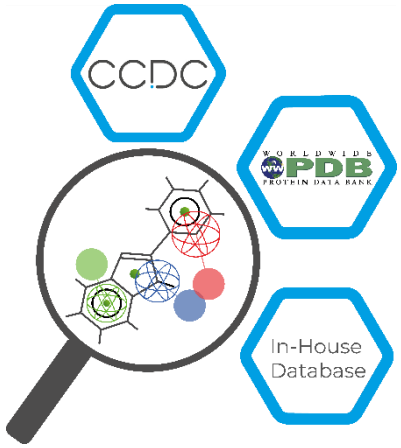
Creating a Pharmacophore Query from a Reference Molecule & Scaffold Hopping in CSD-CrossMiner

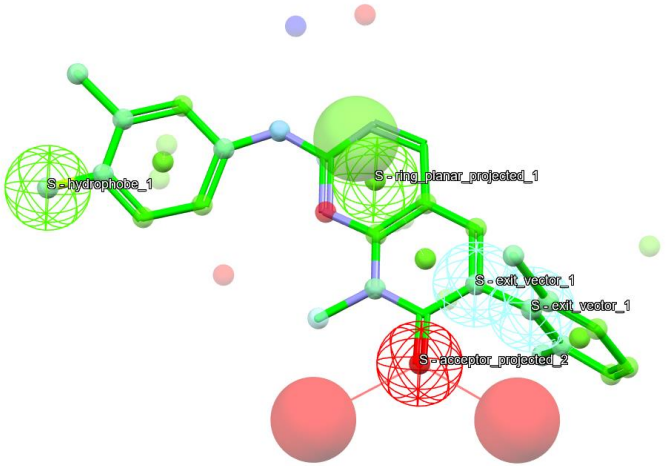
Developed using 2024.2 CSD Release



CCDC
advancing structural science

Table of Contents

Introduction

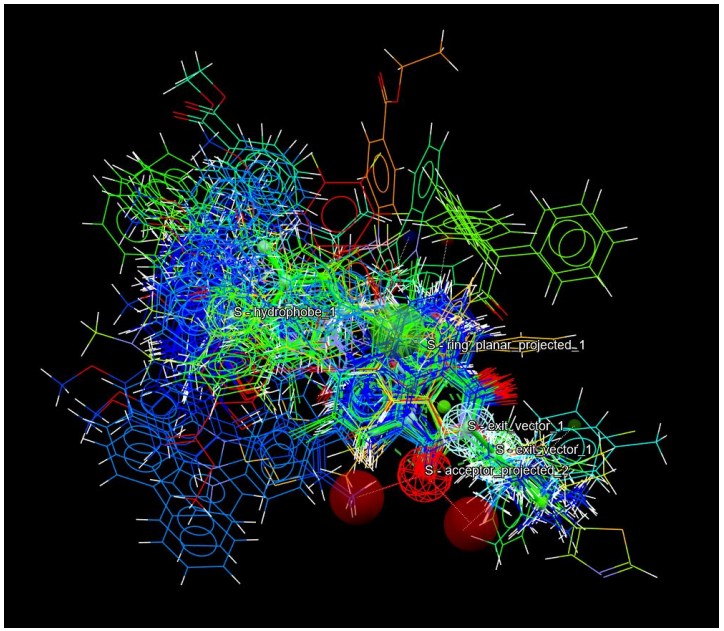


CSD-CrossMiner can be thought of as a pharmacophore-based query tool. However, it is much more powerful than traditional pharmacophore query tools as it allows you to query not only databases of ligands, but also proteins and protein-ligand interactions. CSD-CrossMiner includes a preconfigured database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

Learning Outcomes

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.

This workshop will take approximately ***25*** minutes to be completed. The words in *Blue Italic* in the text are reported in the Glossary at the end of this handout.



Pre-required Skills

There are no pre-required skills for this workshop, however some knowledge of the representation of pharmacophores and features in CSD-CrossMiner is expected. A summary can be found at the end of this handout.

Materials

The files to perform this tutorial are provided in the CROSS-004 folder here.

Please be aware that some of the results may vary depending on your version of CSD-CrossMiner and the associated databases.

Example 1. 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.

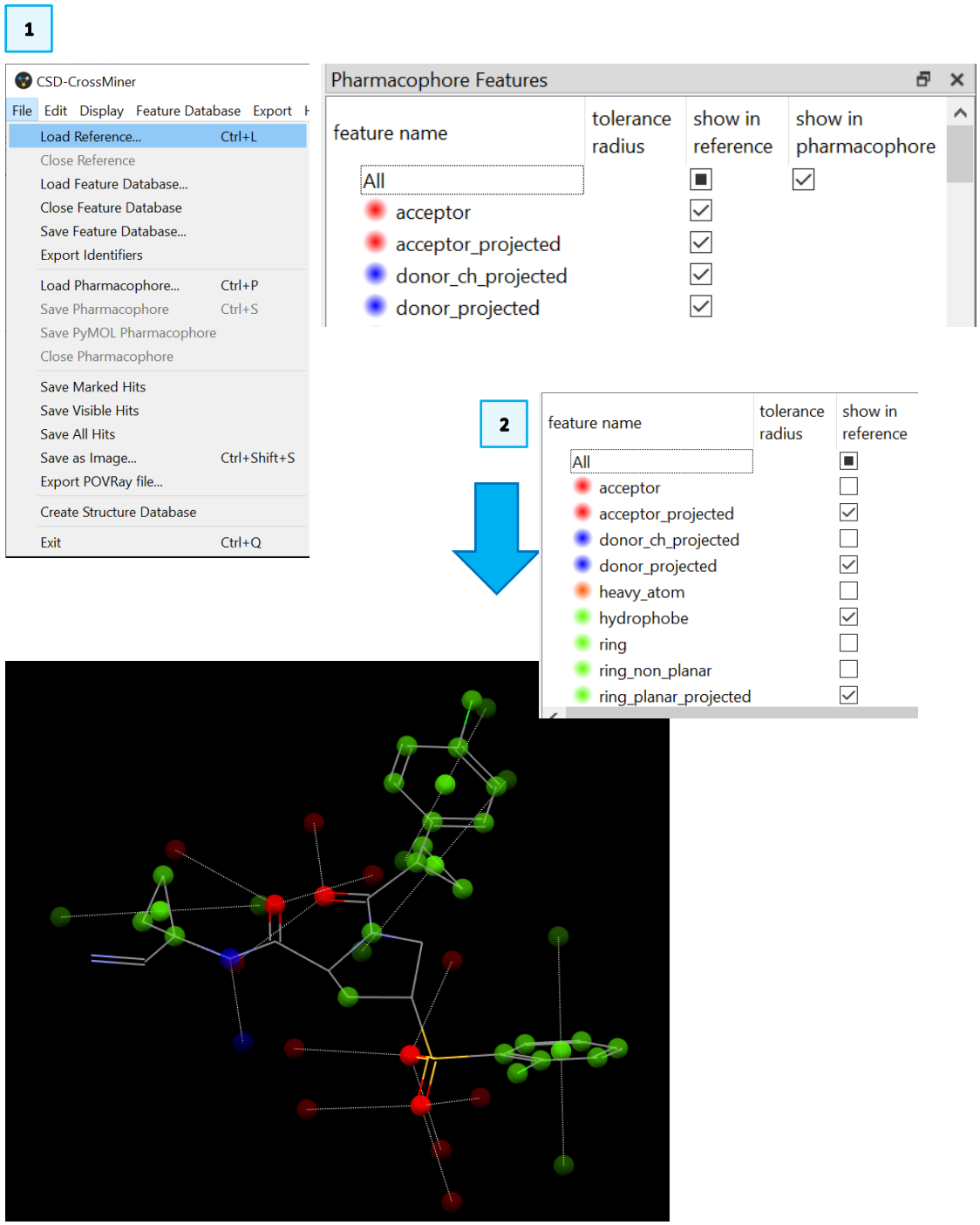
Launch CSD-CrossMiner clicking on the CSD-CrossMiner icon: . Wait a few minutes for loading and initialising. If you have already been working in CSD-CrossMiner, clear the 3D view by clicking on *File > Close Pharmacophore* and/or *File > Close Reference* from the CSD-CrossMiner top-level menu.

1. Click on *File > Load Reference* and select "2xu1_ligand.sdf" provided in the CROSS-004 folder that you have downloaded.

   By default, only (*hydrogen bond*) *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.

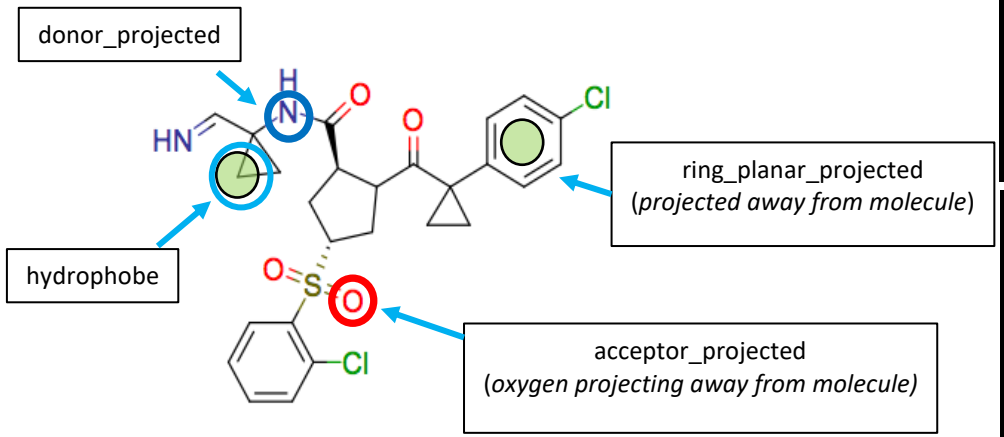
   Note that if a different choice of displayed features is made in the CSD-CrossMiner session (e.g. all features displayed), these new settings will be remembered when a new reference molecule is loaded (e.g. in that case all features of the new reference molecule would be displayed).

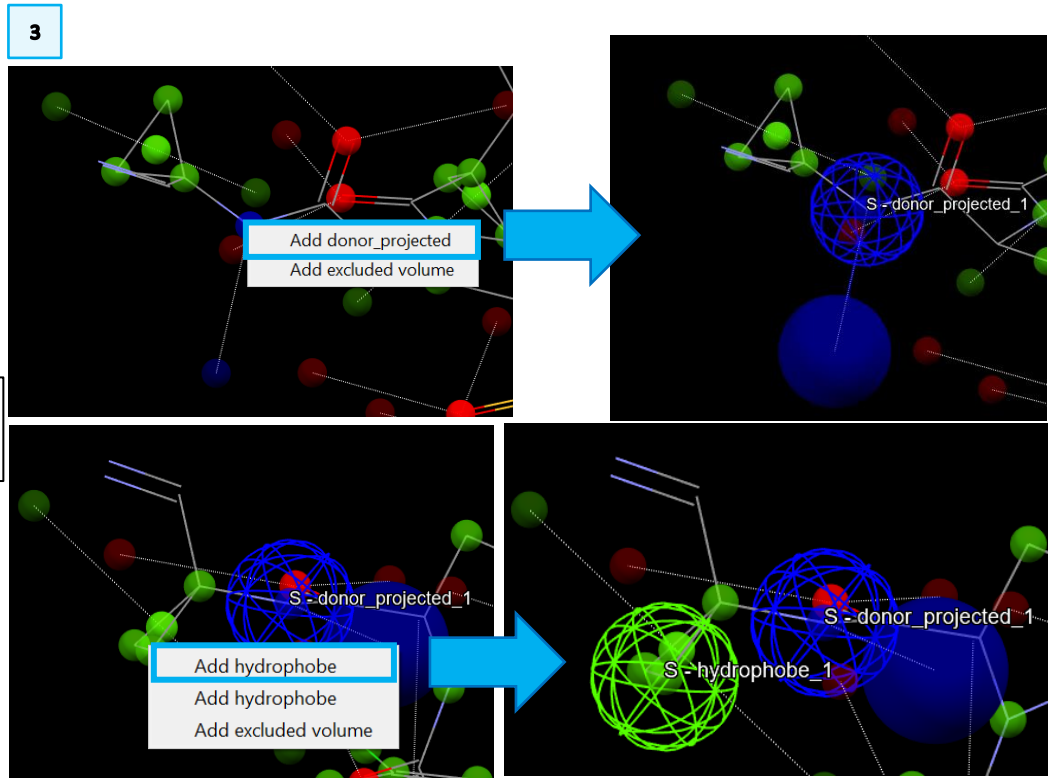
2. To make it easier to choose which atom to add a pharmacophore to, toggle the checkboxes in the *show in reference* column of the **Pharmacophore Features** window. Click two times on the **All** tick-box in the **Pharmacophore Features**, so that no features are shown. Then, you can select individual tick-boxes to display only the required feature types, as needed in the following

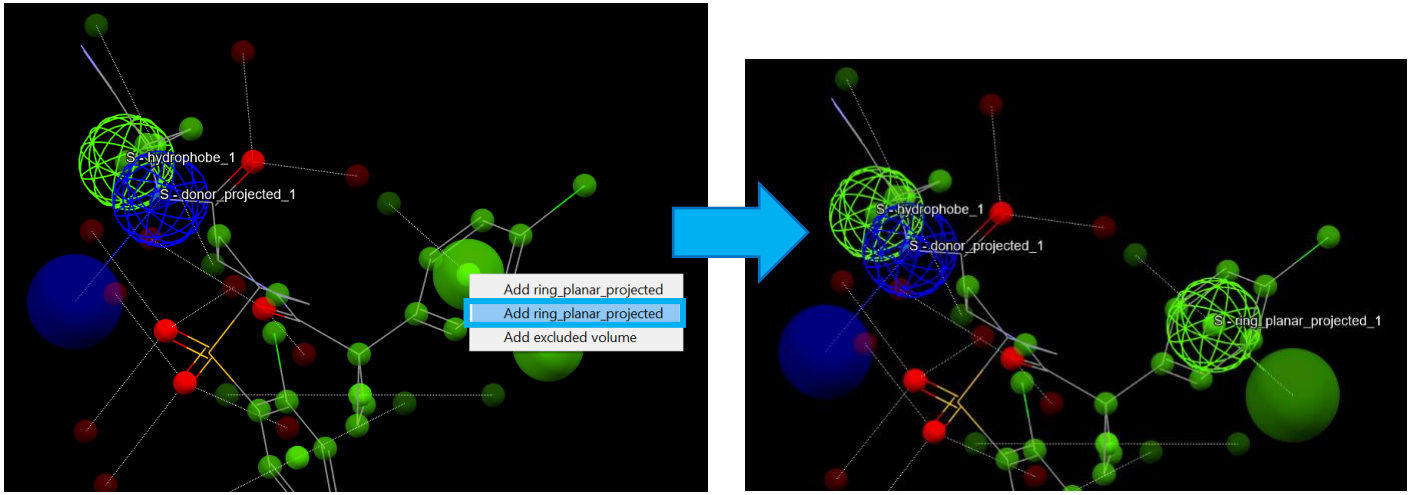


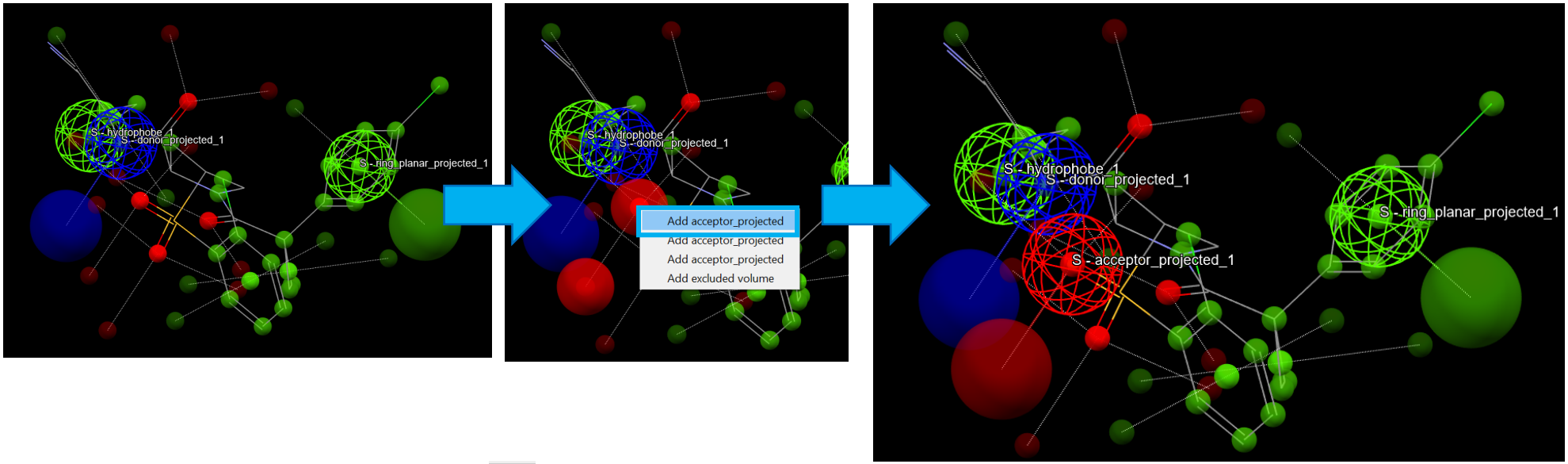
step (i.e., **acceptor_projected**, **donor_projected,** **hydrophobe** and **ring_planar_projected**).

3. Right-click on features of the reference molecule in the 3D view to define *pharmacophore points* as shown in the images below and 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.









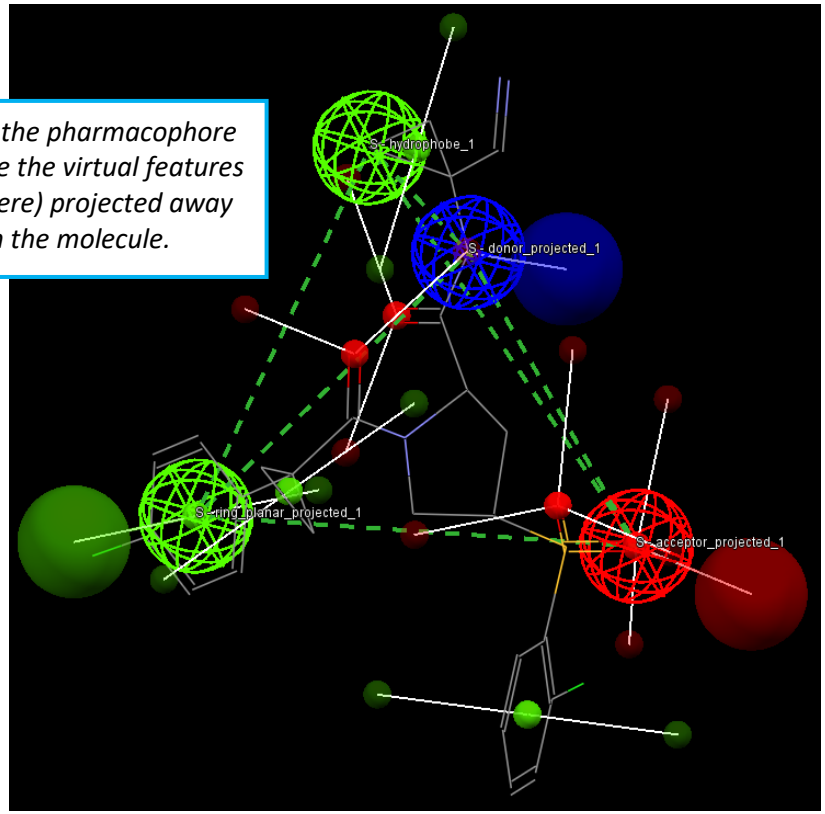
4. When all pharmacophore points are defined, click on  to set all defined pharmacophore points to be part of the same molecule.

5. From the *Feature Databases* window, deselect the CSD databases tick-boxes. These are usually identified with a yellow colour and would have a name similar to **csd545_crossminer** and additional boxes could be present for data updates (see the picture for the 2024.2 version with one data update). 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. This will take a few minutes to complete.

4

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

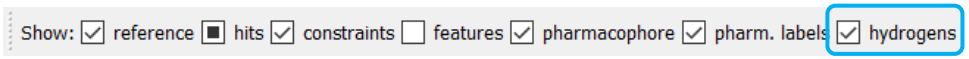


7. Sort the hits in the **Results Hitlist** by *RMSD* by clicking on the *rmsd* column in the **Results Hitlist** browser. The lowest RMSD hit in this example is `2XU1_m1_A_bs_424_A_1221_1`, which corresponds to the co-crystallised ligand structure. Clicking on one hit will display that in the visualiser. Note that the number of results may vary based on the database version you are using.

8. Activate the **Colour:** ☐ **Hits** check-box in the CSD-CrossMiner toolbar to colour the different clusters with different colours colour.

9. You can change the style and the colour of the molecules in the 3D view using the **Style:** toolbar or alternatively by left-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.

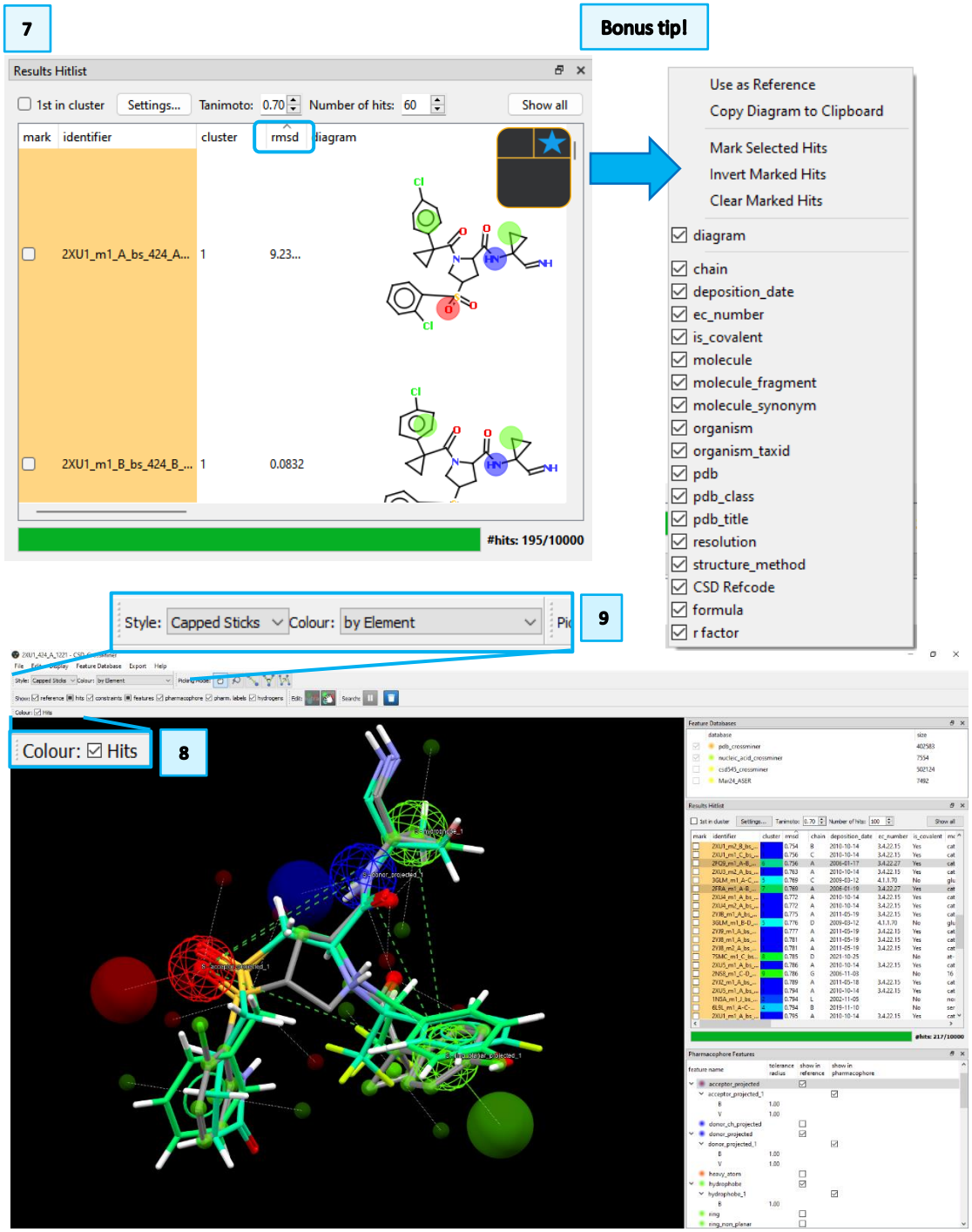
10. Note that you can hide the hydrogen atoms by disabling the **hydrogens** checkbox in the *Show:* toolbar.



11. Scrolling through the hits with higher RMSD reveals interesting analogues. As an example, see `2FQ9_m1_A-B_bs_CRJ_A_999_1`, `2FRA_m1_A-B_bs_CRV_A_999_1`. **Tip**: you could sort the results alphabetically, clicking on the *identifier* header.

Conclusion

In this example, we have used CSD-CrossMiner to identify potential small molecule ligands extracted from the protein-ligand binding sites of PDB structures that match defined pharmacophore features of a reference ligand. A relative quick and straightforward process has revealed numerous candidates for further investigation.



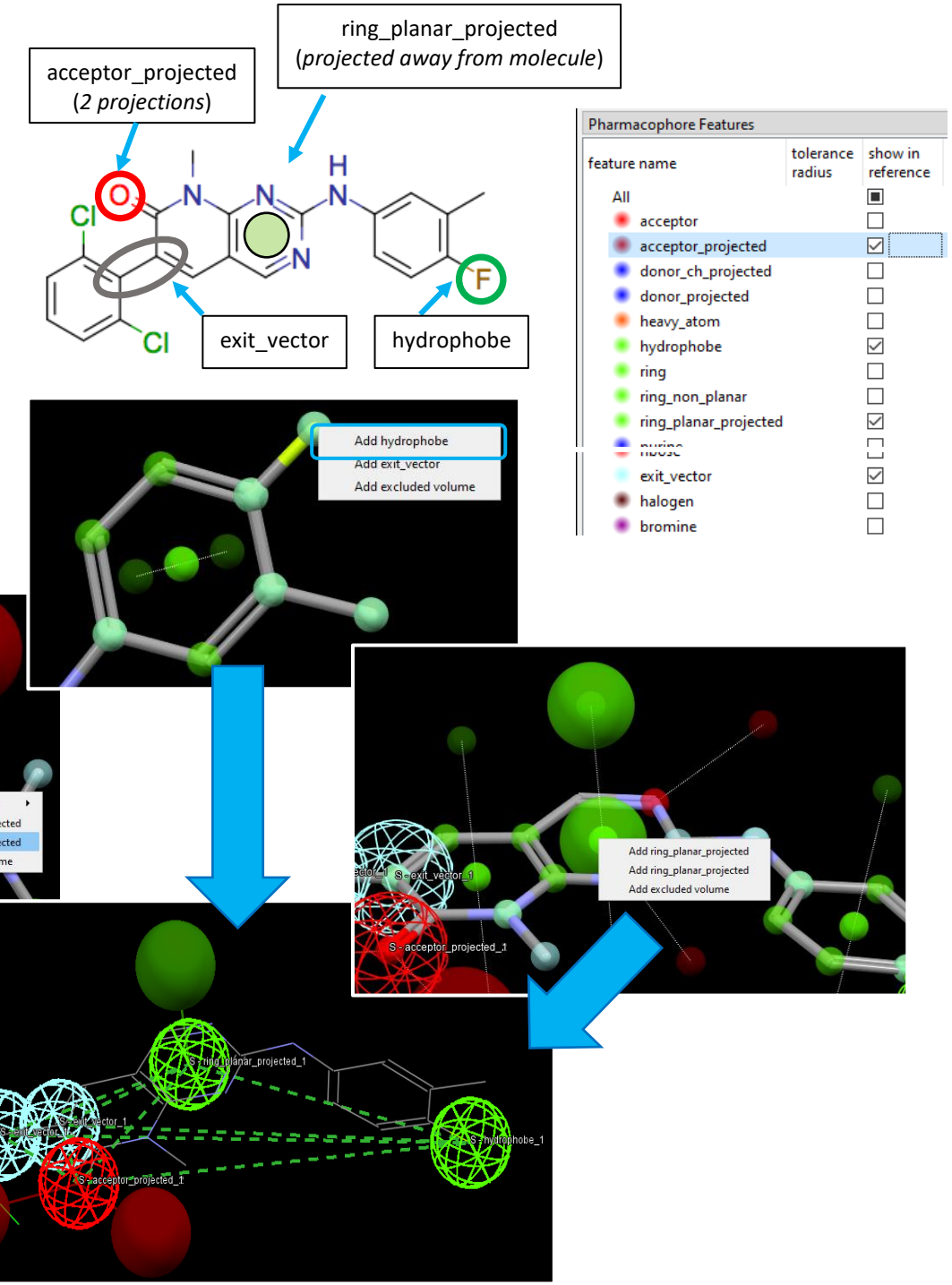
Example 2. Scaffold hopping using CSD-CrossMiner

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. From *File > Load Reference*, load the file "2hzi_ligand.sdf" from the CROSS-004 folder that you downloaded.

3. Define the pharmacophore as indicated in the picture. For the *acceptor_projected*, two points are indicated – ensure that both projected virtual points are set to features. Once set up, 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 all the tick-boxes checked in the **Feature Databases** window. Remember that the number associated with the CSD database and update might change based on the version you have installed.

5. For this example, we are only interested in organic structures therefore, we can customise the search to skip organometallic structures. To do so, go to the the *Edit>Options* menu and make sure the **Skip organometallic structures** tick-box is enabled. From this same menu it is possible to change the number of processors dedicated to the pharmacophore search by changing the **Number of threads**. Tip: Note that the pharmacophore search options (**Options**) cannot be changed when the search is running or when the search is paused.

6. Click on the  button to start the pharmacophore search. The search will take several minutes to complete.

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

Exercise

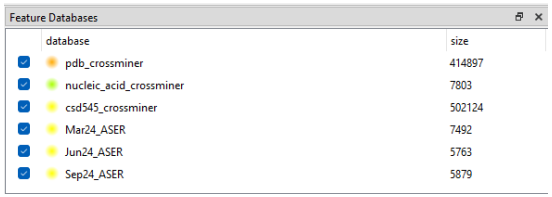
Try experimenting with narrowing this result set down:

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

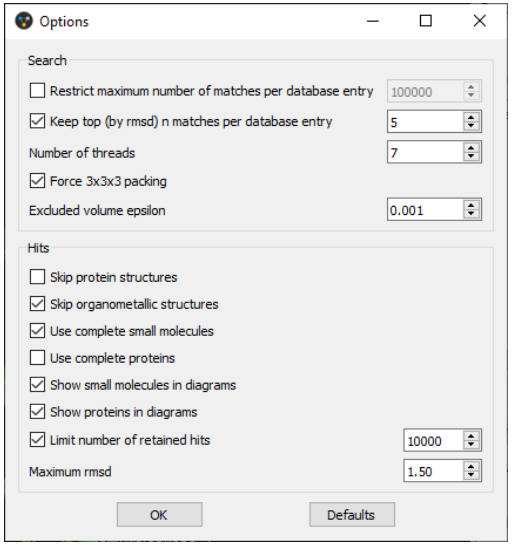
Conclusion

In this example, we have seen how to include extra pharmacophore features (*hydrophobes* and *exit vectors*) in a CSD-CrossMiner search. By including both CSD and PDB structures, a large number of molecules of potential interest for scaffold hopping have been uncovered.

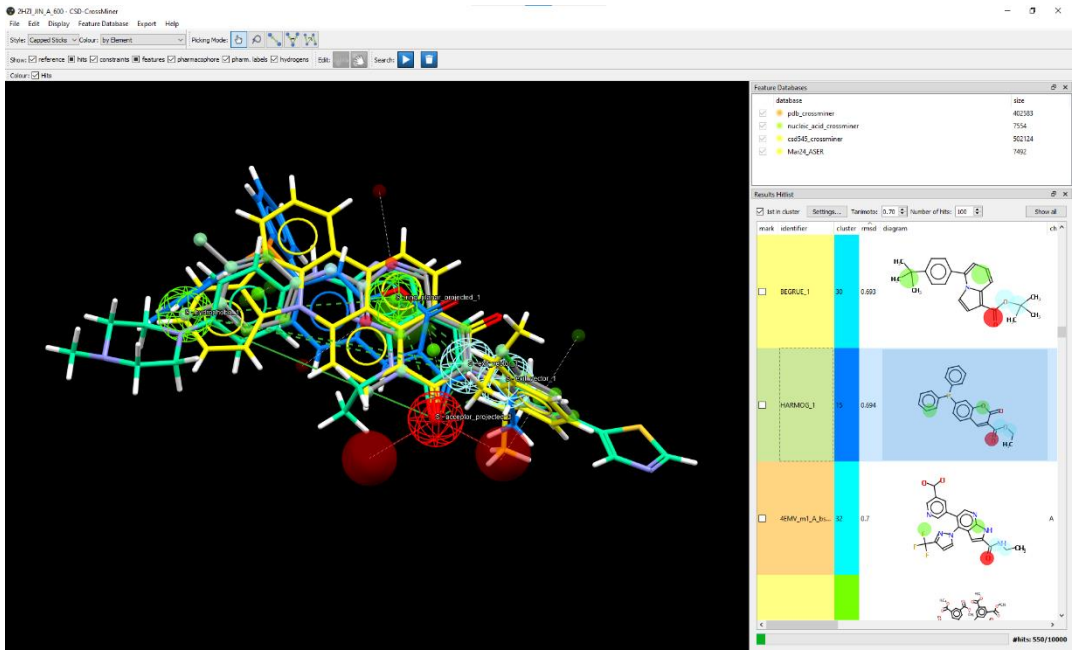
4



5



7



Summary

In this workshop, we have seen how to create a CSD-CrossMiner query by defining pharmacophore features (including directional information, by using *projected* features) from a reference molecule. We have also explored using the different feature databases supplied with the CSD-CrossMiner installation to uncover potential alternative scaffolds that might allow patent breaking or enhanced affinity and selectivity.

It is relatively simple to mine similar compounds in this way and quickly assess the match of the hits generated. While such a search is possible in other CSD applications, such as ConQuest, the query in these tools is more challenging to create; thus CSD-CrossMiner provides a more convenient method for interrogating possibilities in 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/

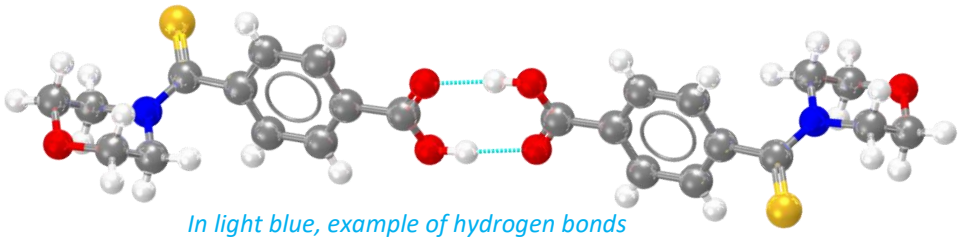
Glossary

Hydrogen Bonds
Hydrogen bonding occurs between donor-acceptor interactions precisely involving hydrogen atoms. The H-bonds interactions are classified as: strong (mostly covalent), moderate (mostly electrostatic) and weak (electrostatic). Their strength is observed to be between 12 and 30 kJ/mol.

Hydrogen Bond Donor/Acceptor
If a typical hydrogen bond is depicted as X—H···Y—Z, where the dots denote the bond, X—H represents the hydrogen bond *donor*. The *acceptor* may be an atom or anion Y, or a fragment of a molecule, Y—Z, where Y is bonded to Z. The acceptor is an electron-rich region such as, but not limited to, a lone pair on Y or a π-bonded pair of Y—Z. [Source: E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. Dannenberg, P. Hobza, H. G. Kjaergaard, A. C. Legon, B. Mennucci and D. J. Nesbitt, *Pure Appl. Chem*., 2011, **83**, 1637 – 1641.]

Hydrophobic/hydrophobe
Hydrophobic molecules effectively "repel" water and thus have a tendency to self-aggregate in aqueous media, excluding water in so doing. On a structural level, these are non-polar groups such as alkyl or aryl moieties. If these functional groups or molecular fragments are also pharmacophore features, then they are called *hydrophobes* in CSD-CrossMiner.

Root Mean Square Deviation (RMSD)
The root mean square deviation (RMSD) is a commonly used measure of the difference between two sets of values (usually comparing observed data to estimated data). The RMSD is defined as the square root of the mean squared error.



In light blue, example of hydrogen bonds for refcode MULWIC.

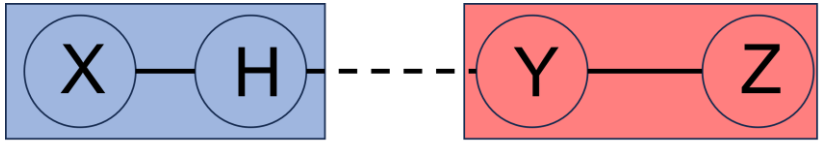
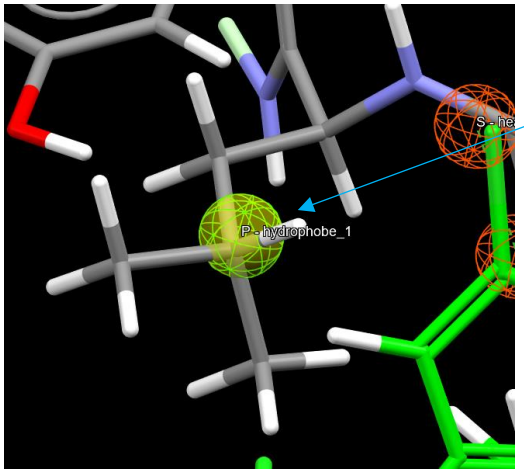


Illustration of a hydrogen bond interaction with between hydrogen bond donor X–H and hydrogen bond acceptor Y–Z.



Hydrophobe pharmacophore point.

An isobutyl group is hydrophobic. The green mesh sphere indicates the position at which such a feature (functionally a hydrophobe) must be found.

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.

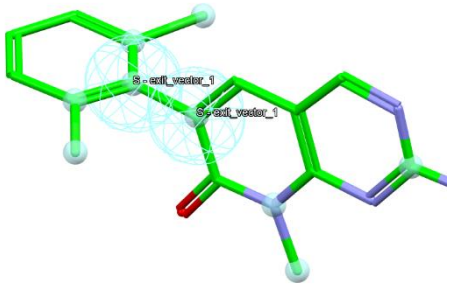
Feature database
A database containing the structures from the structure database, indexed with a set of feature definitions provided by CSD-CrossMiner and any additional features defined by the user. This is the database that CSD-CrossMiner uses to perform the actual 3D search against a pharmacophore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

Hits
The results of the search which satisfy the pharmacophore query.

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.

Scaffold hopping
Scaffold hopping is the identification of isofunctional molecular structures with chemically completely different core structures.

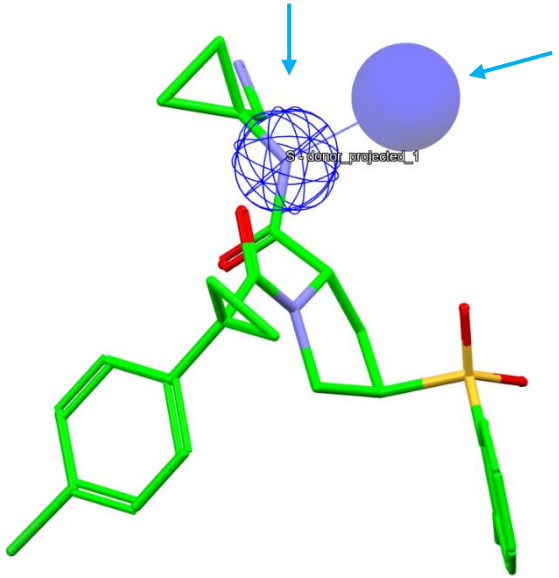
Structure database
Is a database containing the 3D coordinates of small molecule structures and/or protein-ligand binding sites. This database is used to create a feature database.



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.

Tanimoto index

The Tanimoto index is the ratio of the number of features common to both molecules to the total number of features, i.e.

$$T(A,B) = (A \cap B)/(A + B - (A \cap B))$$

where *A* and *B* are the number of attributes of object *a* and *b*, respectively. It can be thought of as a measure of similarity.

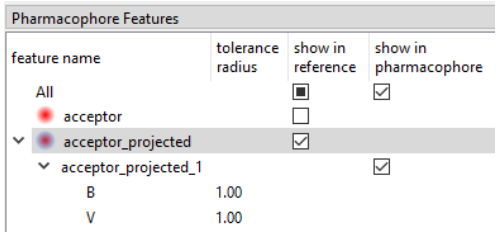
Features and Pharmacophore Representation

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:

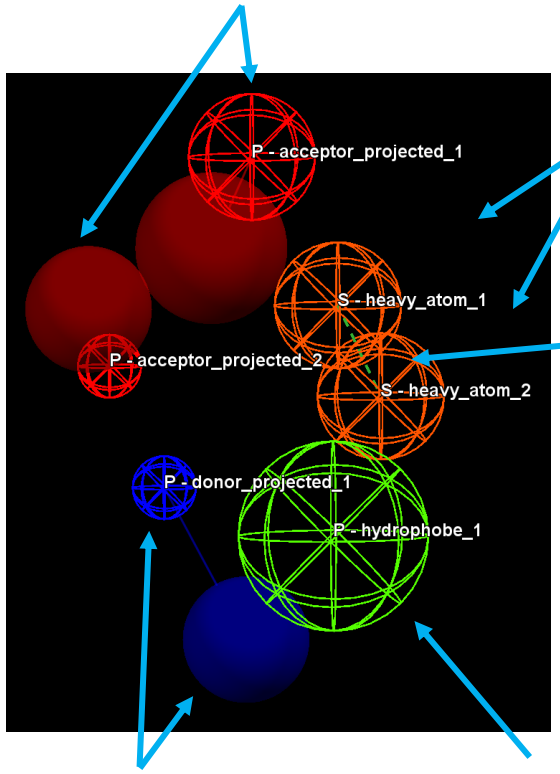
- **P**: Protein pharmacophore point
- **S**: Small molecule pharmacophore point
- **A:** Either a small molecule or protein pharmacophore point
- **Dashed line**: intra and intermolecular constraints. Constrained features must belong to either the same molecule as each other (*intra*, dashed green line)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