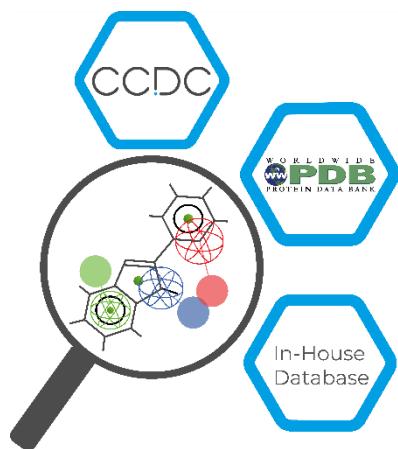


Creating a Pharmacophore Query from a Reference Molecule & Scaffold Hopping in CSD-CrossMiner

Version 2.0 – September 2019
CSD-CrossMiner v1.5.3

Table of Contents

Introduction.....	2
CSD-CrossMiner Terminology.....	2
Overview of CSD-CrossMiner.....	3
Features and Pharmacophore Representation.....	4
Creating a Pharmacophore from a Reference Molecule.....	5
Scaffold Hopping using CSD-CrossMiner.....	8



CCDC
advancing structural science

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 (about 381 018 structures total). The included PDB database is divided in two subsets, one composed by 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 the protein-ligand-nucleic acids binding site is provided, where the binding site is defined as all molecules with an atom within a 6Å radius around the ligand (> 285 000 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 create a pharmacophore query from a reference molecule and apply the tool to obtain scaffold hopping ideas.

The files to perform this tutorial are provided in the tutorial3 folder [here](#).

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 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: . The `csd_pdb_crossminer.feet` feature database, provided in the CSD package (`$CSD_2019/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.
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.

2

Feature Databases		
checkbox	database	size
<input checked="" type="checkbox"/>	pdb_crossminer	285946
<input checked="" type="checkbox"/>	nucleic_acid_crossminer	5427
<input checked="" type="checkbox"/>	csd540_crossminer	381018

3

Pharmacophore Features			
feature name	tolerance radius	show in reference	show in pharmacophore
<input checked="" type="checkbox"/>	acceptor	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	acceptor_projected	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	donor_ch_projected	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/>	donor_projected	<input checked="" type="checkbox"/>	
<input type="checkbox"/>	heavy_atom	<input type="checkbox"/>	
<input type="checkbox"/>	hydrophobe	<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"/>	purine	<input type="checkbox"/>	
<input type="checkbox"/>	pyrimidine	<input type="checkbox"/>	
<input type="checkbox"/>	adenine	<input type="checkbox"/>	
<input type="checkbox"/>	cytosine	<input type="checkbox"/>	
<input type="checkbox"/>	guanine	<input type="checkbox"/>	
<input type="checkbox"/>	thymine	<input type="checkbox"/>	
<input type="checkbox"/>	uracil	<input type="checkbox"/>	
<input type="checkbox"/>	deoxyribose	<input type="checkbox"/>	
<input type="checkbox"/>	ribose	<input type="checkbox"/>	
<input type="checkbox"/>	exit_vector	<input type="checkbox"/>	
<input type="checkbox"/>	halogen	<input type="checkbox"/>	
<input type="checkbox"/>	bromine	<input type="checkbox"/>	
<input type="checkbox"/>	chlorine	<input type="checkbox"/>	
<input type="checkbox"/>	fluorine	<input type="checkbox"/>	
<input type="checkbox"/>	metal	<input type="checkbox"/>	
<input type="checkbox"/>	water	<input type="checkbox"/>	
<input type="checkbox"/>	ALA	<input type="checkbox"/>	
<input type="checkbox"/>	ARG	<input type="checkbox"/>	
<input type="checkbox"/>	ASN	<input type="checkbox"/>	
<input type="checkbox"/>	ASP	<input type="checkbox"/>	
<input type="checkbox"/>	CYS	<input type="checkbox"/>	
<input type="checkbox"/>	GLN	<input type="checkbox"/>	
<input type="checkbox"/>	GLU	<input type="checkbox"/>	
<input type="checkbox"/>	GLY	<input type="checkbox"/>	

Features and Pharmacophore Representation

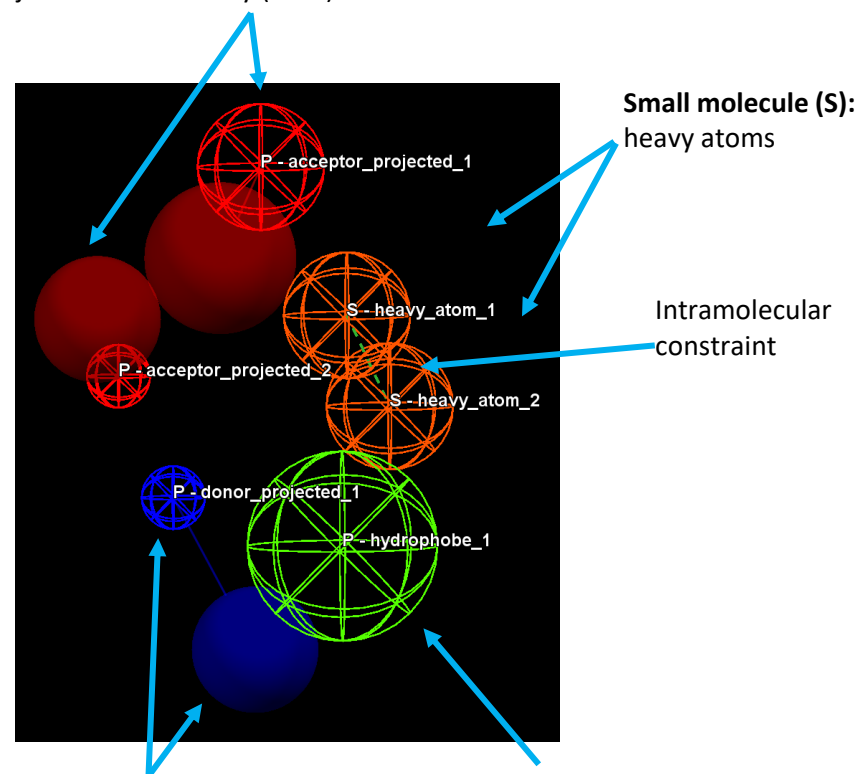
In the CSD-CrossMiner 3D view, features are represented as small translucent spheres coloured as defined in the *Pharmacophore Features* window. A pharmacophore point is represented as a mesh sphere which reflects the uncertainty in the position of the pharmacophore point. In the 3D view:

- **P:** Protein pharmacophore point
- **S:** Small molecule pharmacophore point
- **A:** Either a small molecule or protein pharmacophore point
- **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 of the pharmacophore point.
- **Solid sphere:** the projected virtual point to represent the directionality of e.g. 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 pharmacophore, the mesh sphere (the actual feature itself) is defined as *B* in the *Pharmacophore Features* window (**B**ase feature), and the projected virtual representing the directionality of the feature is defined as *V* (**V**irtual point).

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



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		<input type="checkbox"/>	<input checked="" type="checkbox"/>
• acceptor		<input type="checkbox"/>	<input type="checkbox"/>
• acceptor_projected		<input type="checkbox"/>	<input 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		

Creating a Pharmacophore from a Reference Molecule

A common use of CSD-CrossMiner is to create a pharmacophore from a reference molecule. This could be building a pharmacophore from the ligand of an experimentally determined protein-ligand complex, or manually creating a CSD-CrossMiner pharmacophore from a set of atoms representing a pharmacophore created by another mechanism.

If you already have work in CSD-CrossMiner, clear the 3D view by clicking on **File > Close Pharmacophore** and/or **File > Close Reference** from the CSD-CrossMiner top-level menu.

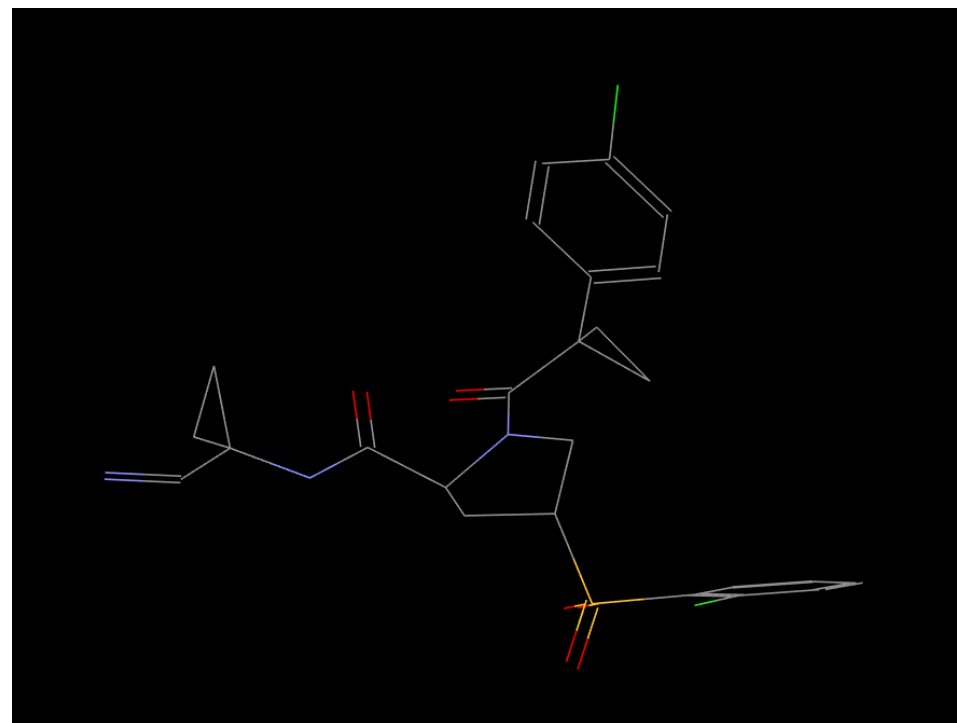
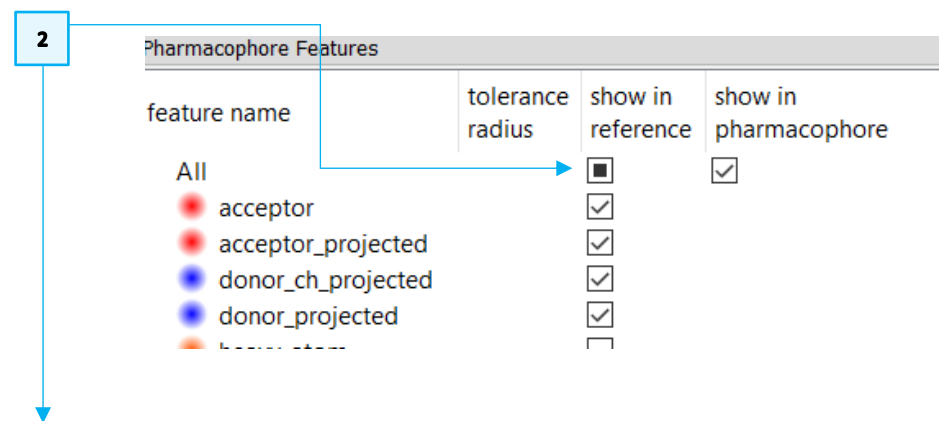
1. Click on **File > Load Reference** and select *2xu1_ligand.sdf* provided in the `tutorial13` folder which you have downloaded.

By default, only donor and acceptor features associated with the reference structure are displayed in the 3D view. The features are represented in the 3D view as small translucent spheres, whose identity and associated colour is shown in the *Pharmacophore Features* window.

Note that if a different choice of displayed features is made in the CSD-CrossMiner session (e.g. all features displayed), these new settings will be remembered when a new reference molecule is loaded (e.g. in that case all features of the new reference molecule would be displayed).



To make it easier to choose which atom to add a pharmacophore to, toggle the check-boxes in the *show in reference* column of the *Pharmacophore Features* window.

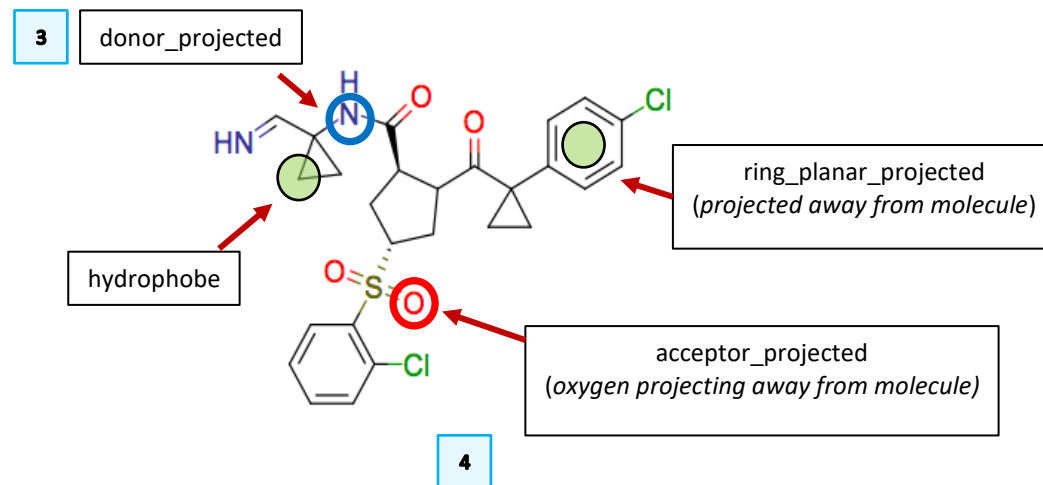
2. Click two times on the *All* tick-box in the *Pharmacophore Features*, so that no features are shown. Then, you can select individual tick-boxes to display only the required feature types, as needed in the following step.



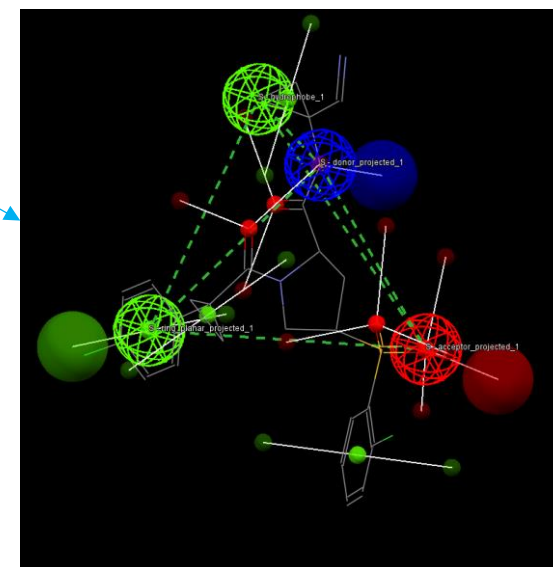
3. Right-click on features of the reference molecule in the 3D view to define pharmacophore points as shown in the image to the right. For the directional features, different locations of the projection can be selected. Select the **Add ring_planar_projected** and **Add acceptor_projected** to have the virtual points far away from the molecule.



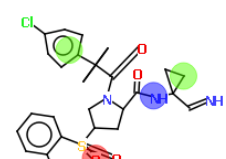
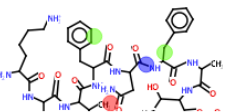
4. When all pharmacophore points are defined, click on  to set all defined pharmacophore points to be part of the same molecule.
5. From the *Feature Databases* window, deselect **csd540_crossminer** tick-box. This will result in using only the feature database derived from PDB structures.
6. Click on  to start the search and let it run to completion.
7. Sort the hits in the *Results Hitlist* by RMSD by clicking on the *rmsd* column in the *Results Hitlist* window. The lowest RMSD hit in this example is 2XU1_m1_A_bs_424_A_1221_1, which corresponds to the co-crystallised ligand structure.



Note that the pharmacophore points have the virtual features (solid sphere) projected away from the molecule



7

mark	identifier	cluster	rmsd	diagram	chain
<input type="checkbox"/>	2XU1_m1_A_bs_424_A_1221_1	1	8.26...		A
<input type="checkbox"/>	1S7X_m1_J_bs_LYS_L_1_1	2	0.701		L

#hits: 166/10000

5

Feature Databases		size
database		
<input checked="" type="checkbox"/>	pdb_crossminer	285946
<input checked="" type="checkbox"/>	nucleic_acid_crossminer	5427
<input type="checkbox"/>	csd540_crossminer	381018

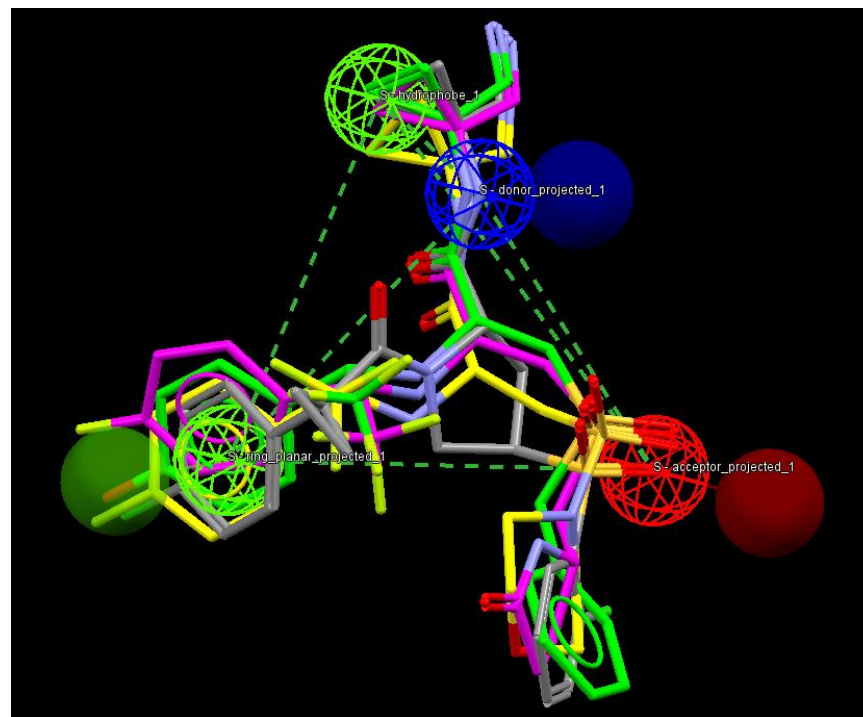
8. Scrolling through the hits with higher RMSD reveals interesting analogues (see 2FRA_m1_A-B_bs_CRV_A_999_1 and 2FQ9_m1_A-B_bs_CRJ_A_999).

You can change the style and the colour of the molecules in the 3D view by clicking on one of the small molecule atoms while pressing *Shift*. This will select all the atoms of the same molecule (the selection is represented as a small yellow sphere). Right-click on one of the selected atoms and pick **Styles**, from the pull-down menu and select the desired style, then right-click again and select **Colour** to select the desired colour

Note that you can hide the hydrogen atoms by disabling the **hydrogens** check-box in the *Show:* toolbar.

This demonstrates how easy it is to get ideas for scaffold hopping using CSD-CrossMiner.


8




Show: reference hits constraints features pharmacophore pharm. labels hydrogens

Scaffold Hopping using CSD-CrossMiner


In this example you will experiment with scaffold hopping of PD180970, the co-crystallised ligand of the ABL kinase domain using CSD-CrossMiner.

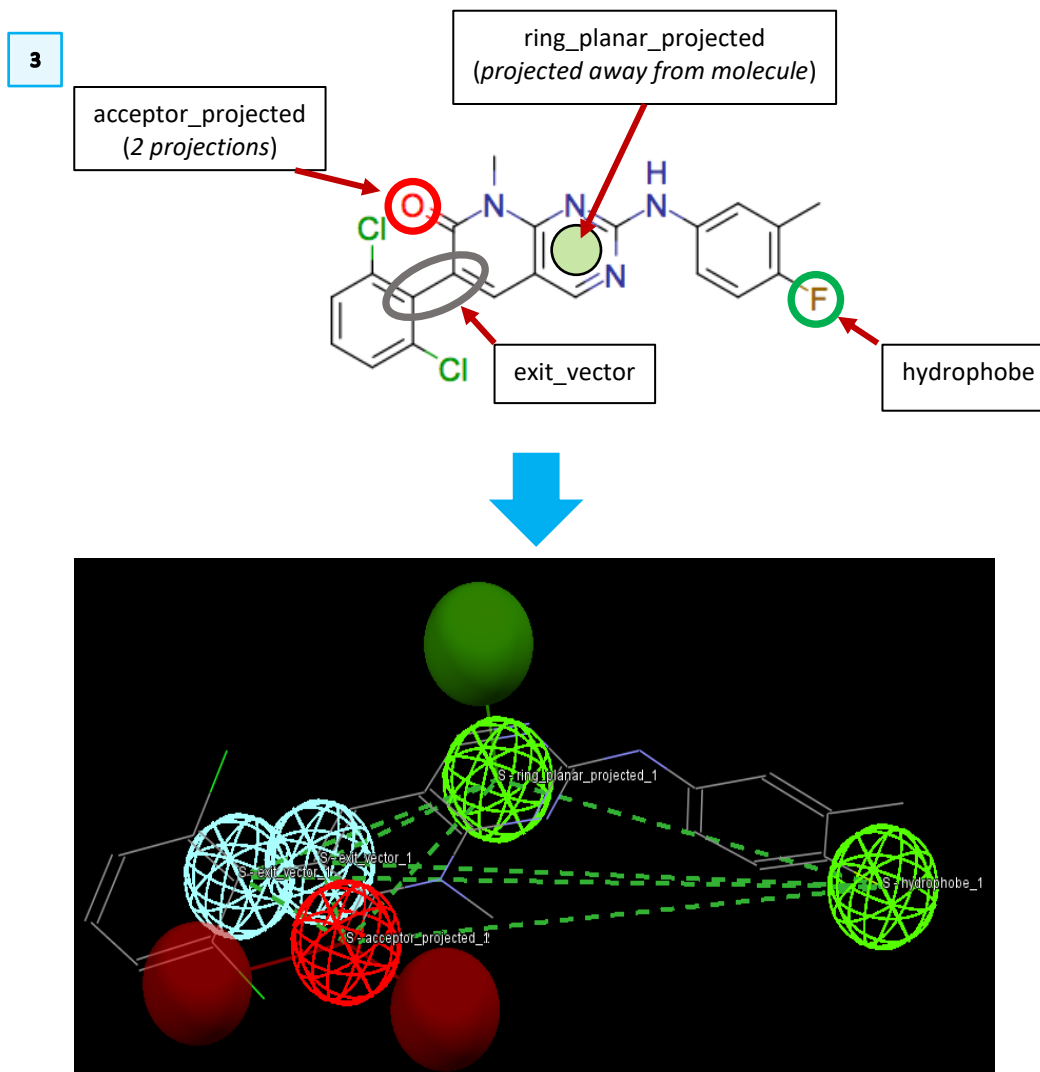
1. Clear the 3D view by clicking on **File > Close Pharmacophore** and/or **File > Close Reference** from the CSD-CrossMiner top-level menu.
2. Load the file *2hzi_ligand.sdf* as you did previously from the *tutorial3* folder that you downloaded.
3. Define the pharmacophore as indicated. For the *acceptor_projected*, two points are indicated – ensure that both projected virtual points are set to features. Ensure that the features are all set to *intra* constraints by clicking on the  button.

4. For this search, we will use both the CSD and PDB databases; therefore, make sure that the **pdb_crossminer**, **nucleic_acid_crossminer** and **csd540_crossminer** tick-boxes are both checked in the *Feature Databases* window.

5. Click on the  button to start the pharmacophore search. The search will take several minutes to complete. 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 selected when the search is running or when the search is paused.

6. You will see that there is no shortage of great results from the CSD database to match this query. When you have few hundreds of hits, pause the search by clicking on .



7. Try experimenting with narrowing this result set down:

- Restrict some of the pharmacophore radii
- Add new features that your chemistry experience says might be important.

The screenshot displays the CSO-CrossMiner software interface. The central window shows a 3D molecular model with various pharmacophore features highlighted in different colors (red, green, blue, yellow) and represented as spheres or vectors. The interface includes several panels:

- Feature Databases:** A list of databases with their sizes:

database	size
pdcb_crossminer	285946
nucleic_acid_crossminer	5427
csd540_crossminer	381018
- Results Hitlist:** A table showing search results with columns for mark, identifier, cluster, mscd, chain, deposition_date, ec_number, and is_covalent.

mark	identifier	cluster	mscd	chain	deposition_date	ec_number	is_covalent
<input type="checkbox"/>	NAALMYL_1	145	0.17				
<input type="checkbox"/>	NAXION_1	158	0.17				
<input type="checkbox"/>	STVT_m1_A_bc_7LV_A_4D1_1	1	0.204	A	2016-11-10	2.7.11.1	No
<input type="checkbox"/>	NADDAZ_1	136	0.228				
<input type="checkbox"/>	MORPHAS_1	63	0.238				
<input type="checkbox"/>	NEDIPK_1	161	0.265				
<input type="checkbox"/>	LORVIC_1	29	0.376				
<input type="checkbox"/>	MEYNUJ_1	90	0.38				
<input type="checkbox"/>	MCMJOP_1	117	0.404				
<input type="checkbox"/>	MAWTOU_1	65	0.499				
<input type="checkbox"/>	MAPDOZ_1	58	0.501				
<input type="checkbox"/>	MKNOW_1	106	0.597				
<input type="checkbox"/>	LURLEN_1	44	0.622				
<input type="checkbox"/>	MUXIAR_1	132	0.627				
<input type="checkbox"/>	NEBIE_1	160	0.63				
- Pharmacophore Features:** A list of features with their tolerance radius and checkboxes for showing them in the reference and pharmacophore.

feature name	tolerance radius	show in reference	show in pharmacophore
ribose		<input type="checkbox"/>	<input type="checkbox"/>
axil_vector_1		<input type="checkbox"/>	<input checked="" type="checkbox"/>
axil_vector_1	1.00	<input type="checkbox"/>	<input checked="" type="checkbox"/>
B	1.00	<input type="checkbox"/>	<input type="checkbox"/>
halogen		<input type="checkbox"/>	<input type="checkbox"/>
bromine		<input type="checkbox"/>	<input type="checkbox"/>
chlorine		<input type="checkbox"/>	<input type="checkbox"/>
fluorine		<input type="checkbox"/>	<input type="checkbox"/>
metal		<input type="checkbox"/>	<input type="checkbox"/>