Creating Substructure Features in CSD-CrossMiner

Version 2.0 – November 2018
CSD-CrossMiner v1.5.0
Customising a Pharmacophore Search Using CSD-CrossMiner

Pharmacophore searching is a key component in many drug discovery efforts and represents an effective mechanism of virtual screening.

In this approach, a pharmacophore query is created to describe features that are essential for the molecule to carry out its function. The query is then used to identify new possible lead compounds by searching a three-dimensional (3D) structural database (Figure 1).

CSD-CrossMiner provides the possibility to search crystal structure databases such as the Cambridge Structural Database (CSD) and the Protein Data Bank (PDB) in terms of pharmacophore queries (Figure 2).

In addition to the common pharmacophore features including hydrophobic, hydrogen bond donor, and hydrogen bond acceptor, with CSD-CrossMiner it is possible to create customised features that allow the inclusion of more specific chemistries in the pharmacophore search.

In this workshop, you will learn how to create a new substructure feature in CSD-CrossMiner and how to make use of this newly generated feature in a pharmacophore search.

The files to perform this workshop are provided in the workshop1 folder here.

Figure 1. Pharmacophore example

Figure 2. CSD-CrossMiner allows CSD and PDB databases to be searched in terms of pharmacophore queries.
Case Study

Introduction

Sodium glucose cotransporters (SGLTs) have recently attracted considerable attention as new drug targets for the treatment of diabetes. In particular, the selective inhibition of the SGLT subtype 2 (SGLT2) could provide a highly effective method of glycemic control. By targeting renal tubular glucose reabsorption, SGLT2-selective inhibitors exhibit a novel mechanism of action resulting in excretion of glucose into the urine.

A recent study published by Yoshihito Ohtake et al. (DOI: 10.1021/jm300884k) proposes a new class of highly potent and SGLT2-selective inhibitors incorporating a unique spiroketal structure. By performing a structural database search using a 3D pharmacophore model based on the superimposition of known inhibitors, Ohtake et al. discovered a new potent O-spiroketal C-arylglucoside scaffold (Figure 4).

In this workshop, we will replicate this work, demonstrating how CSD-CrossMiner can be used to efficiently identify interesting hits that can suggest possible chemistries for use in new lead compounds.

Provided input files in workshop1:
- molecular_overlay.mol2, a structural alignment of four known SGLT2 inhibitors generated using the CSD Ligand Overlay tool (Figure 4).

Your task:
- Create a new substructure feature definition for a specific chemistry.
- Add the new feature definition to the loaded feature database.
- Create a pharmacophore query that includes the new feature.
- Perform a pharmacophore search in CSD database using a pharmacophore query that includes the unindexed feature.

Challenges:
The example used here mimics the situation where a researcher wants to search for specific chemistries or features in a crystal database that has not been indexed with the desired chemistry or feature.
Viewing the query overlay

1. Start CSD-CrossMiner by clicking on the CSD-CrossMiner icon: The csd_pdb_crossminer.feat feature database, provided in the CSD-CrossMiner package ($CSD_2019/crossminer_data), is loaded as default feature database.

2. Open molecular_overlay.mol2 by clicking on the main menu option File and then Load Reference from the resultant pull-down menu. This loads the overlay of four known SGLT2 inhibitors generated using the CSD Ligand Overlay tool available from the CCDC (See CSD Ligand Overlay).

3. By default, only the donor (blue) and acceptor (red) features associated with the loaded reference molecules are shown in the 3D view. The displayed features are represented in the 3D view as small translucent spheres and are ticked in the show in reference column in the Pharmacophore Features window.

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.

4. In the Pharmacophore Features window, uncheck the default displayed features by unticking the corresponding check-boxes in the show in reference column.

5. Check the ring_non_planar check-box. This will show the non-planar ring features present in the overlay.

6. Here, the non-planar ring feature corresponds to a glucose ring that is essentially conserved in all four SGLT2 inhibitors used in the overlay.
Creating a new substructure to a feature definition

Using the *ring_non_planar* feature to create the pharmacophore query will result in hits with different non-planar rings that match the glucose location.

However, from the overlaid structures, we know that among the non-planar rings, it is the glucose that is conserved in the SGLT2 inhibitors.

Therefore, we want to discriminate between all non-planar rings, and find hits with only the glucose ring. To do so, we can use CSD-CrossMiner to easily create a glucose feature on-the-fly and then use it to investigate the loaded database.

1. Right-click in the *Pharmacophore Features* window and click on *Add substructure*.

2. This will present the *Feature Editor* window.

3. Type “sugar” in the *Feature name* box and click the *Colour* button to specify a colour other than white.

4. Under *Feature Point Generators*, click *Add* to create a *simple point* feature.

5. Under *Substructure Definitions*, click the *Add* button. This creates a dummy SMARTS string. Edit the *SMARTS pattern* by double clicking on [*], replacing it with “\*C1OCCCC1” and then pressing Enter. The inserted pattern represents the SMARTS code of the sugar rings in the reference molecular overlay.

6. Double click on the 0 in the *indices* column and type “0 1 2 3 4 5” then press Enter to define a point based feature that will ensure that all six atoms in the glucose ring are included in the substructure.

The new feature definition is shown in the 3D display of the *Feature Editor* window as red (Small Molecule) feature points and listed in the right-hand panel under *Generated features* (small molecule: red) of the *Feature Editor* window.

Note that this list is associated with the four overlaid structures displayed in the 3D display.
7. In the Feature Editor window menu select File > Add Feature to Current Feature Database and then click Close to close the Feature Editor window. This will make the newly created feature available in CSD-CrossMiner session.

Creating the pharmacophore search

1. The ‘sugar’ feature is now displayed in the 3D view as translucent coloured sphere and ticked in the Pharmacophore Features window (scroll to the bottom of the Pharmacophore Features window to find it).

In the Pharmacophore Features window, you will notice a diagonal hatching on the ‘sugar’ feature name; this indicates that the feature has not been pre-calculated and therefore, the loaded feature database has not been indexed with this feature definition.

2. Right-click on one of the sugar spheres in the 3D view and select one of the Add sugar menu items. Note that four choices are available, as the reference file contains four sugar rings overlaid. Click on the last of the four options to create a sugar pharmacophore point.

3. Change the radius of each sugar pharmacophore base point to reduce the uncertainty in the position of the ring; this will ensure that the sugar ring in the search is very localised.

By default, the sphere radii of the pharmacophore points are set to 1.0 Å; change them to 0.4 Å by double-clicking on the tolerance radius of each pharmacophore sugar point in the Pharmacophore Features window.
The overlay model shows that having a properly aligned sugar ring, two aromatic moieties, and special distances between two pairs of pharmacophore points are important in order to achieve high activity. We can make the pharmacophore search more selective by including the phenyl rings in the pharmacophore query. For this workshop, we will include only the phenyl ring adjacent to the sugar ring into the pharmacophore query.

4. Tick the ring_planar_projected feature tick-box in the Pharmacophore Features window to display all the planar ring features present in the reference overlay in the 3D view. Note that there are two ring_planar_projected features in the 3D view adjacent to the glucose ring. Right-click on the ring_planar_projected green sphere pointed in the figure on the right and click on Add ring_planar_projected.

5. Although the aromatic moiety adjacent to the sugar ring is conserved in the SGLT2 inhibitors, its location is not certain. To make the pharmacophore query less restricted on the location of the aromatic ring, change the sphere radius of the ring_planar_projected pharmacophore point in the Pharmacophore Features window from 1.0 Å to 1.3 Å, by double-clicking on ‘B’ and use the spin-box under tolerance radius.

6. Finally, we want to find structures where all pharmacophore points belong to the same molecule. To add all intramolecular constraints, click the Intra button of the Edit: toolbar.

7. Deselect the PDB CrossMiner database, by unticking the pdb_crossminer tick-box in the Feature Databases window) before searching. This will speed up the search, as the hits of most relevance in this search are in the CSD.

8. You can now run the search by clicking the Play button in the Search: toolbar. Note that a different choice of sugar atoms and/or phenyl ring can results in different results of the ones discussed below.
Because the pharmacophore query includes non-indexed feature definitions, the pharmacophore search can require several minutes to complete.

**Inspecting the results**

1. To inspect the results, ensure the *Results Hitlist* window is shown. The **Display > Toolbars** from the top-level menu choose will show all displayed windows and toolbars. From here, hide the *Pharmacophore Features* and *Feature Databases* windows by unticking the tick-boxes. This will provide additional space for the *Results Hitlist* window.

2. Edit the **Number of hits**: in the *Results Hitlist* to 400 by double-clicking on the **Number of hits** spin-box. This will ensure that all matched hits are displayed.

3. When **1st in cluster** check-box is selected in the *Results Hitlist*, the matching hits are clustered based on the Tanimoto value showed in the **Tanimoto** spin-box. In this case, both the *Results Hitlist* and the 3D view show only the cluster representatives of those similar groups. In this workshop we are interested in similar molecular hits, so it is useful to uncheck the **1st in cluster** check-box in the *Results Hitlist* window.

4. Disable the **hydrogens** tick-box in the **Show** toolbar, this will hide the hydrogen atoms of the reference overlay and those of the matched hits.

5. Sort the hits by their rmsd by clicking on the **rmsd** column in the *Results Hitlist* window.

   Note that the results could be slightly different depending on the choice of the planar ring pharmacophore point and on the choice of the sugar ring pharmacophore point.

6. One of the first hits in the *Results Hitlist* window (hits with low rmsd) is VOLMUE (shown here in yellow) which includes a spiroketal moiety. Interestingly, the VOLMUE molecule is the CSD entry that inspired the use of a spiroketon chemistry in the original paper.
7. Navigating through the Results Hitlist browser you will find some other interesting hits:
   - WUKRIE (shown here in green) shows another example of a spiro chemistry.
   - QIXXAX (shown here in purple) shows a similar spiro motif but with a pyridine ring.
8. Later hits include BAGDIW which was also identified as a hit compound in the original paper.

Conclusion

This workshop shows a way of using a new substructure feature in a CSD-CrossMiner pharmacophore search to focus on closely related structures and inspect potential alternative scaffolds that might allow patent breaking or enhancement of 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.

Further exercises

- Experiment with more strongly defined sugar rings by using more elaborate SMARTS substructure definitions, or by adding exocyclic acceptor features to the pharmacophore.
- Explore more extensively the effect of the radii used in the pharmacophore query on the hits generated.
- Try the effect of changing the location of the planar ring projected pharmacophore.
- Try adding the additional hydrophobic planar ring in the overlay to see if you can find any SGLT2 inhibitors in the CSD.