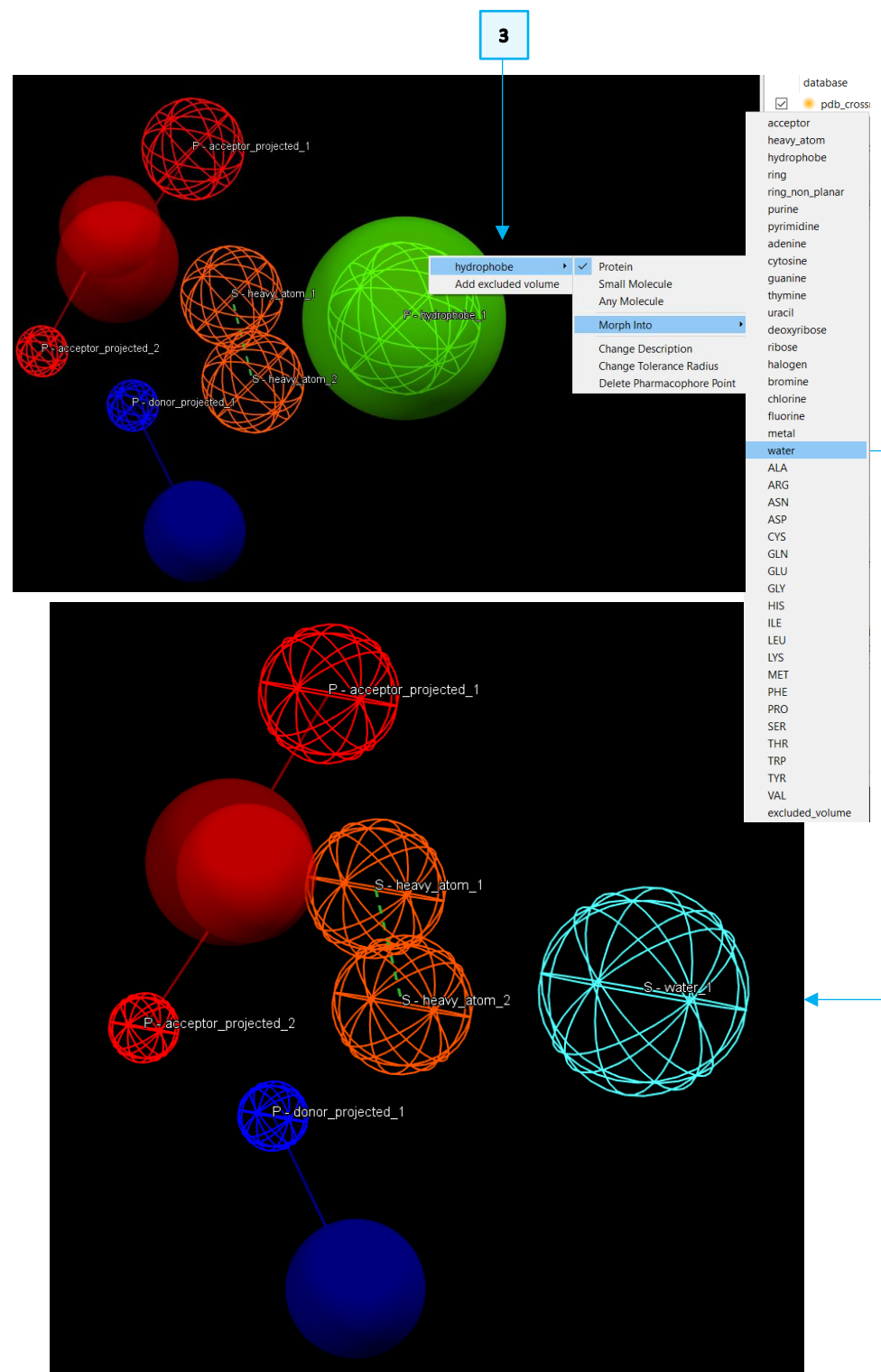
- Define where the feature belongs: **Protein**, **Small Molecule**, or **Any**.
- Change the feature type (**Morph Into**).
- Change the label of the feature (**Change Description**).
- Change the tolerance radius of the pharmacophore point (**Change Tolerance Radius**).
- Delete the pharmacophore point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



- 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.

- Be sure that the **1st in cluster** tick-box is checked, this will show only the cluster representatives 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.

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:

- 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.

Change the radius of the water pharmacophore points to 1.50 by using one of the methods described hereafter:

- 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

4

feature name	tolerance radius	show in reference	show in pharmacophore
All		<input type="checkbox"/>	<input checked="" type="checkbox"/>
acceptor		<input checked="" type="checkbox"/>	
acceptor_projected		<input checked="" type="checkbox"/>	
acceptor_projected_1			<input checked="" type="checkbox"/>
B	1.00		
V	1.00		
acceptor_projected_2			<input checked="" type="checkbox"/>
B	0.50		
V	1.00		
donor_ch_projected		<input checked="" type="checkbox"/>	

5

The screenshot shows the CSD-CrossMiner software interface. The main 3D view displays a molecular model with several pharmacophore points represented as spheres. The points are labeled: acceptor_projected_1, acceptor_projected_2, acceptor_projected_3, water, and donor_ch_projected. The water point is highlighted in red. The Results Hitlist window on the right shows a table of search results. The table has columns: mark, identifier, cluster, mtd, chain, deposition_date, cc_number, and n_covalent. The table is currently empty. The Pharmacophore Features window is also visible, showing the same table as in the previous screenshot.

the radius of the water sphere and automatically restart the search.

- If you do not have a three-button mouse or have trouble getting this to work, you can use the *Pharmacophore Features* window (See **Step 3**) to change the radius of the water pharmacophore sphere. You can also change the radius of a single pharmacophore point by using the **Change Tolerance radius** through the pharmacophore context menu (see **Step 2**).

You will start to see the *Results Hitlist* window repopulating with new matching hits. Note that 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).

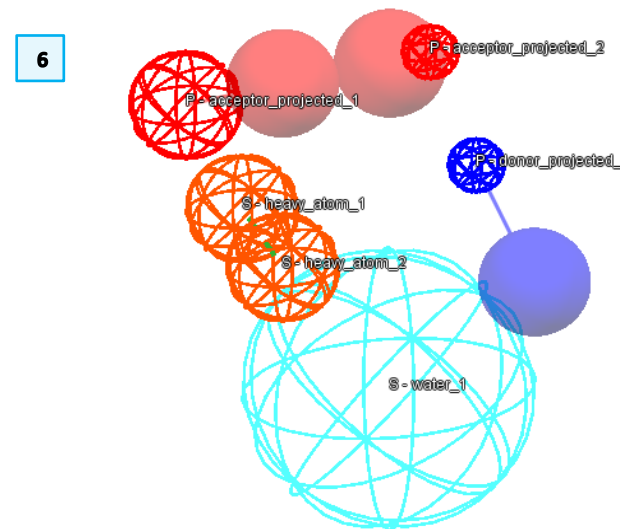
Tip: Every CSD-CrossMiner user has accidentally grabbed a pharmacophore point and dragged it across the screen when they intended to rotate the molecule viewer. In such cases **Ctrl + z** (or **Edit > Undo**) is very helpful. It will undo the last change made to the pharmacophore. However, the search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

not using it (open hand  button means editing is off).

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.

- Display the 2D diagram of the matched hits by right-clicking in the *Results Hitlist* window and then click on **+ diagram**. This will help you to see the nature of the hits matching the pharmacophore query.
- Right-click on one of the hit 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.



mark	identifier	cluster	rmsd	chain	deposition_date	ec_number	is_cova
<input type="checkbox"/>	2P3T_m1_B_bs_993_B_50...	1	0.000	B	2007-03-09	3.4,21.6	No
<input type="checkbox"/>	4MNY_m1_A_bs_GLY_C_1	1	0.000	C	2013-09-11		No
<input type="checkbox"/>	4C4D_m2_A_bs_BGC_A_4	1	0.000	A	2013-09-05	3.2,1.91	No
<input type="checkbox"/>	3TR0_m1_A_bs_5GP_A_2	1	0.000	A	2011-09-09	2.7,4.8	No
<input type="checkbox"/>	6JBS_m1_A_bs_NAG_A_9	1	0.000	A	2019-01-26		Yes
<input type="checkbox"/>	2I41_m1_B_bs_5GP_B_12	1	0.000	B	2006-08-24	2.7,4.8	No
<input type="checkbox"/>	3WA2_m1_X_bs_PEG_X_7	1	0.000	X	2013-04-22	1.4,3.21	No
<input type="checkbox"/>	4C4D_m2_A_bs_G2F_A_4	1	0.000	A	2013-09-05	3.2,1.91	No
<input type="checkbox"/>	1GKY_m1_A_bs_5GP_A_1	1	0.000	A	1991-12-23	2.7,4.8	No
<input type="checkbox"/>	4JFA_m1_D_bs_TRP_D_70	1	0.000	D	2013-02-28	6.1,1.2	No
<input type="checkbox"/>	1X8X_m1_A_bs_TYR_A_9	1	0.000	A	2004-08-19	6.1,1.1	No
<input type="checkbox"/>	2XC4_m1_A_bs_IVK_A_12	1	0.000	A	2010-04-16	3.4,21.6	No
<input type="checkbox"/>	4MMP_m1_A_bs_SLB_A_4	1	0.000	A	2013-09-09		No
<input type="checkbox"/>	1Z8Q_m1_A_bs_HEM_A_4	1	0.000	A	2005-03-31		No
<input type="checkbox"/>	1EX7_m1_A_bs_5GP_A_1	1	0.000	A	2000-05-01	2.7,4.8	No
<input type="checkbox"/>	1SRH_m1_A_bs_MOB_B_1	1	0.000	B	1994-02-17		Yes
<input type="checkbox"/>	4AWU_m1_A_bs_FMN_A_1	1	0.000	A	2012-06-05	1.6,99.1	No
<input type="checkbox"/>	4O2A_m1_A-B_bs_2RR_B	1	0.000	B	2013-12-17		No
<input type="checkbox"/>	4KH7_m1_A-B_bs_GSH_A	1	0.000	A	2013-04-30		No

#hits: 410/10000

9. You can display all the features associated with the reference structure by clicking on the **feature** ☐ tick-box in the CSD-CrossMiner *Show:* toolbar at the top (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.
10. If you wish to hide all features, click on **features** ☒ (which will turn to ☐) or alternatively by unticking the *All* tick-box in the *Pharmacophore Features* window.
11. Change the style of the small molecule in the 3D view by left-clicking on one of the small molecule atoms while pressing *Shift*. 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 and select **Capped Sticks**.

9

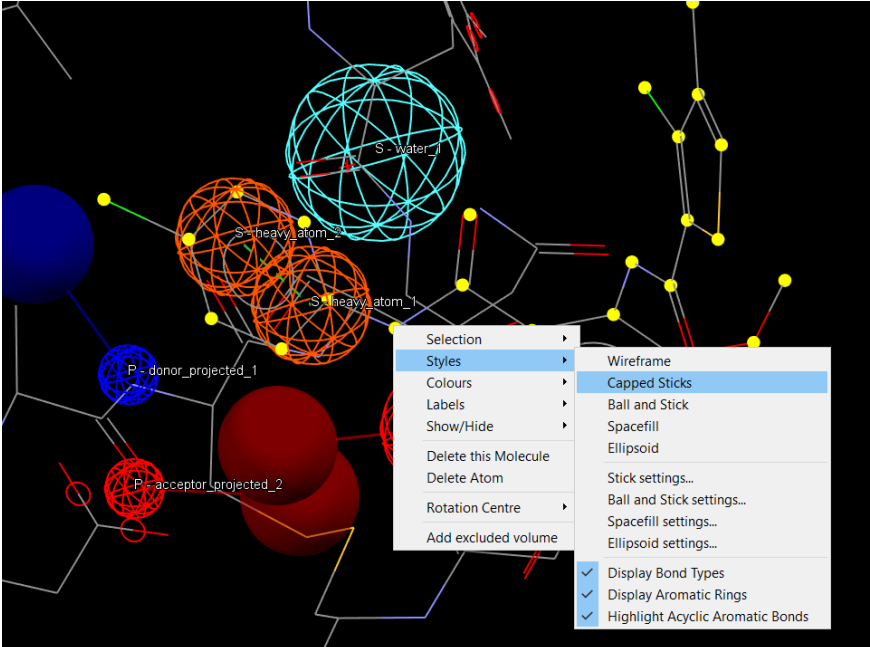
Show: ☒ reference ☐ hits ☒ constraints ☐ features


Pharmacophore Features			
feature name	tolerance radius	show in reference	show in pharmacophore
All		<input type="checkbox"/>	<input checked="" type="checkbox"/>

10

Show: ☒ reference ☐ hits ☒ constraints ☒ features

11



12. 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. Take a look at the new results.

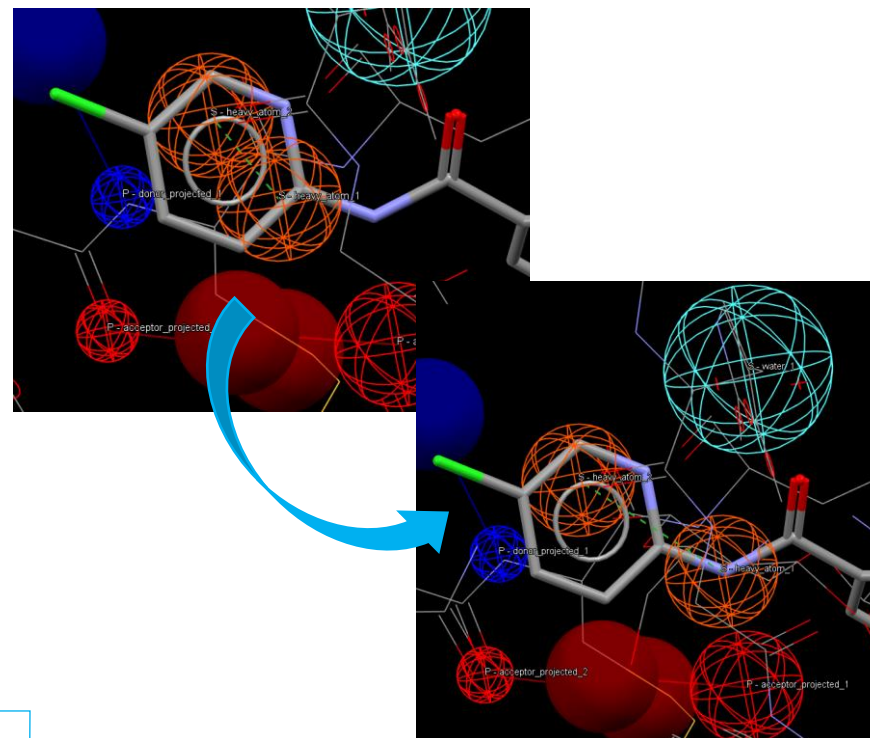
13. You can hide the reference molecule by disabling the **reference** tick-box in the *Show:* toolbar.
14. You may wish to save your edited pharmacophore by clicking **File > Save Pharmacophore**. Name the file *pharmacophore_edited.cm*.

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 exercise 2.

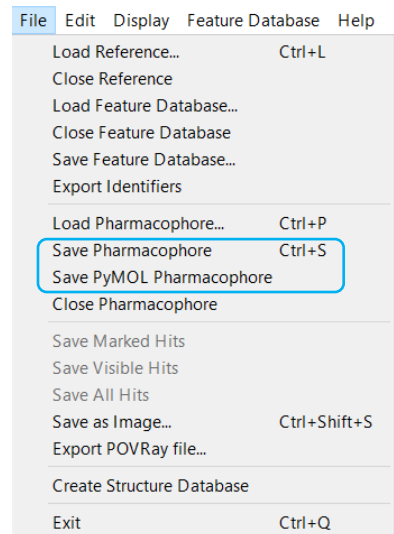
12





13

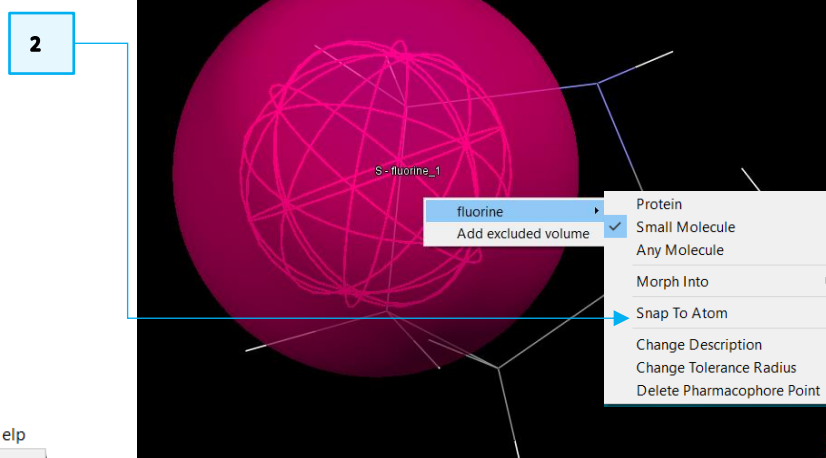
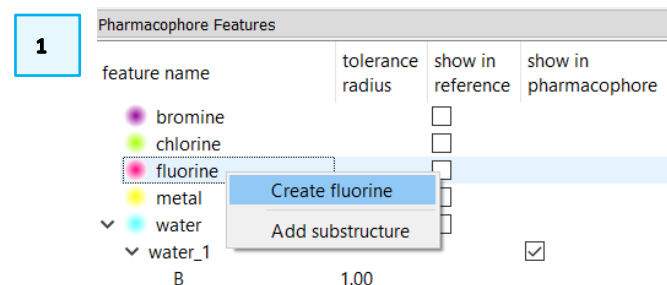
Show: ☒ reference ☐ hits ☒ constraints ☐ features ☒ pharmacophore ☐ pharm. labels ☐ hydrogens

14

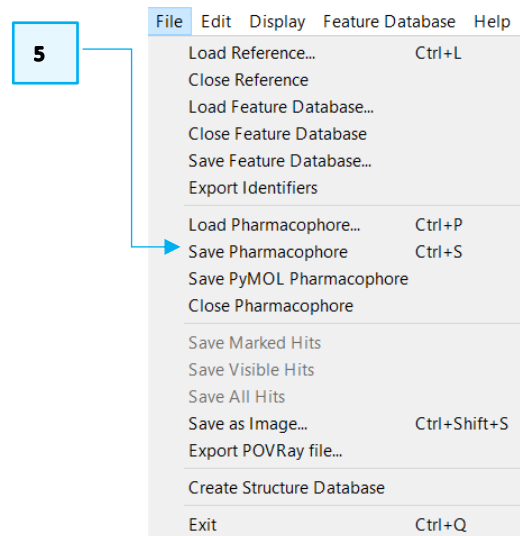


Bonus exercise: Adding new pharmacophore features

1. 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.
2. Align the new fluorine near to the closest atom by right-clicking on the fluorine feature and select **Snap To Atom**.
3. Start the search by clicking on . This new search probably won't find any results. This step was just to demonstrate how to add and manoeuvre new features.
4. Try to change the nature of the **fluorine** pharmacophore points by using **Morph Into** from the right click pharmacophore menu or try to move the pharmacophore point using the **Pharmacophore editing** mode on (). Do you have nay hits?
5. When you are happy with your pharmacophore query you can save it by clicking **File** > **Save Pharmacophore**. Name the file *pharmacophore_edited.cm*.



This ends the bonus exercise.



Conclusions

In this workshop we used CSD-CrossMiner to run and edit a pharmacophore search. You should now be able to:

- Navigate the CSD-CrossMiner interface.
- Load pharmacophores from a list of pre-defined ones.
- Understand the pharmacophore features displayed.
- Run, pause, stop a pharmacophore search.
- Edit the pharmacophore features.
- Remove, add, order the annotation fields in the *Results Hitlist*.
- Use one of the results as a new reference.

Next steps

If you finished the exercises early, you could head to the bonus exercise. Otherwise, ask the workshop tutors for new challenges!

Feedback

We hope this workshop improved your understanding of CSD-CrossMiner and you found it useful for your work. As we aim at continuously improving our training materials, we would love to get your feedback. We will be sharing a link for a feedback survey with you at the end of the session to fill in. It will take less than **5 minutes** to complete. The feedback is anonymous. Thank you!

