

Searching and Subsets in ConQuest

Developed using
2020.2.0 CSD Release

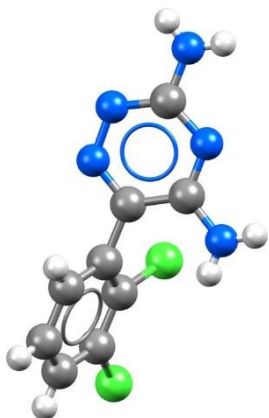


Table of Contents	
Introduction.....	2
Objectives	2
Example 1. Basic Searching in ConQuest.....	3
Example 2. Restricting searches to a subset	5
Conclusions.....	5
Further Exercises	6
Conclusions.....	6
Next steps.....	6
Bonus Exercise. Loading Subsets.....	7
Review. ConQuest Interface.....	9
Review. Draw Window	10
ConQuest sketching conventions.....	10

Introduction

ConQuest is the desktop search interface to the Cambridge Structural Database (CSD). All textual, numeric and structural data stored within the CSD can be searched using ConQuest. ConQuest provides an extensive range of flexible search options including searching based on compound name, formula, elemental composition, and literature search to name a few. With over a million entries in the database, individual structures can be missed. Subsets break the database down into more manageable subcategories and is an efficient way to search a certain category of compounds. As of the 2020.2 release, the latest additions to the subsets are the CSD Drug, pesticide, and the COVID-19 subsets.

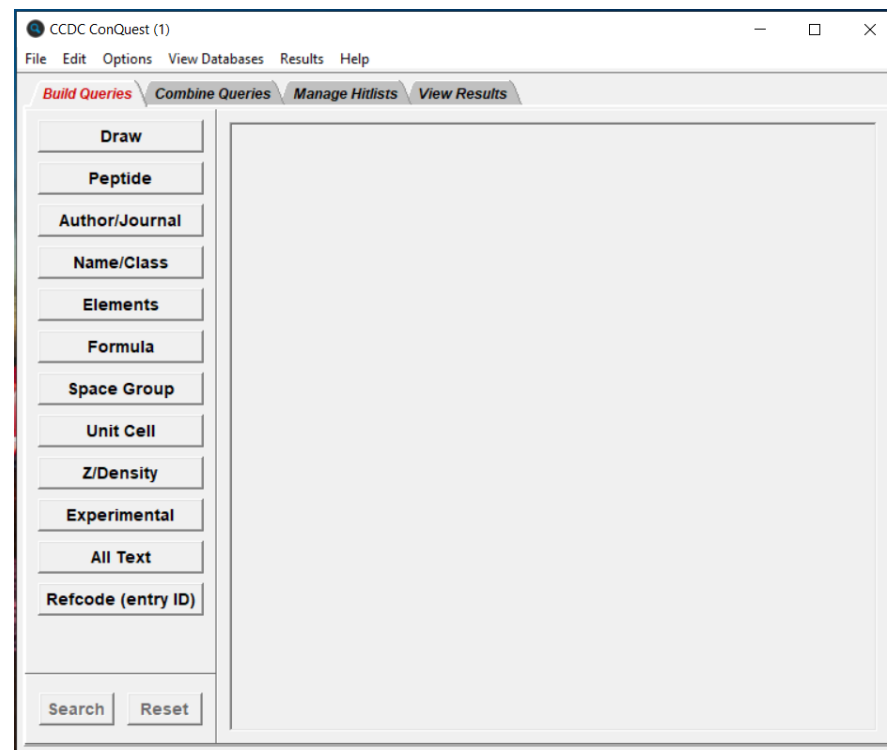
Before beginning this workshop, ensure that you have a registered copy of CSD-System or CSD-Enterprise installed on your computer. Please contact your site administrator or workshop host for further information

Objectives

This tutorial will guide you through sketching and using curated subsets in the CSD. At the end of this tutorial, you should be able to:

- Draw a 2D structure in ConQuest and conduct a basic search
- Restrict a search to a chosen subset
- Locate and load subsets
- Search structures at the intersection of two subsets

This workshop will take approximately **20** minutes to be completed.




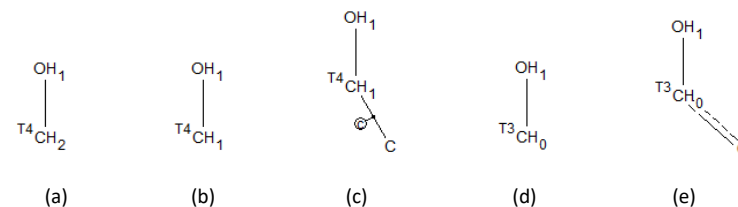
Example 1. Basic Searching in ConQuest

All drawing takes place in the central white area of the *Draw* window. In addition to creating 2D chemical structure sketches, the *Draw* window allows for the inclusion of 3D parameters for searching or for filtering.

For this workshop, we will sketch and search for fragments (a) and (b). If you complete the workshop early and are interested, you can draw the remaining three fragments and conduct a search.

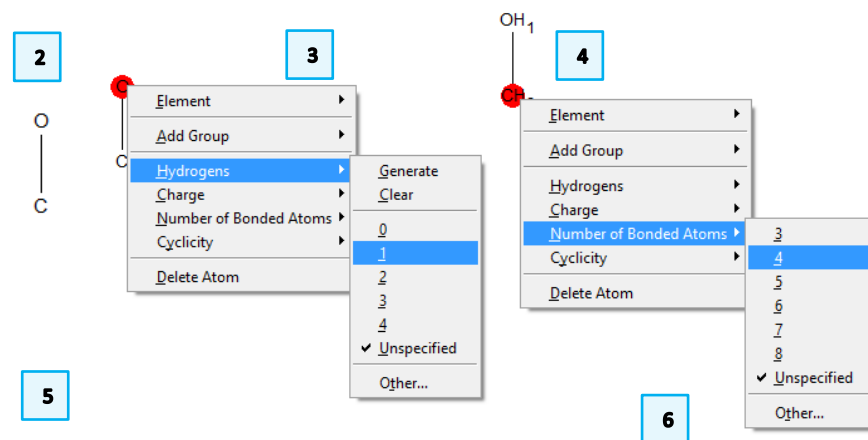
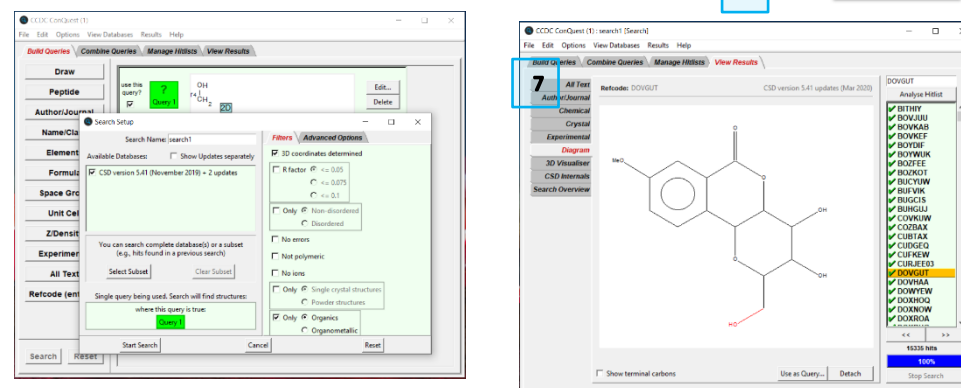
This example will demonstrate sketching techniques in ConQuest.

1. Launch ConQuest by clicking the ConQuest Icon  on your desktop or launching it from the Start or Applications menu and open the sketcher by clicking the **Draw** button.
2. Sketch a simple alcohol fragment by first adding a carbon atom and then adding an oxygen atom to it with a single bond.
3. Right click the O, choose *Hydrogens* and then 1 from the drop-down menu to add the H atom.
4. To ensure that the carbon atom is bonded to at least 2 hydrogen atoms and one other atom in addition to the OH group, first add two H atoms as in step 3 above. Then right-click the C again and choose *Number of Bonded Atoms*, and then 4 from the drop-down menu. This will produce fragment “a” from the list at the right.
5. Click the **Search** button, tick the boxes for *3D coordinates determined* and *Only Organics* in the **Search Setup** window. Then click the **Start Search** button.
6. You should get over 15,300 hits with this query. Use the arrow keys to scroll through the refcode list to view the results. *What do you find?*



These subtly different definitions would retrieve the following types of functional group:

- primary alcohols
- cyclic and acyclic secondary alcohols
- cyclic secondary alcohols only
- unsaturated cyclic and acyclic alcohols and carboxylic acids
- unsaturated cyclic alcohols only e.g. phenols

7 8

7. When you are done looking at your results, click the *Build Queries* tab to return to the Query window. You should see your first search fragment listed as Query 1 as in Step 5 to the right.
8. Click the **Edit...** button in the query window to return to the **Draw** window.
9. Right-click on the C atom and choose *Hydrogens*, and 1 from the drop-down menu as in step 3 above. Leave the *Number of Bonded Atoms* set to 4 to match a secondary alcohol (fragment “b” above).
10. Click **Search**. A pop-up warning will ask if you want to overwrite the existing Query. Click **No** to save this as a new Query.
11. Make sure the “3D coordinates determined” and “Only Organics” filters are ticked and start the search.
12. You should get over 27,900 hits with this new query. Use the arrow keys to scroll through the refcode list to view the results. Notice how the hits are different from the first search.

You should now be familiar with sketching substructures using the ConQuest *Draw* window; how to change bond and atom types; and how to view the data for each hit in your results list.

8

10

12

Example 2. Restricting searches to a subset

There are three ways to use subsets; you can load subsets in ConQuest to view the structures, restrict your search to a specific subset or through the CSD Python API. To begin with, we will focus on how to restrict the search to a subset. You can find how to load subsets in the bonus exercise. If instead you are interested in accessing subsets through the API, we recommend looking at the [API workshop](#) and [user guide](#).

For the purposes of this workshop, we will assume that you are interested in CSD entries with “amide” in the compound name in the CSD COVID-19 subset.

1. Click on the *Build Queries* tab and click on the **Name/Class** button to start a text-based search. Type the word *amide* in the “Compound Name” box. Click **Add** to include the word amide in the “Contains:” box. Click **Search** to begin the search.
2. In the *Search Setup* dialog box, we will select the subset we want to search from. Click on the **Select Subset** button to open the *Restrict Search* window. In this window, select *Entries in a pre-defined hitlist* to load the CSD defined subsets. From the drop-down menu, select **CSD COVID-19** subset. Once the subset is loaded, click **OK** to close the window
3. In the *Search Setup* window, make sure to tick the boxes for “3D coordinates determined” and “Only Organics” to apply these filters. Then click **Start Search** to begin the search.
4. There should be at least 11 structures returned. You can look through the results to assess where the amide groups are located and use such information to either refine or expand future searches.

Conclusions

Now you have seen a couple of ways to use subsets in your ConQuest searching. We have looked at three subsets in the CSD for this workshop, however, there are 11 subsets in total that you can explore. For example, the MOF subset can be

The screenshots show the following steps:

- Step 1:** The 'Build Queries' tab is active. The 'Name/Class' button is selected. The 'Compound Name' field contains 'amide'. The 'Contains' field is empty. The 'Add' button is highlighted.
- Step 2:** The 'Search Setup' dialog box is open. The 'Search Name' is 'search2'. The 'Available Databases' section shows 'CSD version 5.41 (November 2019) + 3 updates' selected. The 'Select Subset' button is highlighted.
- Step 3:** The 'Restrict Search' dialog box is open. The 'Current Restrictions' section shows 'Type: Refcode List', 'Title: unknown', and 'Percent: Unknown'. The 'Restrict based on pre-defined hitlists' option is selected. The 'CSD COVID-19 subset' is chosen from the drop-down menu. The 'OK' button is highlighted.
- Step 4:** The 'Search Setup' dialog box is open. The 'Search Name' is 'search2'. The 'Available Databases' section shows 'CSD version 5.41 (November 2019) + 3 updates' selected. The 'Select Subset' button is highlighted.

The final screenshot shows the search results in the 'Search Results' window. The 'Refcode: DIBBUL' is displayed. The 'CSD version 5.41 (November 2019)' is selected. The 'DIBBUL' hitlist is shown on the right, listing 11 hits. The chemical structure of the first hit is displayed in the center.

extremely useful if you want to search for MOFs but are unsure of how to draw a substructure that will find the right results.

Further Exercises

- Can you find structures that are common to the CSD Drug and MOF subset?
- How would you conduct a name search for other functional groups in the CSD COVID-19 subset?
- Can you export your results to Mercury and conduct data analysis on the structures?

Conclusions

This workshop introduced how to do basic searches and how to use subsets in ConQuest. You should now be familiar with:

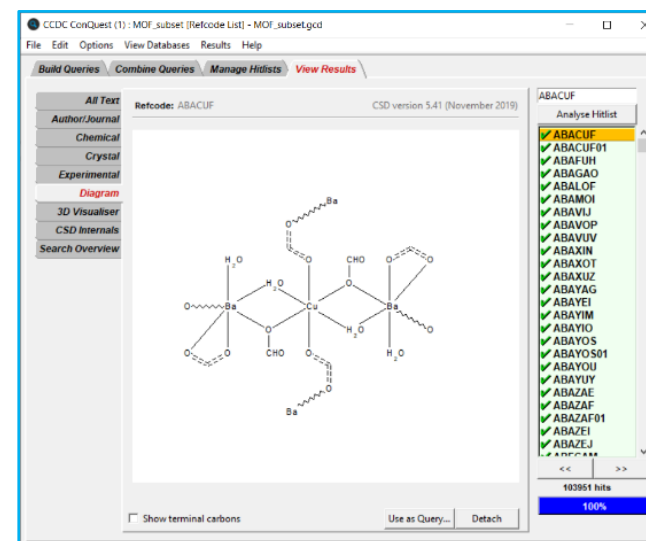
- Sketching substructures using the ConQuest *Draw* window
- Restricting searches to a specific subset

In the bonus exercise you will see in more detail:

- Loading pre-defined subsets
- Combining subsets

Next steps


The ConQuest user guide and other available ConQuest documents can be found [here](#).



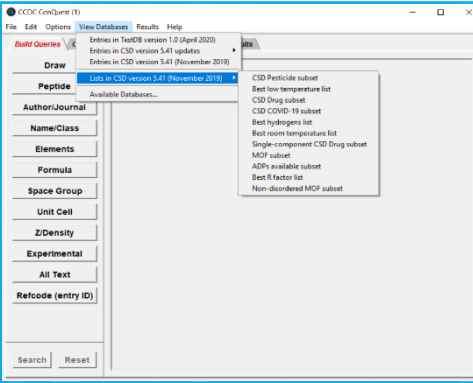
Bonus Exercise. Loading Subsets

There are three ways to use subsets; you can load subsets in ConQuest to view the structures, restrict your search to a specific subset or through the CSD Python API. For this bonus exercise, we will focus on the first method. If you are interested in accessing subsets through the API, we recommend looking at the [API workshop](#) and [user guide](#).

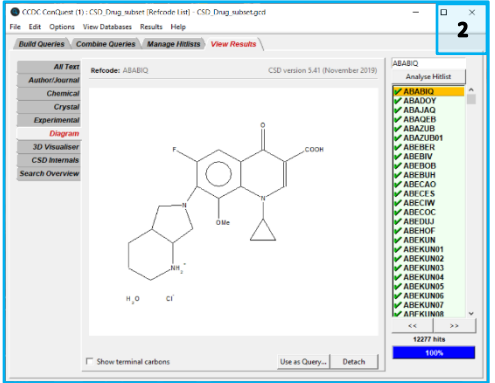
For the purposes of this exercise, we will assume that you are interested in CSD entries with a specific functional group in the approved Drug list by the DrugBank and pesticides subsets.

1. Launch ConQuest by clicking the ConQuest Icon  on your desktop or launching it from the Start or Applications menu. In the main ConQuest window, click on **View Databases** then click on *Lists in CSD version 5.41 (November 2019)*. A list of available subsets will be displayed.
2. You will first load the drug subset by clicking on *CSD Drug subset*. It will take a couple of seconds to load and there should be at least 12,277 structures in the subset. Click on **View Databases** again then click on *Lists in CSD version 5.41 (November 2019)* but load the *CSD Pesticide subset*. There should be at least 972 entries. With both subsets loaded, *how can you determine how many structures are in both subsets?*
3. You can answer this question by using the combine hitlist function from the Manage Hitlists tab. Click on the *Manage Hitlists* tab. We want to find out how many structures are in both subsets. Under the *Combine Hitlists* menu, List A should be CSD_Drug_subset. In the drop-down box under List B, choose CSD_Pesticide_subset. The radio next to “common to List A and List B” should be selected under *Generate a List of Entries*. Click **OK**, a new hitlist named “combination1” should be generated with 170 hits.
4. To view the structures in “combination1” in the main ConQuest window, Click the **View** button. We are interested in looking for a structure with an amide functional group that belongs in both the drug and pesticide subset.

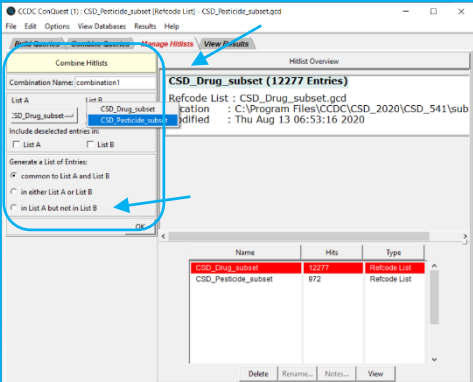
1



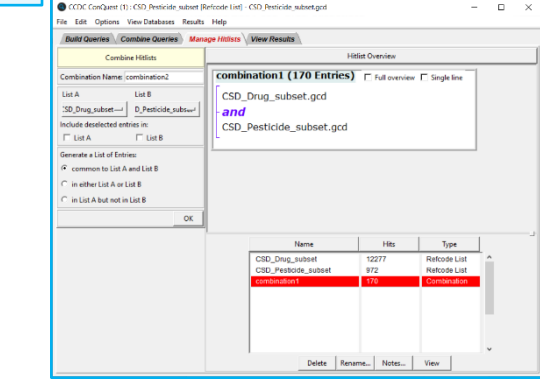
2



3



3



4

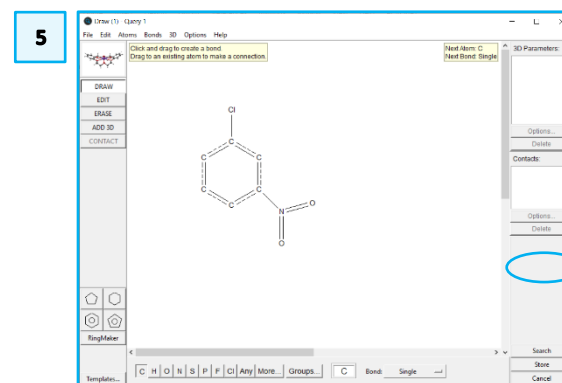
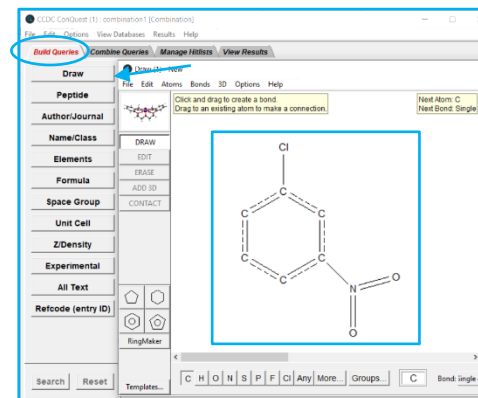
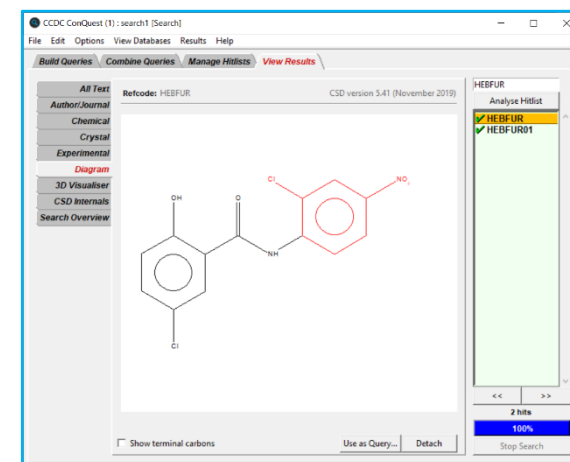
Name	Hits	Type
CSD_Drug_subset	12277	Refcode List
CSD_Pesticide_subset	972	Refcode List
combination1	170	Combination

Delete Rename... Notes... **View**


How can we conduct such search, limiting the search to the results from the combination in the previous step?

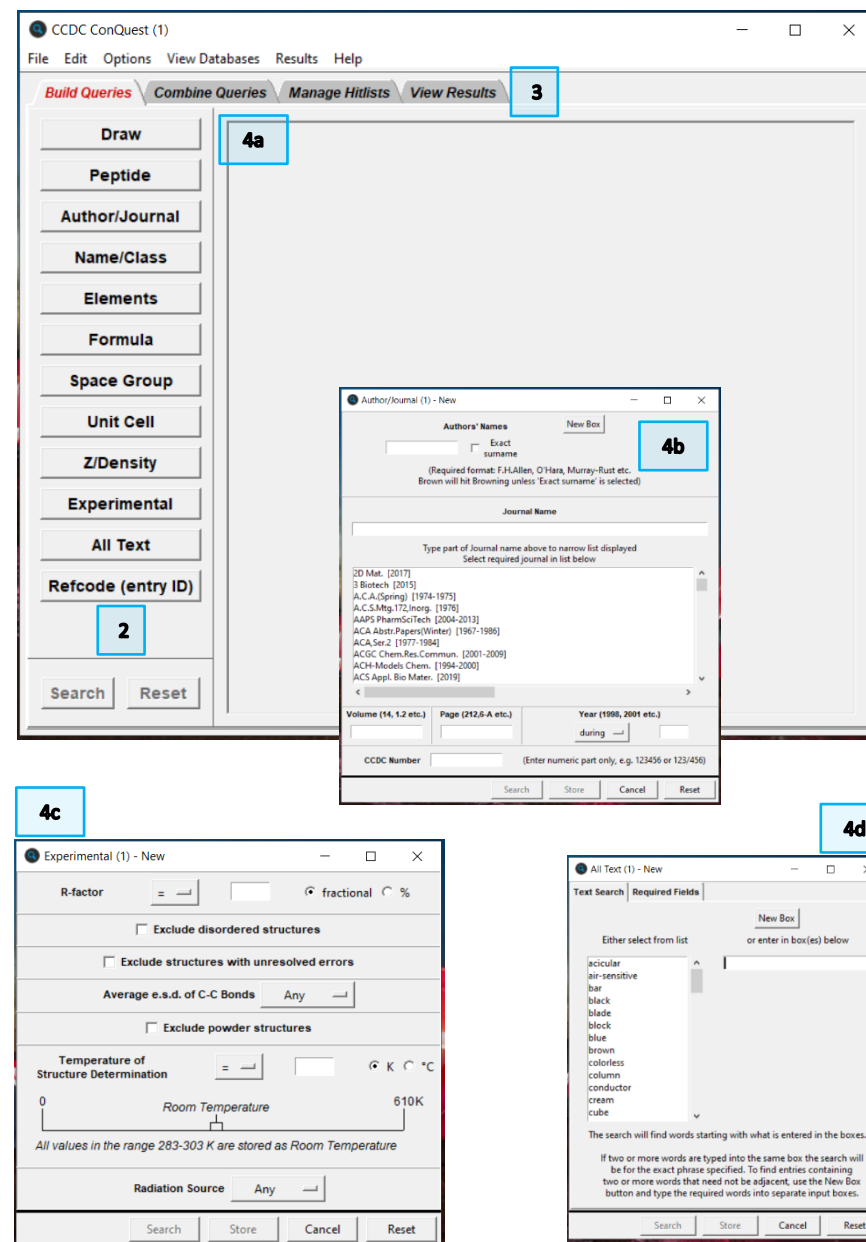
- In the main ConQuest window, click on the **Build Queries** tab, click on the **Draw** button, and sketch the motif in the figure to the right. If you need a refresher on the Draw window, I recommend revisiting the CQ-001 workshop. Once you are done sketching, click **Search** to proceed to the **Search Setup** window.
- The **Search Setup** window is where we can select the subset to be used for the search. Click the **Select Subset** button to open the **Restrict Search** window. The subset we want to load is the “combination1” hitlist that was previously generated. Since the hitlist was generated this session, select *Entries in a hitlist loaded this session*. There should be a drop-down menu to select which of the hitlists generated this session you want to use. Click on the menu and select **combination1**. Once the hitlist is loaded, click **OK** to close the window.
- Now in the **Search Setup** dialogue box, note that under **Available Databases**, there is an additional line below the CSD version that indicates the search is restricted to 170 refcodes. Make sure both boxes are ticked. Be sure to also tick the boxes for “3D coordinates determined” and “Only Organics” to apply these filters. Then click **Start Search** to begin the search.

Two structures should be returned, HEBFUR and HEBFUR01. Note that the two structures returned are in the same family.

Review. ConQuest Interface

1. Launch ConQuest by clicking the ConQuest Icon  on your desktop or launching it from the Start or Applications menu.
2. The ConQuest main window shows all the search routines you can perform on the left-hand side of the window.
3. The row of tabs across the top of the window will guide you through the steps of the search process.
4. Some example searches are
 - a. Draw – substructure and 3D information searching
 - b. Author/Journal – bibliographic searching
 - c. Experimental – experimental set up searching
 - d. All Text – generic text-based searching
5. The majority of the searching we will do in these tutorials will be substructure searching, so we will focus on the Draw tab here.



Review. Draw Window

All drawing takes place in the central white area of the *Draw* window. In addition to creating 2D chemical structure sketches, the *Draw* window allows for the inclusion of 3D parameters for searching or for filtering.

ConQuest sketching conventions

- Left click in the sketcher to insert the selected atom type
- Left click and drag to sketch two bonded atoms
- Use the **Edit** button to modify properties of or delete atoms, bonds or entire substructures
- Right-click on atoms or bonds to modify their properties
- Use the **Templates...** button to pick from a list of CSD editor devised and drawn substructures
- Use the **More...** button to find less frequently used element types, or generic atom type groups (e.g. halogens), or define custom element combinations (e.g. C or N or O).

