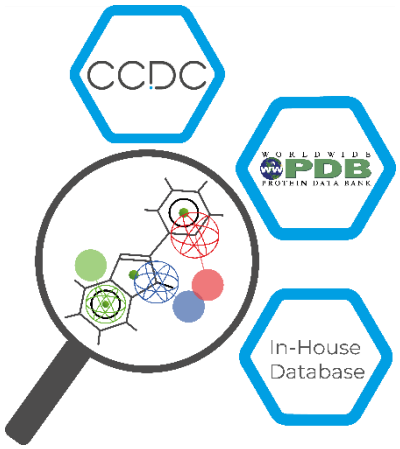
Using CSD-CrossMiner to Create a Feature Database (CROSS-005)

Developed using 2024.1 CSD Release

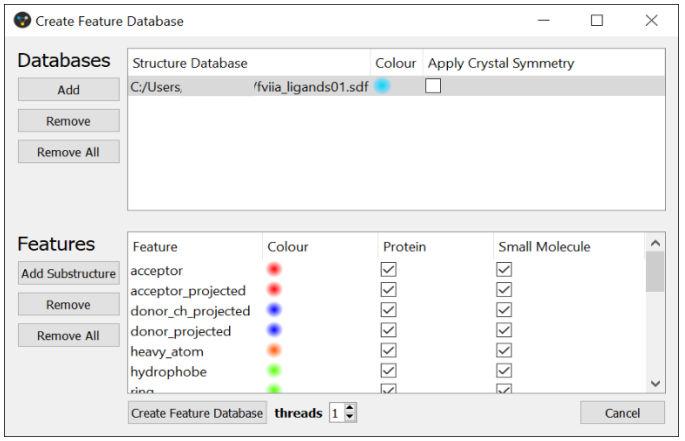


CCDC
advancing structural science

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). However, CSD-CrossMiner can also create such databases when supplied with a set of input molecules, and these can be searched with a pharmacophore query in the same way. We will see how this can be achieved in this workshop.

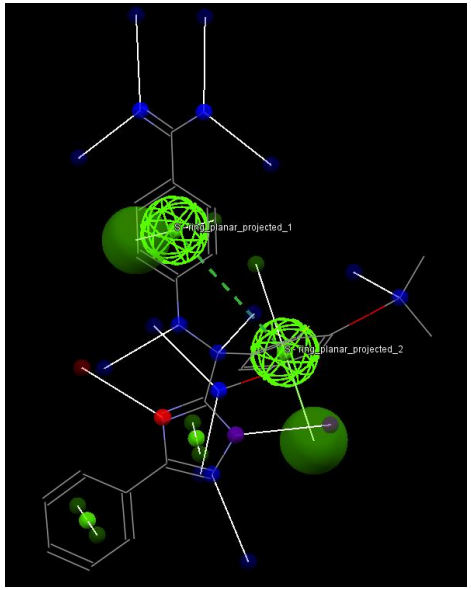


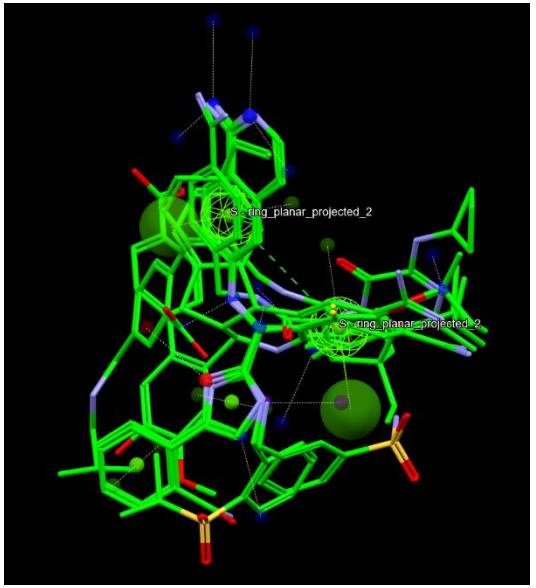
Learning Outcomes

After completing this workshop, you will be able to:

- Create a Feature Database from a group of structures
- Use a custom Feature Database in a pharmacophore search

This workshop will take approximately ***30 minutes*** to be completed. The words in *Blue Italic* in the text are reported in the CSD-CrossMiner Terminology at the end of this handout.





Pre-required Skills

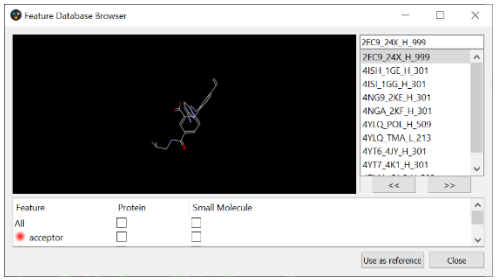
This tutorial is geared towards CSD-CrossMiner users who are already familiar with the software and its terminology, however a summary is provided at the end of this handout. If you are unfamiliar with CSD-CrossMiner, we suggest working through workshop CROSS-002.

Materials

The files to perform this tutorial are provided in the CROSS-005 folder here.

Example 1. Building your own Feature Database

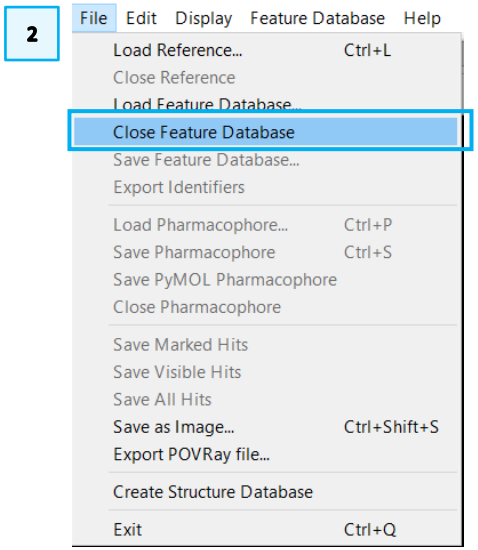
For this exercise you will be building your own *Feature Database* from a set of 3D coordinates of several ligands. When creating a feature database, *feature definitions* are applied to the structure database. All *Structure Databases* must be converted simultaneously into a single feature database to ensure homogeneity in the feature definitions.

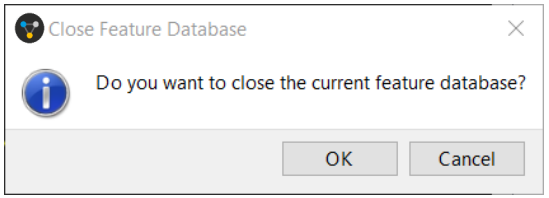


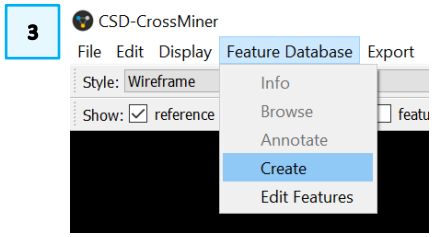
The Feature Database created from 10 structures in this workshop.

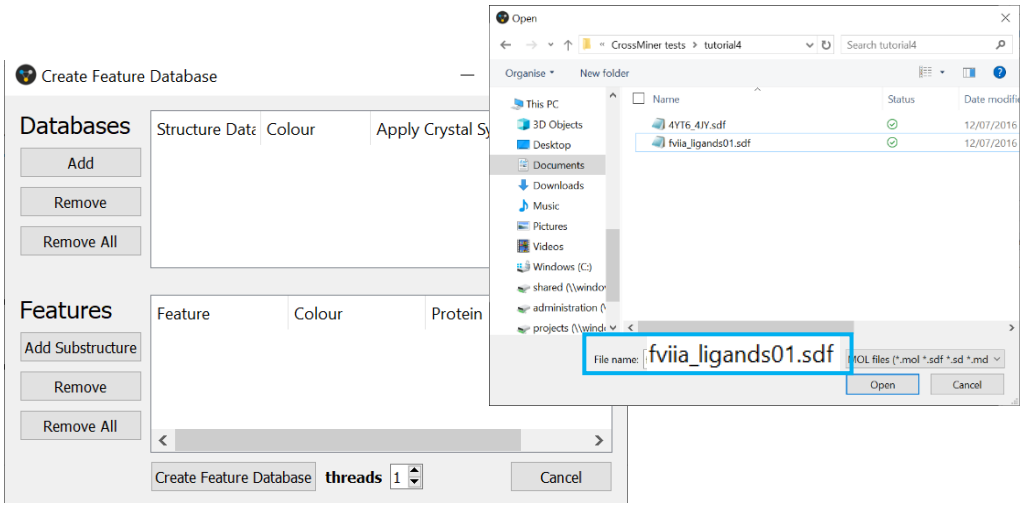
1. Launch CSD-CrossMiner clicking on the CSD-CrossMiner icon: .
2. If you have a CSD-CrossMiner session open with a feature database already loaded, select *File > Close Feature Database* to close the feature database. Then click **OK** in the *Close Feature Database* pop-up dialogue. CSD-CrossMiner may become unresponsive for a short time whilst the feature database is being closed.

   Alternatively, if you are starting a new CSD-CrossMiner session, you can stop the default loading of the feature database by clicking on **Cancel** in the *Loading…*pop-up dialogue, and then **OK**. This will open CSD-CrossMiner without any feature database loaded.
3. Select *Feature Database > Create* from the CSD-CrossMiner top-level menu**.**
4. In the *Create Feature Database* dialog, structure databases (*Databases*) and feature definitions (*Features*) can be specified to create a new feature database.
5. Click on **Add** in the *Databases* section of the **Create Feature Database** window and navigate to the CROSS-005 folder you downloaded to load the *fviia_ligands01.sdf* structure database.









6. Select the loaded structure database by clicking on it in the **Create Feature Database** window. Click on the sphere under the *Colour* column to associate a specific colour to the new feature database. Leave the *Apply Crystal Symmetry* unticked.

Now we must load the feature definitions that will be assigned to all the structures contained in the loaded structure database.

7. Click on the **Add Substructure** button in the *Features* section of the **Create Feature Database** window. From the **Select Feature Definitions** window go to the

   <CCDC installation folder>\ccdc-software\csd-crossminer\feature_definitions\any

   folder and select every entry in the "any" folder (you can use the **Shift** key and select the first and last file of the folder or **Ctrl-A** to select all the content of the folder) and then click **Open.** This will load the features applied to protein, nucleic acids and small molecule structures in the feature database provided with CSD-CrossMiner.

   The loaded features will be listed in the *Features* section of the **Create Feature Database** window, along with their name and colour.

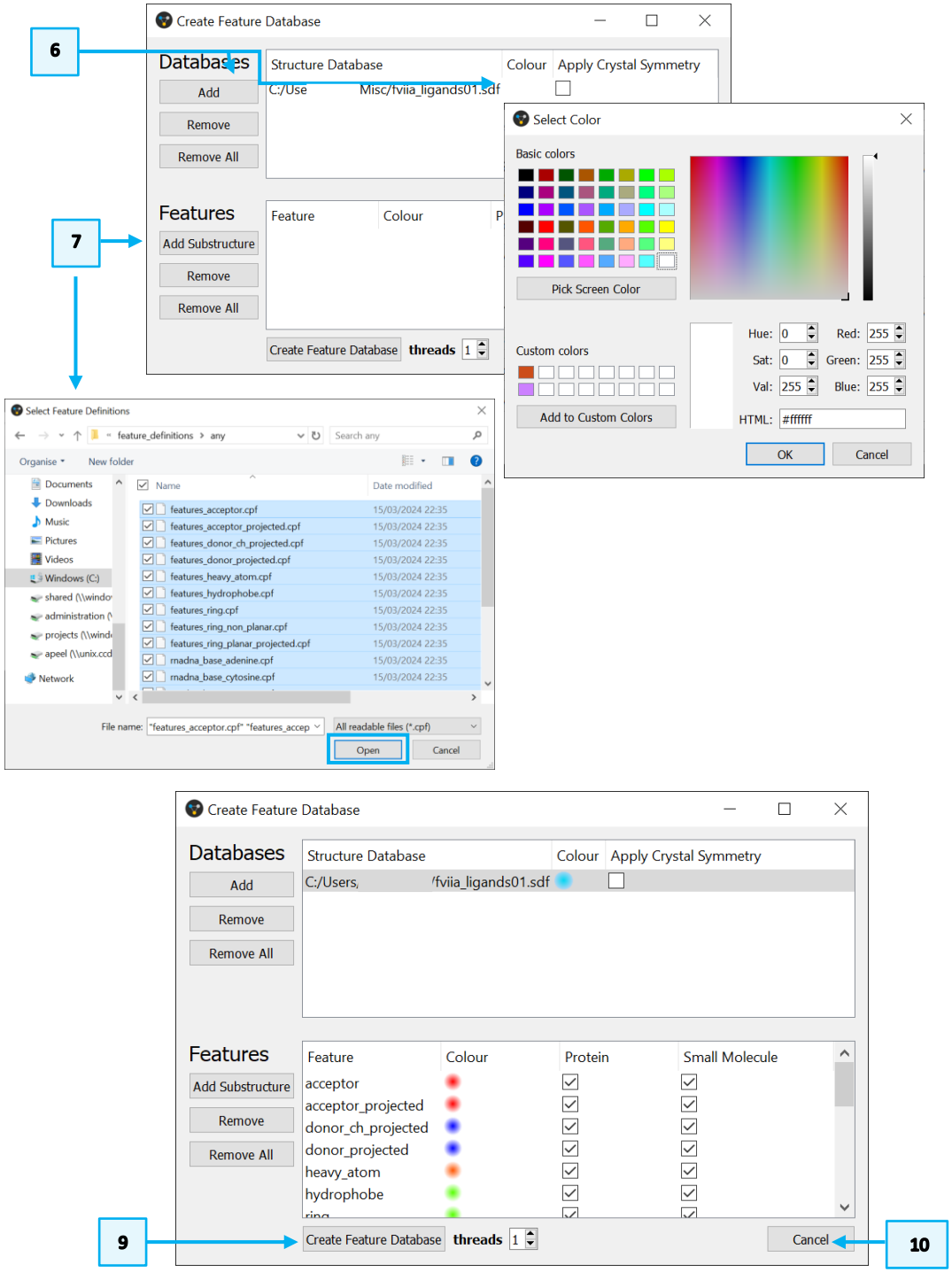
8. Repeat the procedure described in the Step 6. for the features contained in the "small_molecule" folder.

Because the structure database provided in the CROSS-005 folder contains only small molecules, we can ignore the features associated exclusively to proteins contained in the protein folder.

9. Click **Create Feature Database**, to create the database. It will ask you for a save location, please indicate the tutorial data directory and name the database *fviaa_ligands01.feat*.

Note that, depending on the number of structures in the structure database(s) and the number of feature definitions loaded, the creation of a feature database can be computationally expensive. Therefore, the feature database creation can be distributed across multiple CPU cores by specifying the desired number of cores in the **threads** spin-box.

10. Once the indexing has completed, click **Cancel** to close the **Create Feature**



Database window.

11. Try out the new database by clicking on *File > Load Feature Database* from CSD-CrossMiner top-level menu to load it. Now in the **Feature Databases** window, only your database is listed.

Note that the new feature database will be then loaded by default the next time you start a CSD-CrossMiner session.

12. The simplest test is to see if you can pull back a molecule from the original data set. Load the reference molecule *4YT6_4JY.sdf* from the CROSS-005 folder clicking on *File > Load Reference* from the top menus.

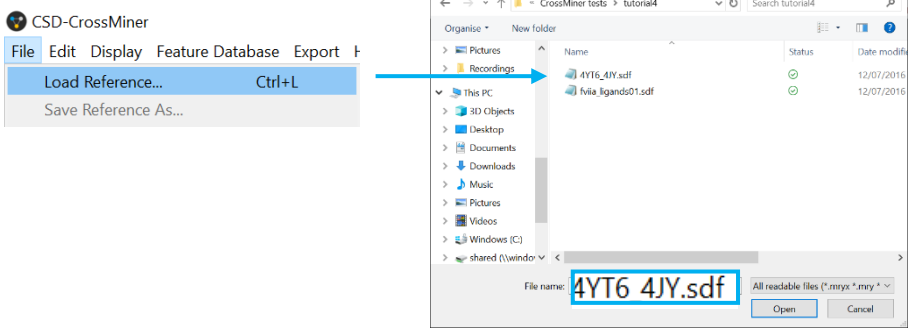
13. The displayed features of the reference structure are ticked in the *show in reference* column in the *Pharmacophore Features* window.

    Tick the *ring_planar_projected* feature tick-box to show all the molecule's planar ring in the 3D view.

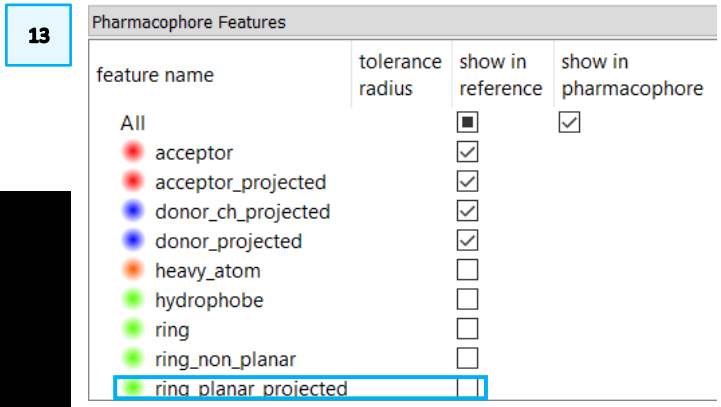
14. Define two *ring_planar_projected* pharmacophore points as indicated in the image on the right by right-clicking on the *ring_planar_projected* features in the 3D view and selecting the top menu item in each case. Make sure they have *intra* constraints by clicking on  and the click on  to start the search.

11

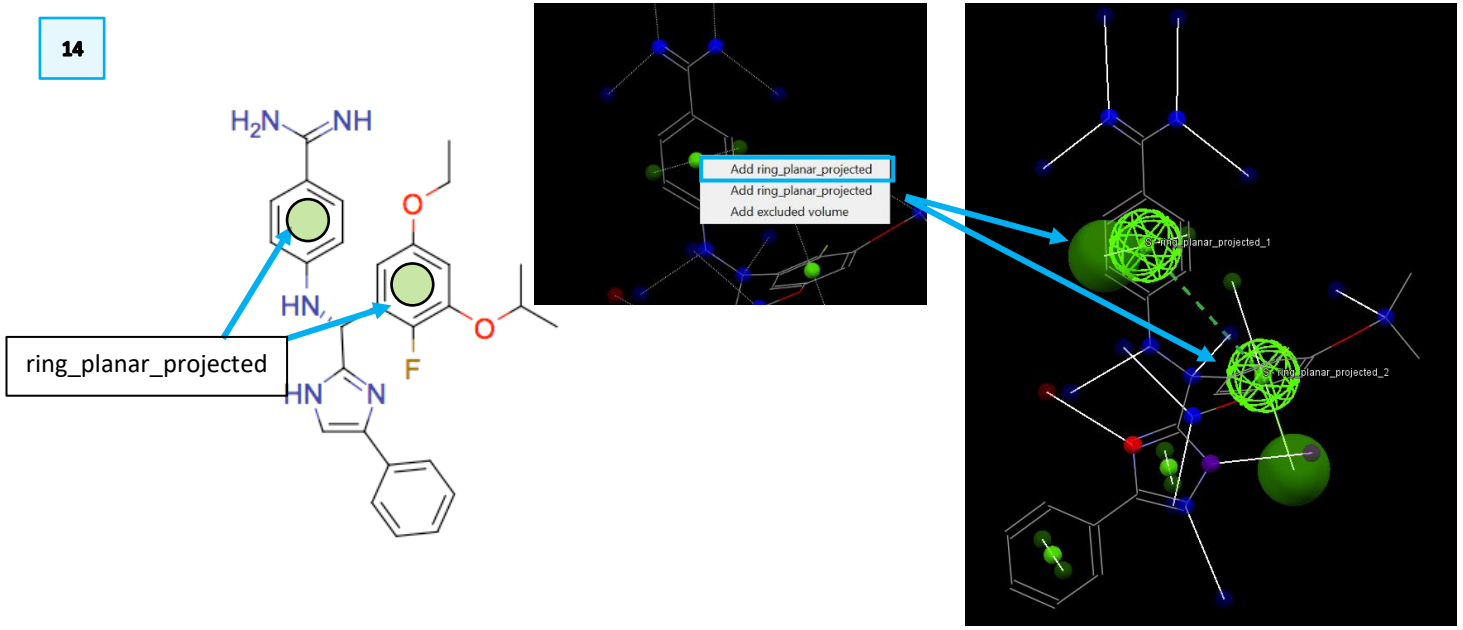
12



13



14

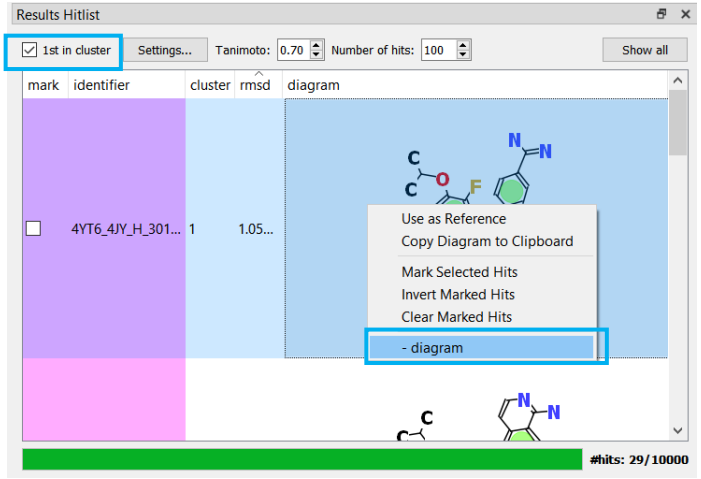


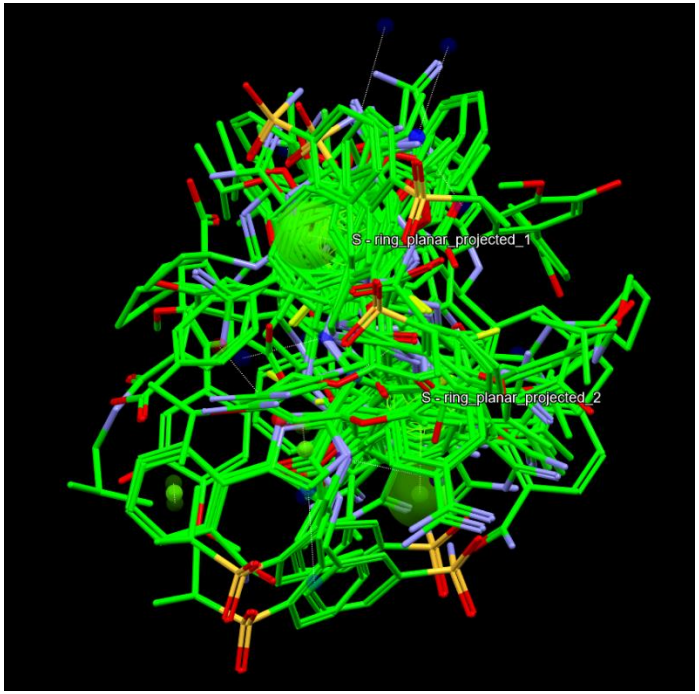
Note that the two options in the *Add ring_planar_projected* define a different placement of the projected ring indicating the potential position of the coupling feature.

- If you pick the *Add ring_planar_projected* options as shown in the 3D image, you should get 29 hits.
- You can hide the 2D diagram in the *Results Hitlist* window by right-clicking on one of the hits and select – *diagram* from the options available.
- By doing so you will see that with the 29 structures are clustered in 5 results. To see all 29 results untick the **1st cluster** tick-box.

If you play with different combinations of the **Add ring_planar_projected** options, you will have different number of matching hits.

When you don't have any information about the location of the coupling feature, any of these combinations is equally valid and all possible combinations should be explored.





Conclusion

It is quite straightforward to create a custom Feature Database from a set of structures files and the features can be indexed in the same way as the default databases.

Summary

In this workshop we have seen how to create a CSD-CrossMiner Feature Database. You should now be able to create such a database from a set of structure files, which might be your own or a specific set of structure from the CSD.

For your reference, you can find the user manual at this link.

Next Steps

After this workshop, you can continue learning about CSD-CrossMiner with more exercises available in the self-guided workshops available in the CSD-Discovery workshops area on our website.

https://www.ccdc.cam.ac.uk/Community/educationalresources/workshop-materials/csd-discovery-workshops/

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-005. Thank you!

CSD-CrossMiner Terminology

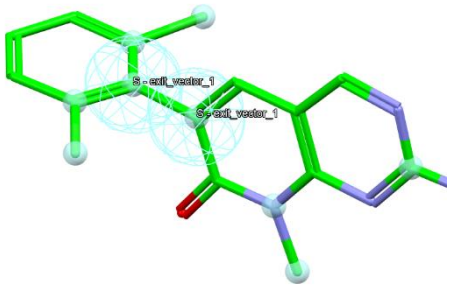
Exit vector
A two-point feature that represents a single, non-ring bond between two heavy atoms features; and it will be represented as two mesh spheres. In the case of CSD-CrossMiner, directionality in an exit vector does not matter.

Features
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
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.

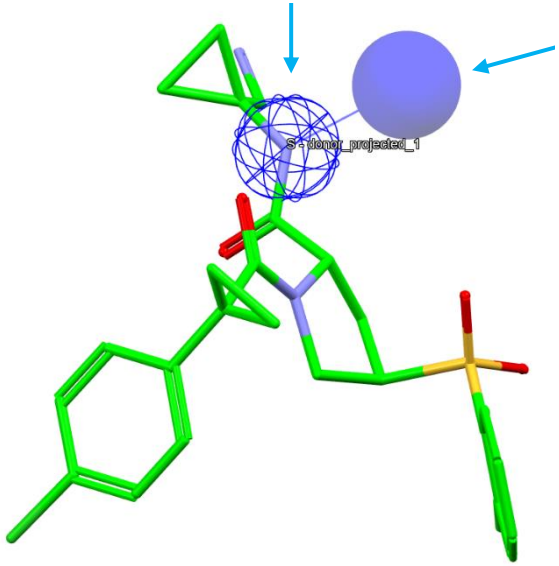
Feature database
A database containing the structures from the structure database, indexed with a set of feature definitions provided by CSD-CrossMiner and any additional features defined by the user. This is the database that CSD-CrossMiner uses to perform the actual 3D search against a pharmacophore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



An exit vector (light blue mesh spheres) defined by the position of two carbon atoms.

Base point (the heavy atom of the donor group)

Virtual point - defines the direction the X-H group should point (Base point →Virtual point)



A molecule with a donor_projected pharmacophore point defined.

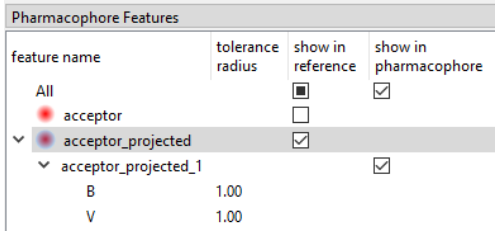
Features and Pharmacophore Representation

In the CSD-CrossMiner 3D visualiser, 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:

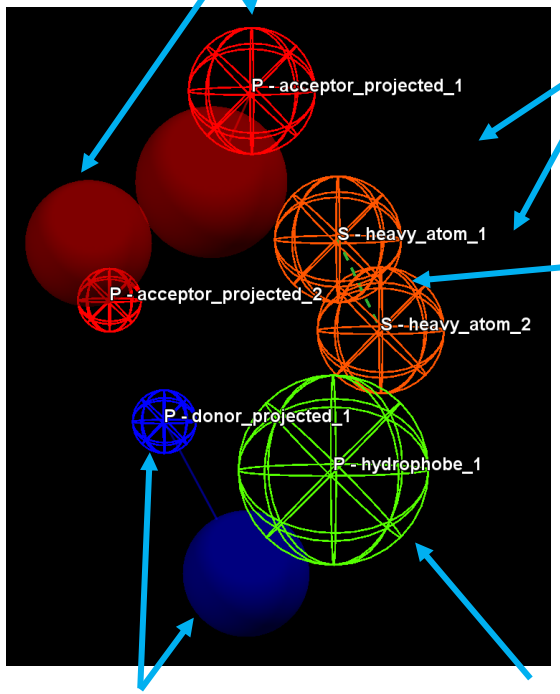
- **P**: Protein pharmacophore point
- **S**: Small molecule pharmacophore point
- **A:** Either a small molecule or protein pharmacophore point
- **Dashed line**: intra and intermolecular constraints. Constrained features must belong to either the same molecule as each other (*intra*, dashed green line) 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 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