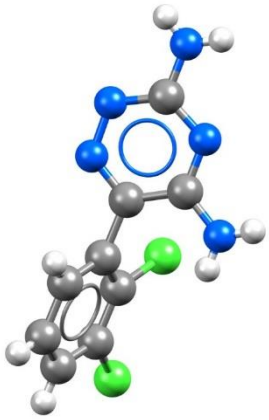
Introduction to ConQuest

Version 1.0 – April 2016



The Cambridge Crystallographic Data Centre

Table of Contents

Introduction

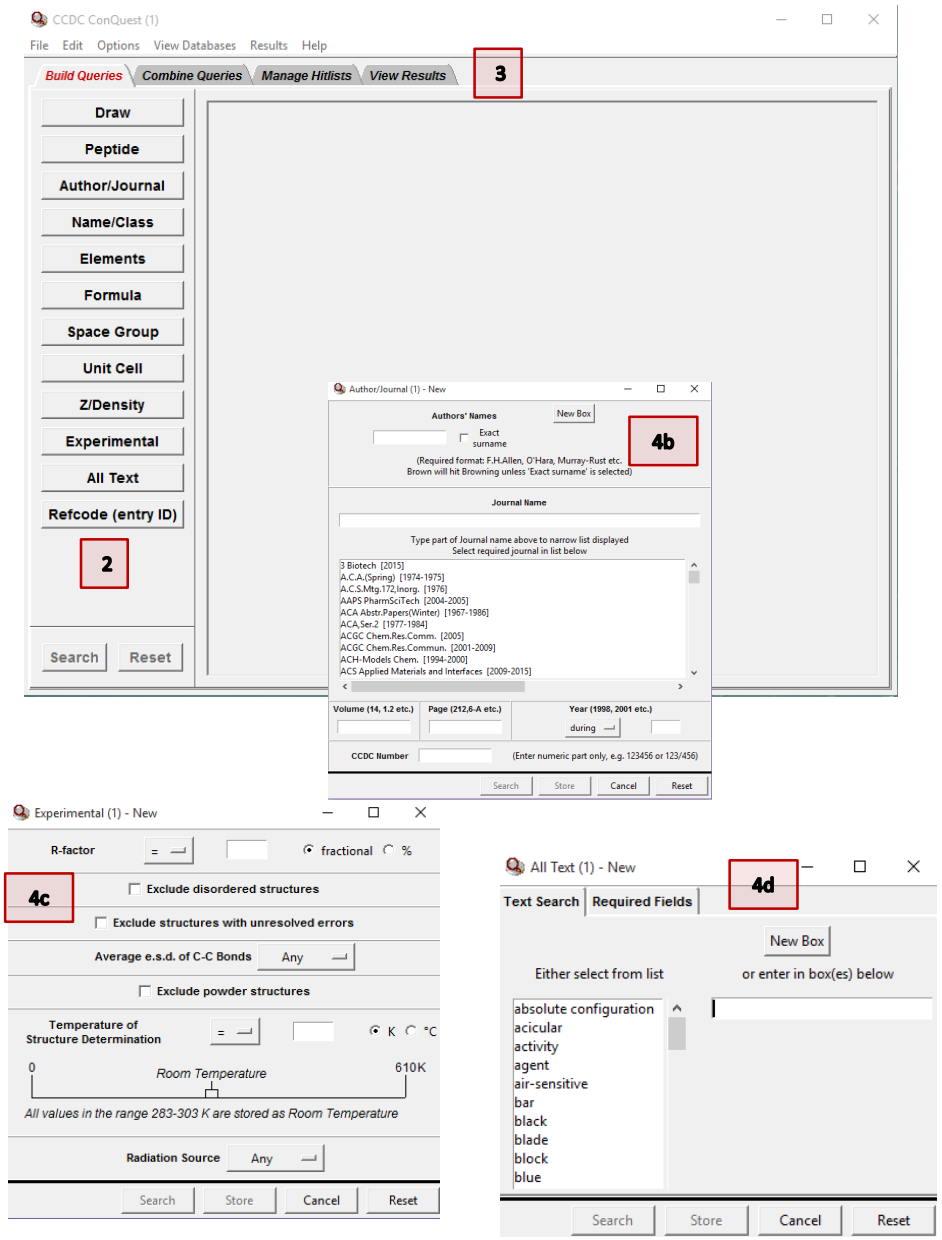
This set of tutorials will guide you through the use of ConQuest for substructure searching, for searching for and filtering by 3D data, and for viewing and analyzing your results. This tutorial was produced using CSD-Enterprise v5.37, ConQuest v1.18 and Mercury v3.8.

Before beginning this tutorial, 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.

ConQuest is the search interface to the Cambridge Structural Database (CSD). All textual, numeric and structural data stored within the CSD can be searched using ConQuest.

Overview of ConQuest

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.

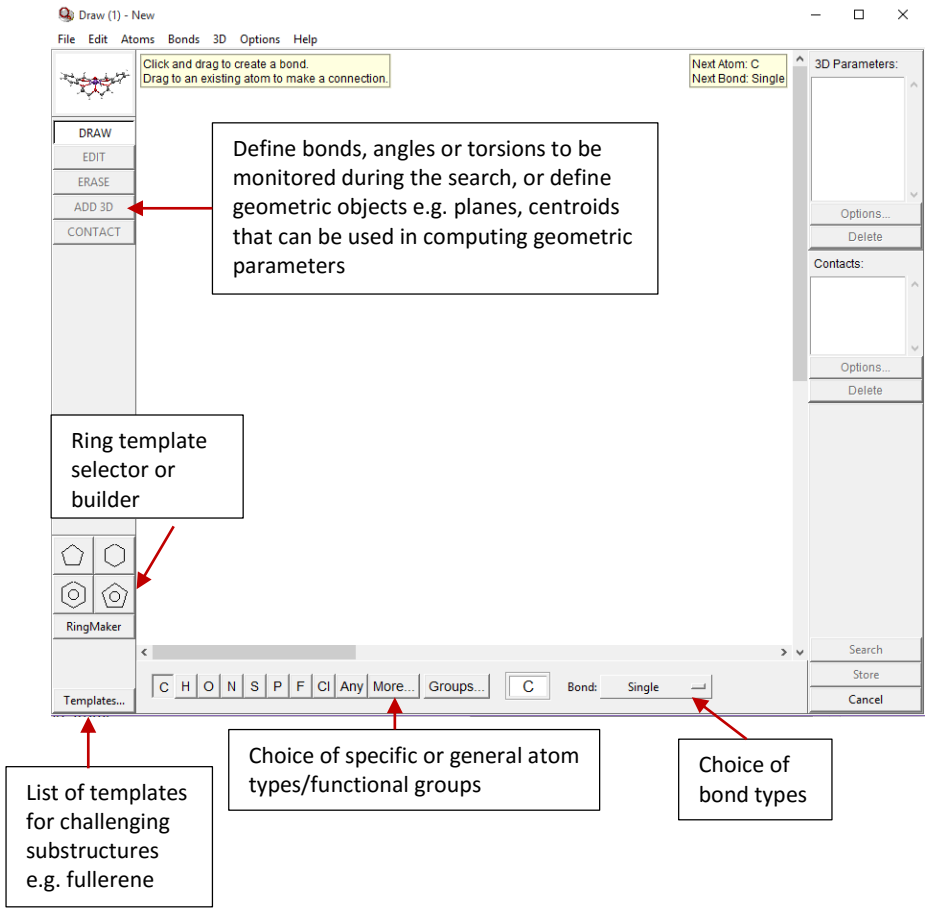


Introduction to the Draw Window

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 (eg. C or N or O).

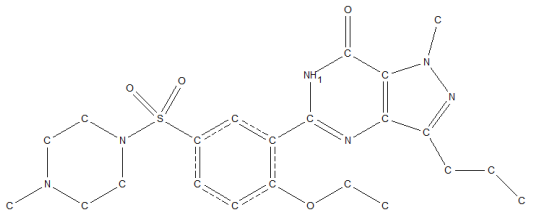


Example 1: Searching for Similar Molecules

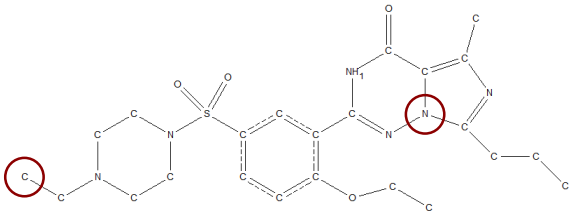
Sildenafil citrate or "Viagra" is a high progile drug. Since sildenafil's release in 1998, other competitor products have been released, e.g. vardenafil, that are structurally similar but have subtle chemical differences (see right).

Are there other entries in the CSD that contain similar fused ring substructures to those observed in both sildenafil and vardenafil?

This tutorial will take you through the steps needed to search the CSD for such similar compounds**. You will learn how to sketch and edit a fragment and how to view your results.**



Sildenafil (neutral form of active ingredient)



Vardenafil (neutral form of active ingredient)

1. Launch ConQuest and open the sketcher by clicking the **Draw** button.

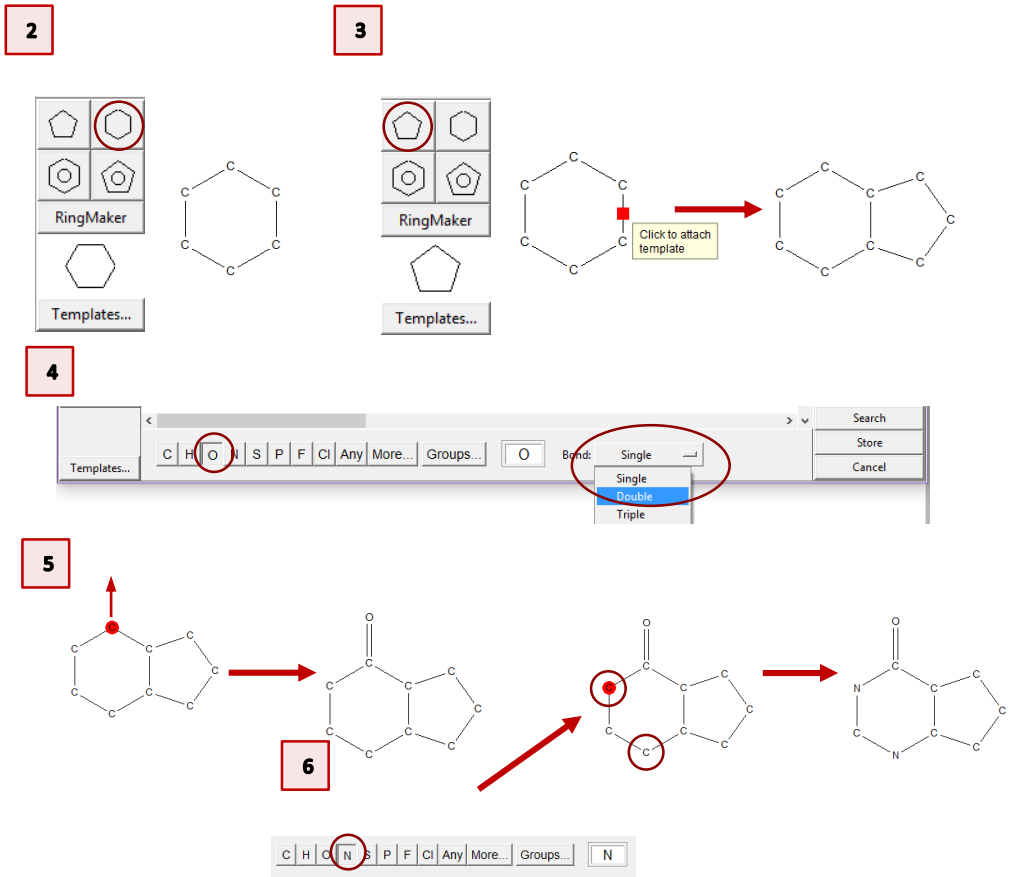
2. Start to sketch the sildenafil substructure shown on the right by first clicking on a six-membered ring template and clicking in the sketcher area.

3. Now click on a five-membered ring template, hover over the right hand C-C bond to attach the five-membered ring to the six-membered ring.

4. Add the carbonyl oxygen to the substructure by first selecting the O atom as shown below, and then changing *Single* to *Double* in the **Bond** pull down menu.

5. Click on the C atom to which the O atom should be bonded and then drag upwards while holding the left mouse button down.

6. Introduce two nitrogen atoms into the six membered ring at the positions shown by left-clicking the N button and then left-clicking on the atoms to be modified.



7. Add variable atom types (C or N) to the five-membered ring. Select *Other Elements* from the **More** pull-down menu, then activate the Multi Pick radio button and select **C** and **N** from the periodic table.

8. A new QA atom type has been produced. Select the atoms shown and modify these to be QA (i.e. either C or N) by left-clicking them.

9. Add a double bond to the six-membered ring shown by ensuring *Double* is selected from the **Bond** pull-down menu, then left-clicking on the bond.

10. Modify the bond types shown. Click on the **Bond** type pull down menu and select *Variable*. Tick the boxes for *Single* and *Double* as shown. Then click **OK**.

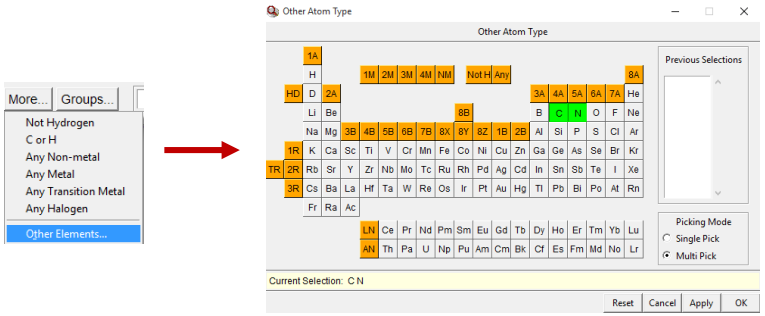
11. Now click on all the bonds in the five-membered ring to change them to the variable 1, 2 bond type.

12. Exclude any further ring fusion by right-clicking on the N atom adjacent to the carbonyl, clicking on *Hydrogens* from the drop down and selecting *1* from the list.

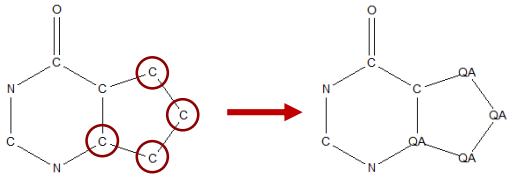
13. Now that you've finished sketching the fragment, start the search by clicking the **Search** button.

14. Tick the boxes for the "3D coordinates determined" and "Only Organics" filters in the *Search Setup* window, then click **Start Search**.

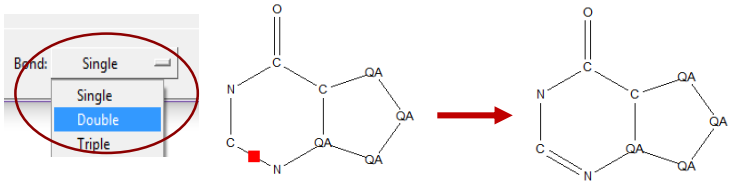
7



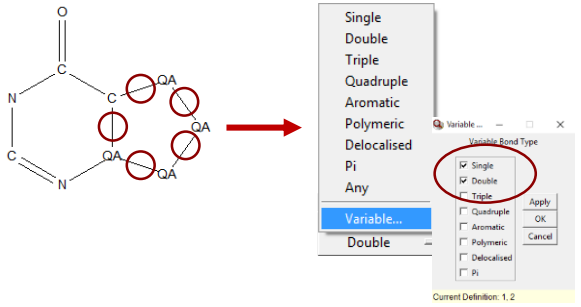
8



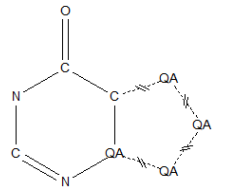
9



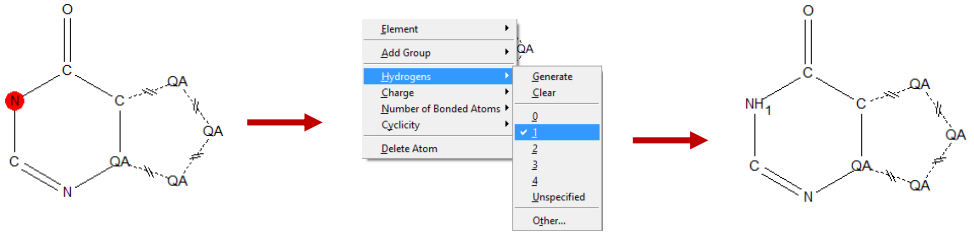
10



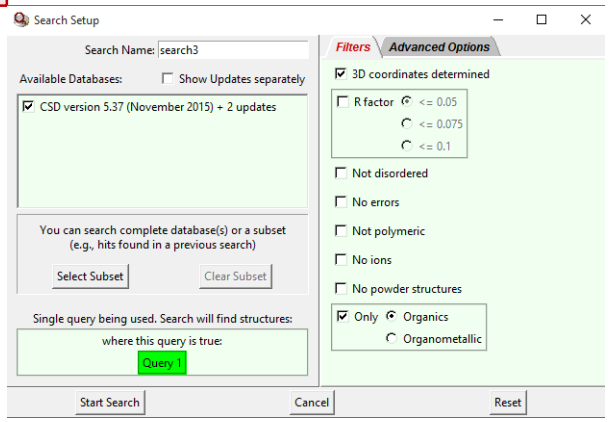
11



12



14



Viewing search results

15. All data stored with each CSD entry can be accessed from the tabs on the left, e.g. publication details.

16. The matched substructure is highlighted in red in the 2D diagram.

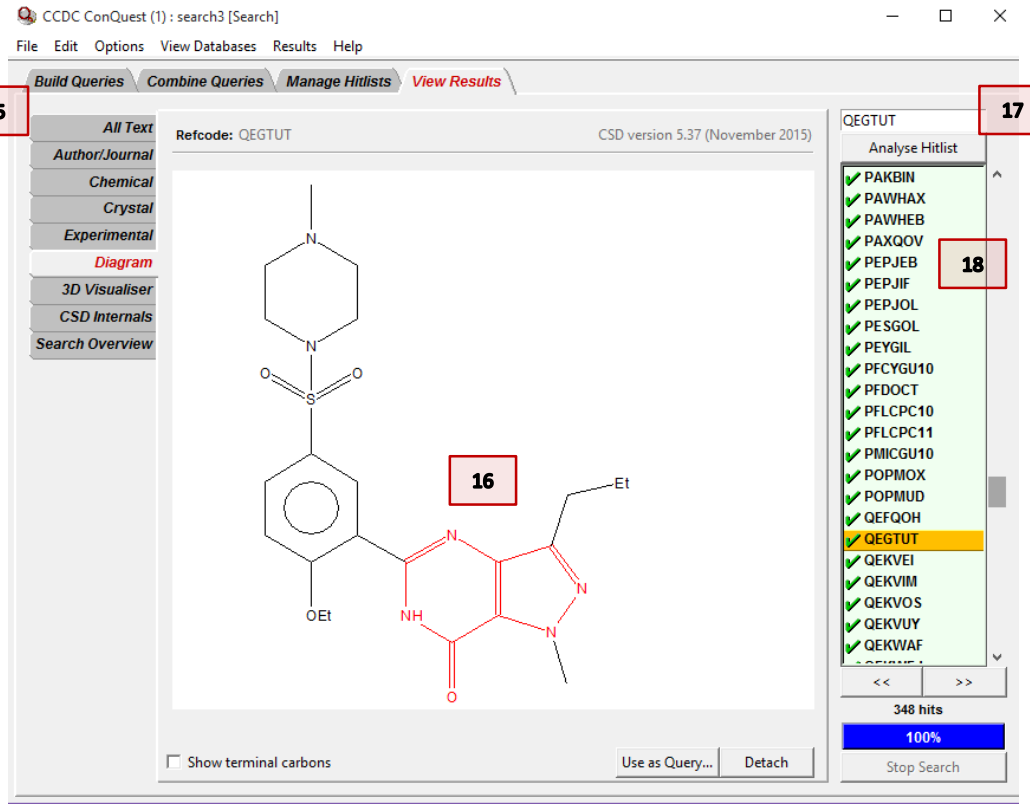
17. To inspect crystal packing, launch visualizer (Mercury) from the **Analyse Hitlist** button.

18. Hitlist of CSD Entries that contain the matched substructure are shown on the right side of the window.

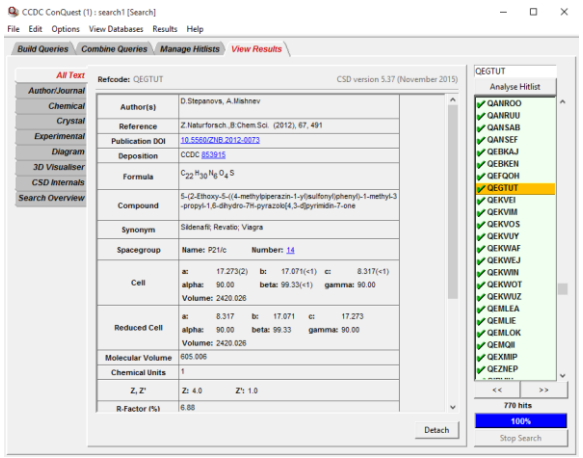
19. The *Alt Text* tab displays all textual information stored with the CSD entry.

20. Structures can also be viewed in 3D using the *3D Visualiser* tab.

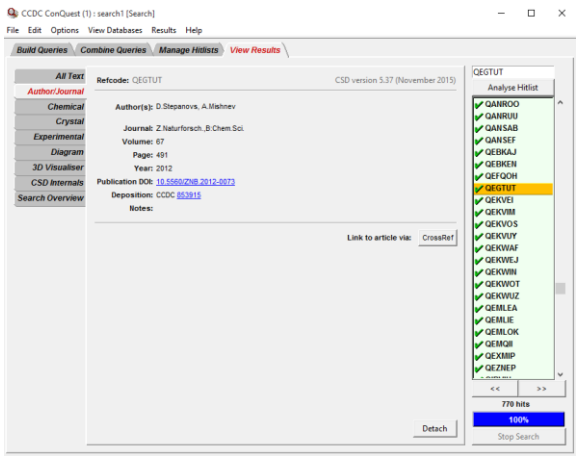
21. The *Author/Journal* tab shows publication details including links to the manuscript DOI where available.



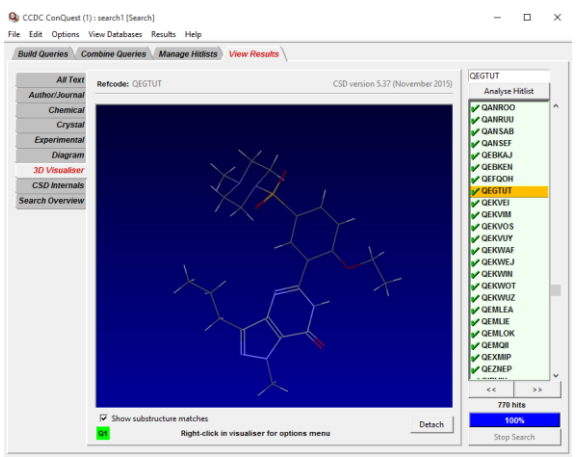
19



20



21



Conclusions

- There are 348 CSD entries (CSD version 5.37 + Nov. and Feb. 2016 updates) that contain similar substructures to the five and six membered fused ring core found in sildenafil and vardenafil.
- Sildenafil itself appears in the search results: CSD refcode QEGTUT
- Vardenafil is not in the CSD, although two closely related compounds, BITZOT and PEYGIL, are.
- Many other hits have drug-like activity or are natural products
- In the case of the five-membered ring, the bond and atom types in the substructure were left variable (single and double, and C, N, respectively). These could be further refined, for instance to match exactly the atom types and bonding found in sildenafil if additional specificity were required.

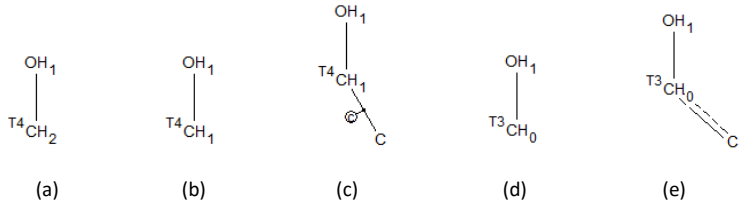
You should now be familiar with sketching substructures using the ConQuest *Draw* window; how to change bond and atom types; how to use ring templates; and how to view the data for each hit in your results list.

Further Exercises

- Try changing the filters to exclude disordered, ionic, polymeric and powder structures as well as those with errors. How do your results change?
- Change the 'QA' atom type to include O as well as C and N. How does this change your results?
- Can you design a similar search to focus on a different fragment of the sildenafil molecule? How many hits do you get?
- Explore the **Templates…** button in the **Draw** window to see what other useful templates are available.
- Try designing a substructure search for your own compounds of interest.
- Advanced: Can you use the *Manage Hitlists* tab to combine hitlists to find common molecules between the above search results?

Example 2: More Searching in ConQuest

We can demonstrate the effect small changes on our search fragment have on the results of our searches by looking at the series of alcohols shown on the right.



This example will demonstrate additional sketching techniques, combining different queries with Boolean operators, and exporting your results.

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

(a) primary alcohols
(b) cyclic and acylic secondary alcohols
(c) cyclic secondary alcohols only
(d) unsaturated cyclic and acyclic alcohols and carboxylic acids
(e) unsaturated cyclic alcohols only e.g. phenols

1. Launch ConQuest 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.

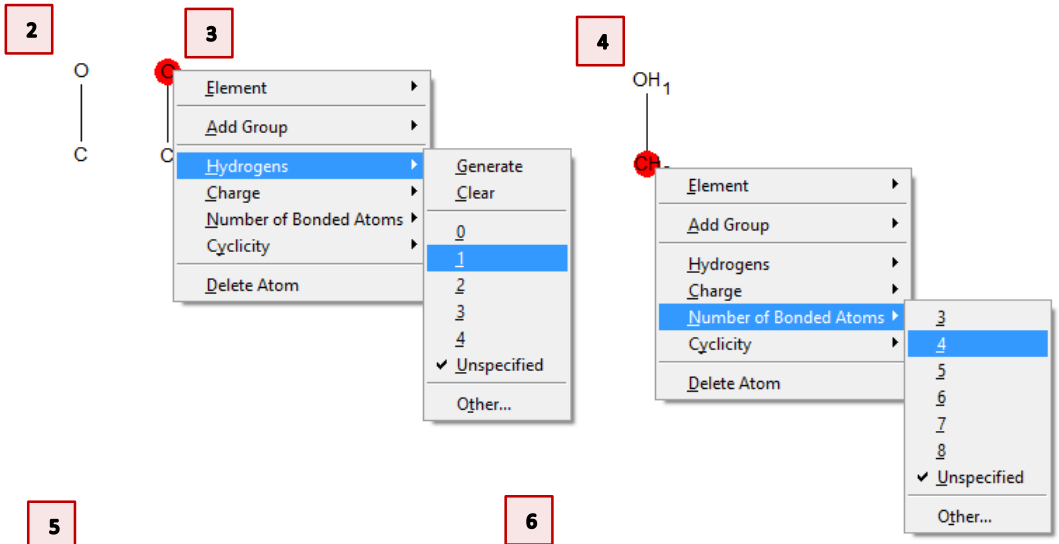
3. Right click the O, choose *Hydrogens* and then *1* from the drop-down menu to add the H atom.

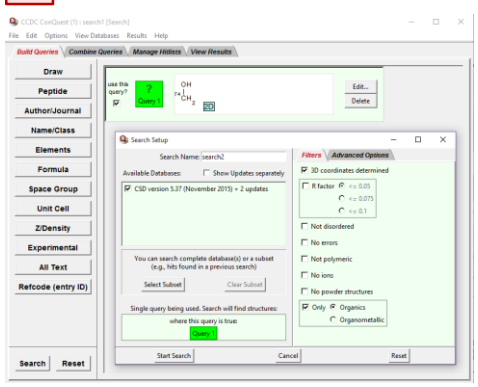
4. To ensure that the carbon atom is bonded to at least 2 hydrogen atoms and one other atom in addition to the OH group, first add the 2 H atoms as in step 3 above. Then Right-click the C again and choose *Number of Bonded Atoms*, and then *4* from the drop-down menu. This will produce fragment "A" from the list at the right.

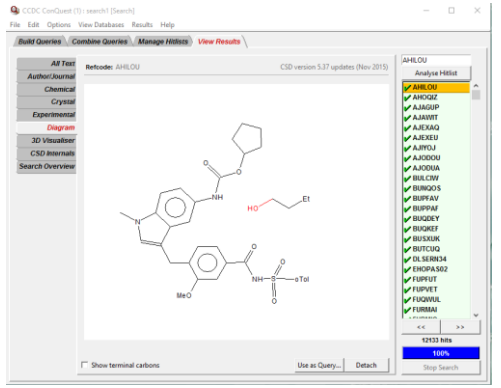
5. Click the **Search** button, tick the boxes for *3D coordinates determined* and *Only Organics* in the **Search Setup** window. Then click the **Start Search** button.

6. You should get over 12,000 hits with this query. Use the arrow keys to scroll through the refcode list to view the results. What do you find?

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 22,000 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.

13. Repeat steps 7 and 8 above to return to the **Draw** window with the fragment from Query 2 showing.

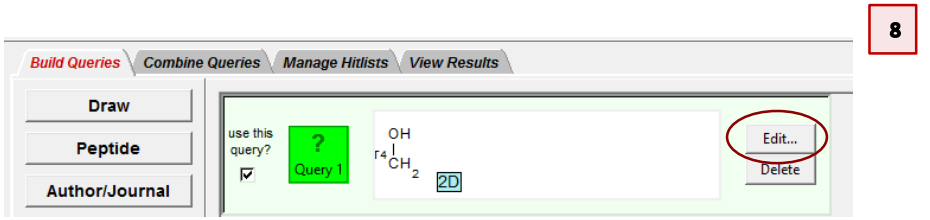
14. To draw a cyclic secondary alcohol, first draw another C bonded to the first one.

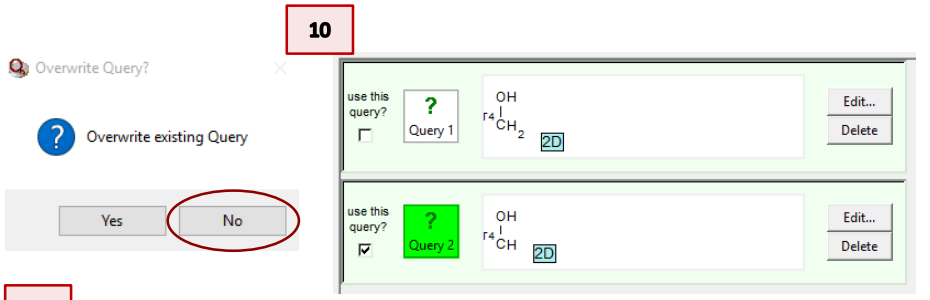
15. Now, right-click on the C-C bond, choose *Cyclicity* and *Cyclic* from the drop-down menu.

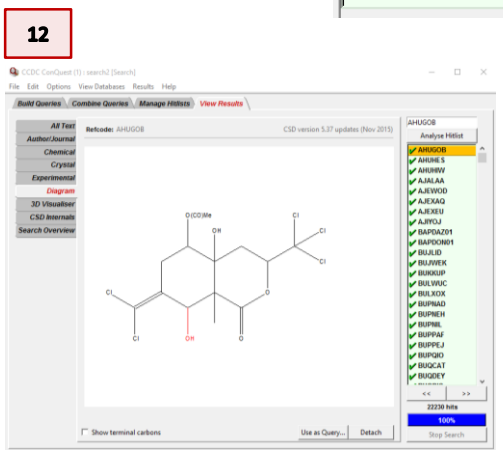
16. This will mark the bond as *cyclic* or belonging to a ring system.

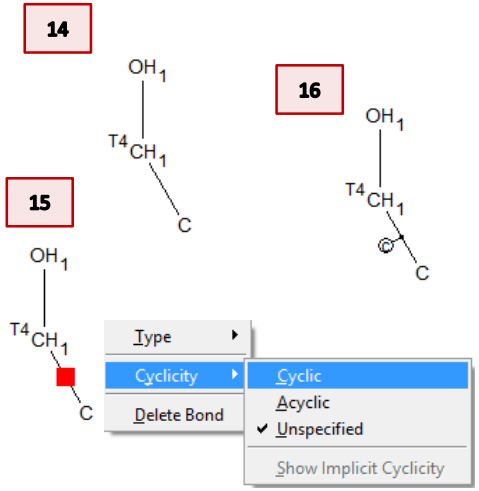
17. Follow steps 10 and 11 above to save this as Query 3 and start the search.

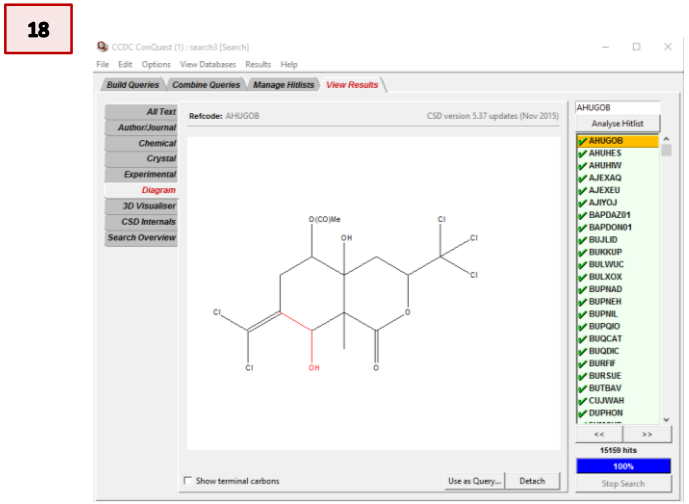
18. This time you will get over 15,000 hits. Again, scroll through the refcode list to explore your results. Notice that refcose AHUGOB appears in both searches.











19. Again repeat steps 7 and 8 above to return to the **Draw** window showing the fragment form Query 2.

20. Right-click the C atom and choose *Hydrogens* and *0* from the drop-down menu. This will give you a fragment that matches tertiary alcohols (fragment "C" above).

21. Click **Search** and then **No** to write this out as a new Query.

22. Start the search again with "3D coordinates determined" and "Only Organics" filters set.

23. You should have over 14,000 hits for this query. Again, scroll through the refcode list to explore your results.

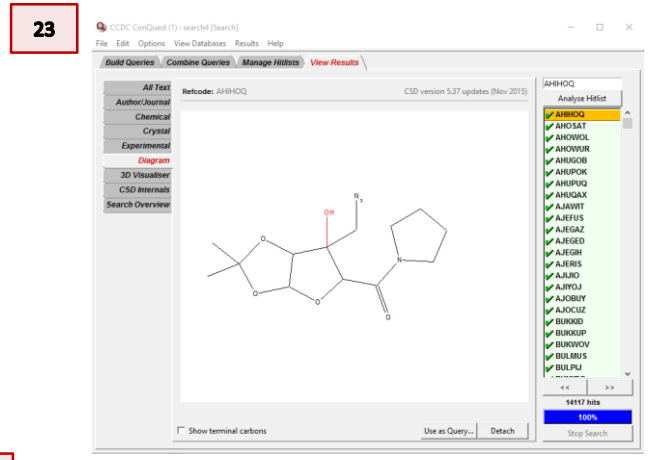
24. Finally, we want to search for a phenol alcohol (fragment "D"). Again, follow steps 7 and 8 above to edit Query 3.

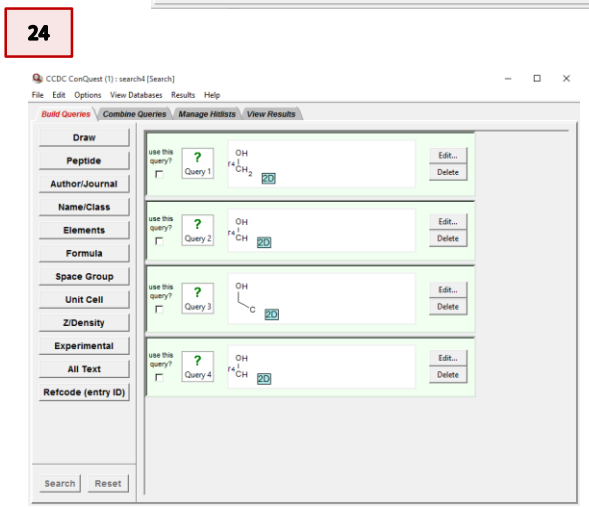
25. In the **Draw** window, right-click the single bond between the two carbons. Select *Cyclicity* and then *Unspecified* from the drop-down menu. Right-click the single bond again and choose *Type* and then *Aromatic* from the drop-down menu.

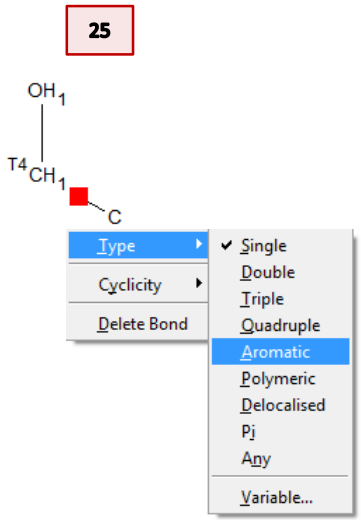
26. Right-click the central C atom, choose *Hydrogens* and then *0* from the drop down menu. Then right-click the central C again, choose *Number of Bonded Atoms* and the *3* from the drop-down. This will set up the proper bonding for a phenol group

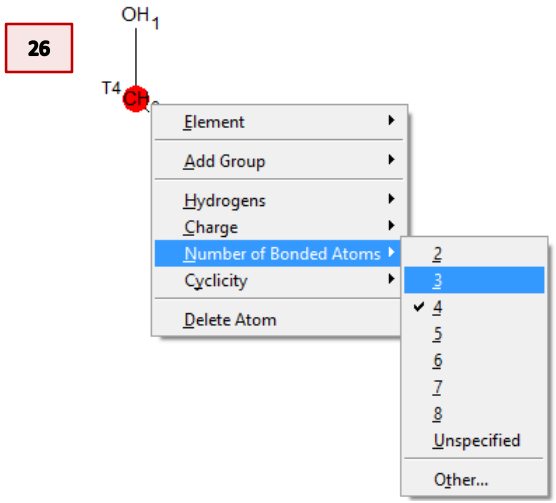
27. Click **Search** and then **No** to save this fragment as a new Query. Make sure the "3D coordinates determined" and "Only Organics" filters are set and start the search.

28. You should now have over 21,000 hits for this Query. Scroll through the refcode list to explore your results.









Combining Queries

1. Now you should have five separate Queries representing the five different alcohol fragments above. You can check this by clicking on the *Combine Queries* tab of the window.

2. Let's look at finding compounds with both a cyclic secondary alcohol (Query 3) AND a phenol (Query 5). To do this, click the box that says "Query 3" and drag it to the top box labeled "must have (boolean AND)".

3. Do the same this with the box that says "Query 5" so that you have both Query 3 and Query 5 in the top "must have" box.

4. Click **Search** and make sure you have the "3D coordinates" and "Only Organics" filters ticked. You should get 750 hits. Scroll through the refcode list to explore your results.

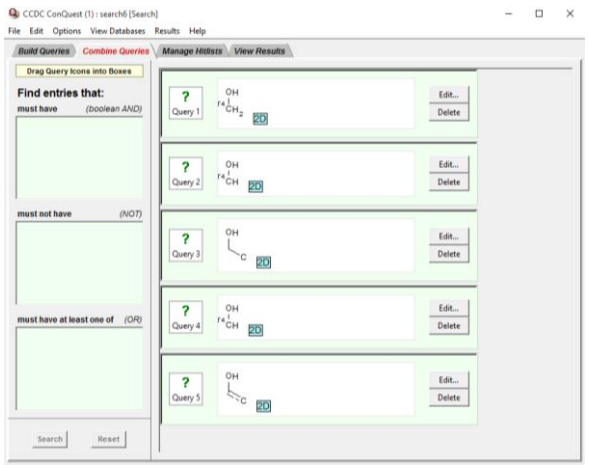
5. Query 2 returns all secondary alcohols while Query 3 returns only cyclic secondary alcohols. You can use the Boolean NOT operator to return only acyclic secondary alcohols.

6. Return to the *Combine Queries* tab. First, click and drag the Query 3 and Query 5 boxes back to the right side of the window. Then drag the Query 2 box to the "must have" box – we must have a secondary alcohol. Next drag the Query 3 box to the "must not have" box – we don't want any cyclic secondary alcohols.

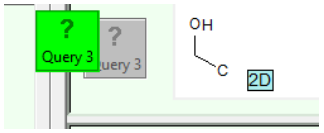
7. Start the search as in step 4 above. You should obtain 7071 hits. Scroll through your results to see that none of them contain cyclic secondary alcohols.

8. Finally, to find all compounds containing an alcohol group in this set, drag each Query box to the bottom "must have at least one of (OR)" box. Start the search as usual. It will return over 59,000 hits!

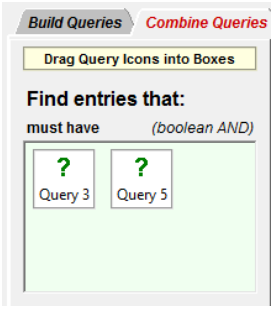
1



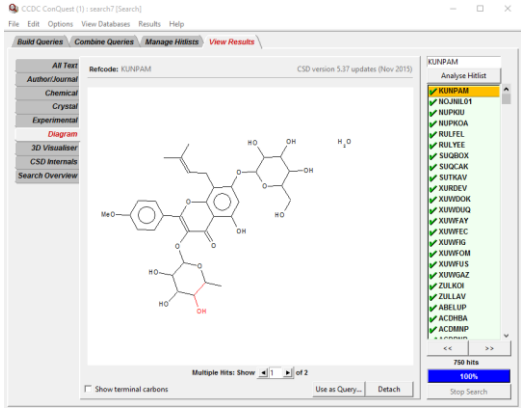
2



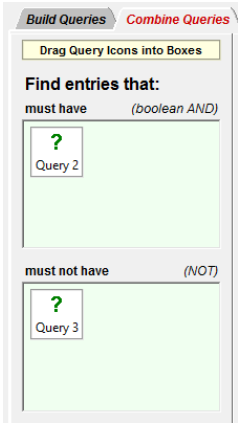
3



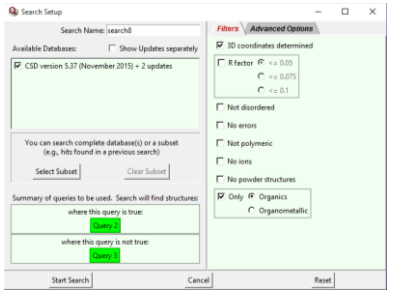
4



6



7



Combining Searches

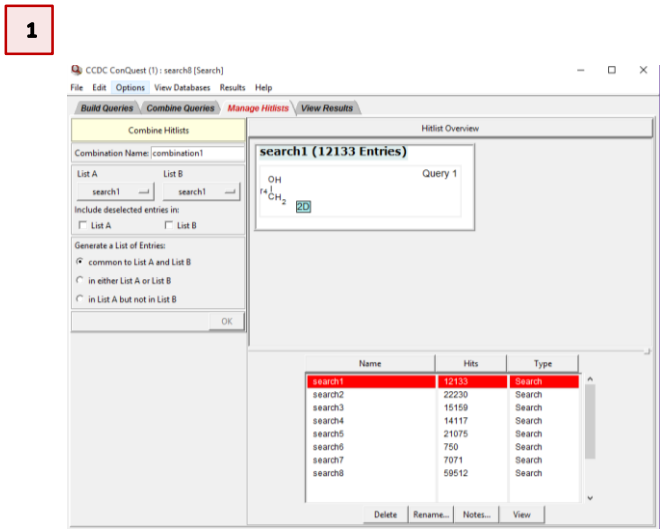
1. You can also combine search results using the *Manage Hitlist* tab. You will see a list of all the searches you performed. If it helps to keep them straight, you can select a search, click the **Rename…** button and enter a new name for the search.

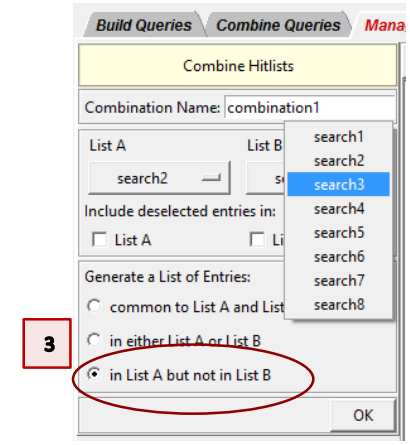
2. If we wanted to generate the list of refcodes for the set of acyclic secondary alcohols, we could simply subtract the hits from search 3 (cyclic secondary alcohols) from the list of hits in search 2. To do this, in the drop-down box under List A, choose *search2*. The in the box under List B choose *search3*.

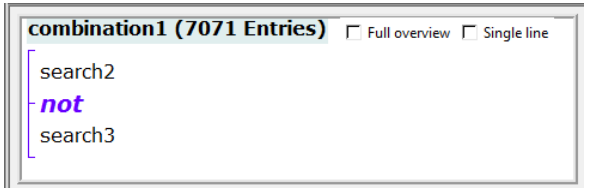
3. Now we need to select the combination. For this we want the hits in search 2 (all secondary alcohols) but not the hits from search 3 (only cyclic alcohols). Tick the box for "in List A but not in List B"

4. Click "OK" and you will see a new item "combination 1" which shows how many hits are returned. You can see immediately here that this combination returns the same number of hits (7071) as our previous Boolean combination of queries. Click the **View** button at the bottom of the window to view the highlighted combination or any of the searches.

5. You can also use these "combination" sets within new combinations to create very refined hitlists.







Exporting Results

1. In the *View Results* tab, you can choose which entries to save by toggling between the green check (selected) and the red cross (deselected).

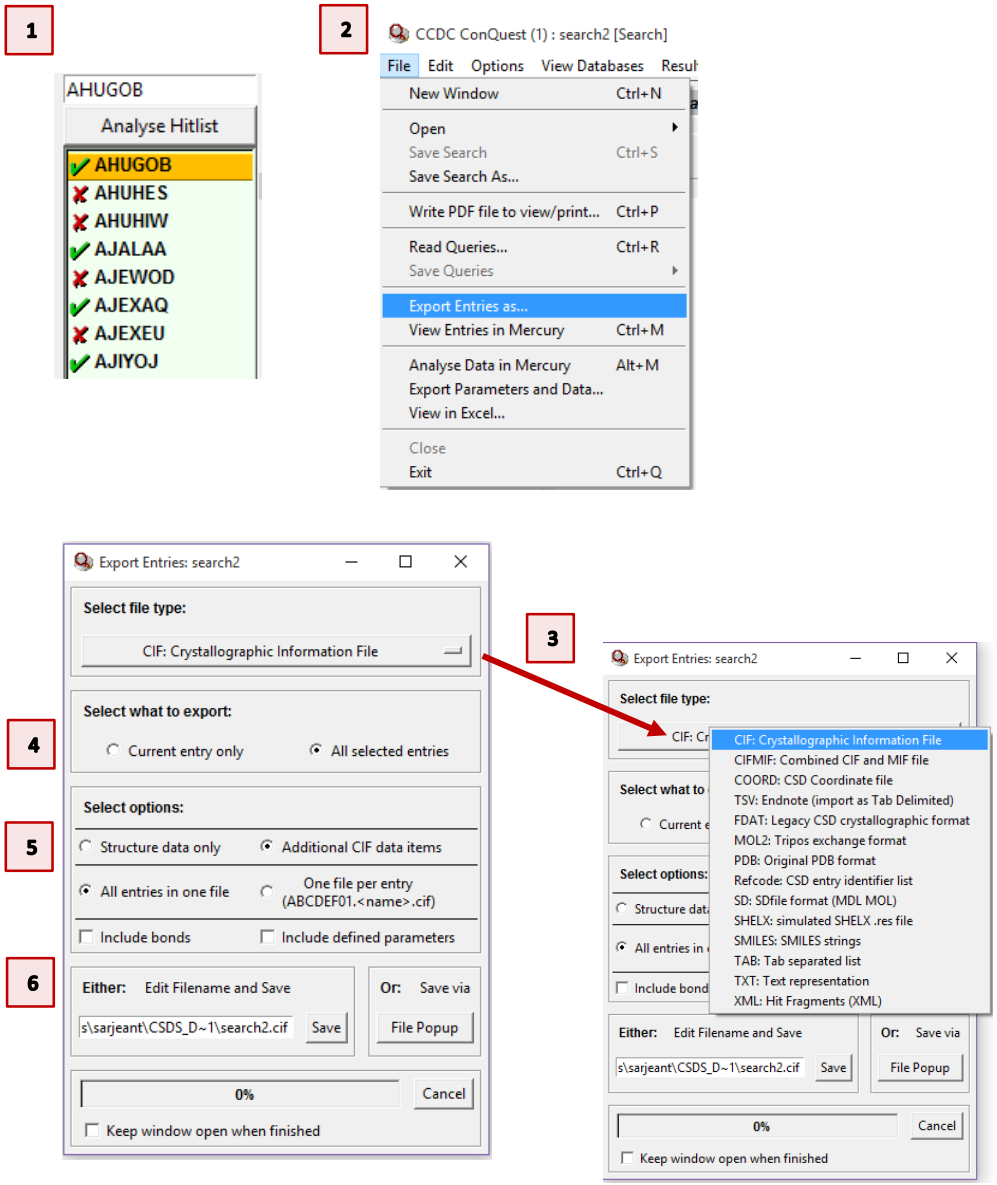
2. To export one or all of your hits in various file formats, from the *View Results* tab, Choose "File" from the top menu, then *Export Entries as…*

3. This will bring up a window where you can choose your desired file format. Choose the appropriate format for your work.

4. Use the dialogue to choose to export the Current entry only or All selected entries.

5. Different file types ask for different options, select these if appropriate.

6. Either edit the file name and save, or choose Save via **File Popup** to choose the location manually.



Conclusions

In this tutorial, you have seen how to make subtle changes to a search fragment and what effect those changes can have on the hitlists returned. You have also learned how to combine queries with Boolean operators, and how to create combinations of different searches using the hitlist manager. Finally, you learned how to export your results as different file types for use in other applications.

Further Exercise

- Try creating a search fragment that would return all the acyclic secondary alcohols without using combinations.

Example 3: Intermolecular Interactions with ConQuest

You can use the entries in the CSD to look for structures that exhibit particular interactions. By setting up a search to look for intra- or intermolecular interactions, you can then analyze the data you obtain from the search *via* ConQuest and Mercury.

For the purposes of this tutorial, we will assume that you are studying the molecule acetyl-methionyl-glycine and are curious about interactions between amine and carboxylic acid groups.

This tutorial will show you how to set up a search for intermolecular interactions and how to analyze the data you obtain on these interactions.

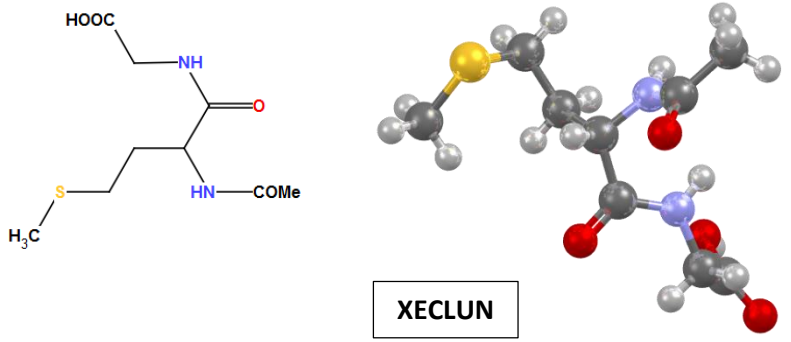
1. Start out by drawing a generic carboxylic acid and amide group using the ConQuest **Draw** window. Make sure to put the amide N-H near the carboxylic acid carbonyl.

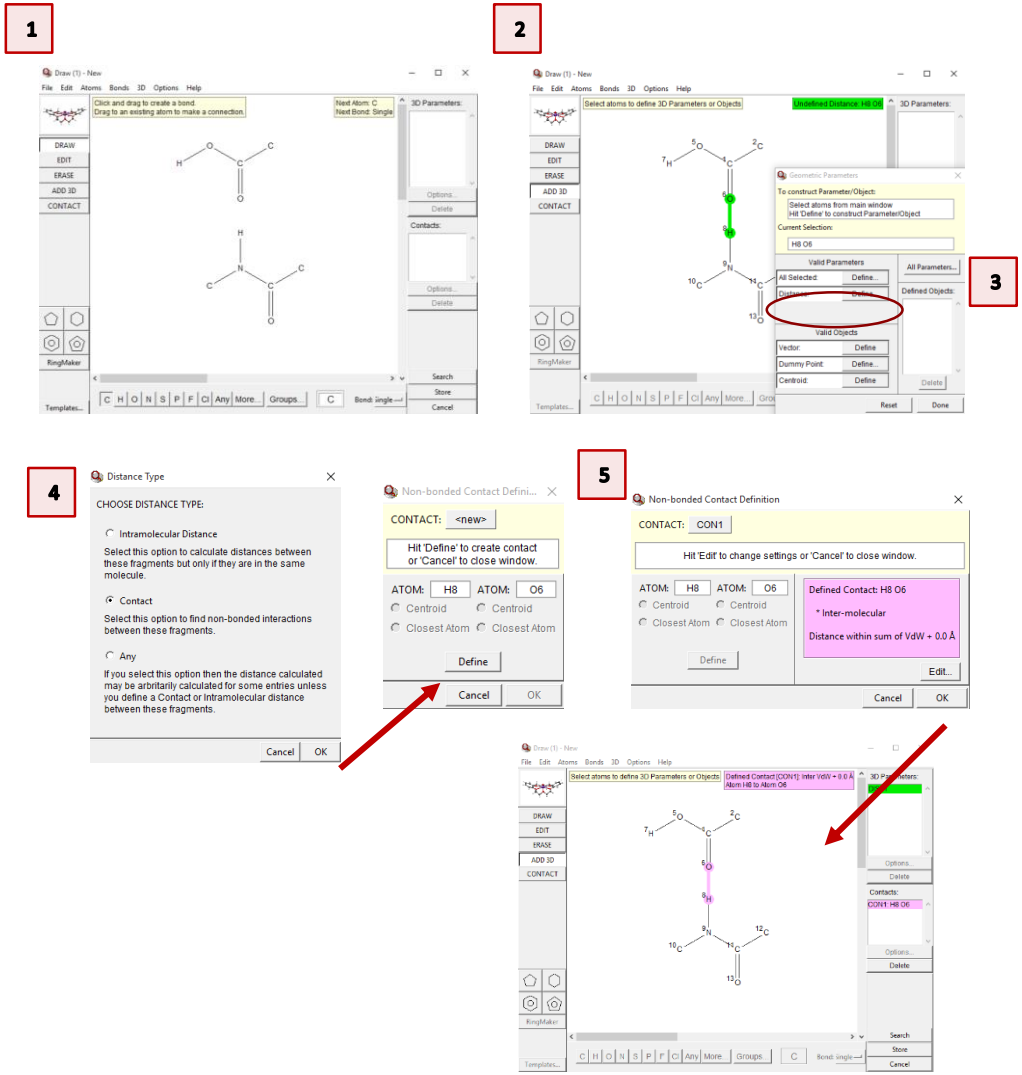
2. To include geometric parameters in your query, click the **ADD 3D** button. This will open a dialogue box. To add an interaction, click the amide H and carboxylic acid carbonyl O. Selected atoms and the distance between them will be highlighted in green.

3. Under the Valid Parameters heading in the dialogue box, click **Define** next to "Distance:"

4. You will then need to choose the type of distance interaction. In the *Distance Type* dialogue box, choose "Contact" for non-bonded interactions. Click **OK**. Then, click **Define** in the small *Non-bonded Contact Definition* box.

5. Another *Non-bonded Contact Definition* dialogue box appears. You can use this to edit the parameters of the search, but we will skip this step for now. Click **OK** to define the interaction. You will see the contact highlighted in pink in the main **Draw** window.





6. Now we want to define the N-H---O angle. In the **Draw** window click **ADD 3D**. Then click, in order, the N, H, and O atoms to define the angle around the H atom.

7. Click the **Define** button next to "Angle:" in the *Valid Parameters* dialogue box.

8. We don't need to change any of the parameters of this angle, so we can simply click **Done** in the *Geometric Parameters* dialogue box.

9. Back in the **Draw** window, you should see both the Distance and Angle queries listed.

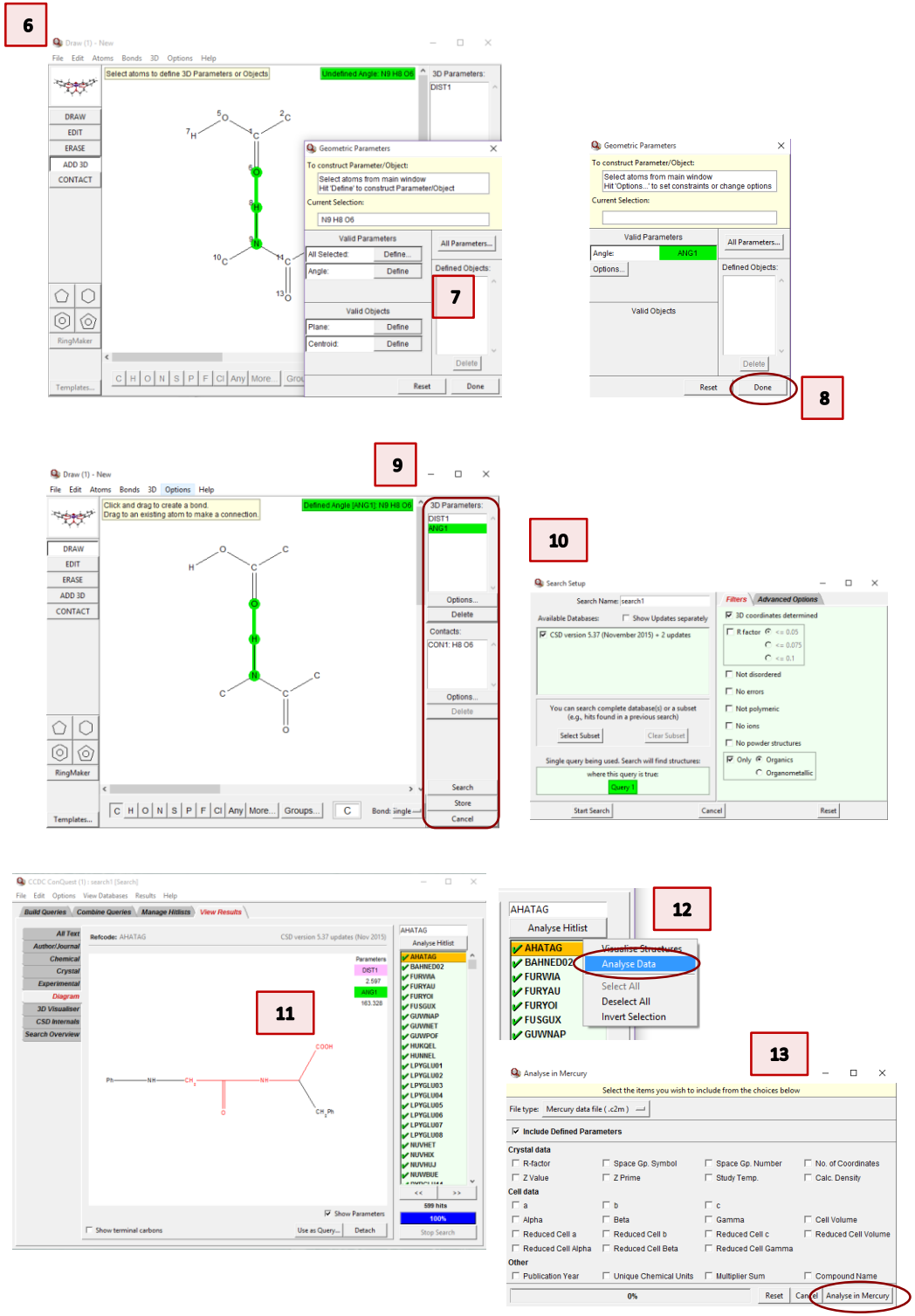
10. Click **Search** to start the search. Be sure that you tick the boxes for 3D coordinates determined and Only Organics to apply these filters. Then click **Start Search** to start the search.

11. The search should return 599 hits. Note that the values for the distance and angle parameters you defined are shown in the upper right corner of the *View Results* tab.

12. You can use the arrow keys to scroll through the refcodes to investigate the hits returned by this query. However, we can use Mercury's **Data Analysis** module to further analyze the data. To launch this, click the **Analyse Hitlist** button and choose *Analyse Data* from the drop-down menu.

13. If you are interested in analyzing other parameters, you can tick them off in the dialogue box that appears. We will skip this step for now. Simply click **Analyse in Mercury**.

14. This will launch the Mercury app with the refcode list loaded and the **Data Analysis** window. The **Data Analysis** window is often hidden behind the main Mercury app, so you may have to minimize Mercury to see it.



15. Look at the **Data Analysis** window. This shows all the data from your geometric parameters, sorted by refcode.

16. To plot the data for analysis, click the headers DIST1 and ANG1 to select both columns.

17. Then click Plots and then *Scatterplot* from the drop-down menu.

18. This will produce a scatterplot in the bottom half of the **Data Analysis** window. The datapoints in this plot are hyperlinked to the Mercury app. You can click the top bar of the plot window and drag it out of the **Data Analysis** window to make it easier to view the data.

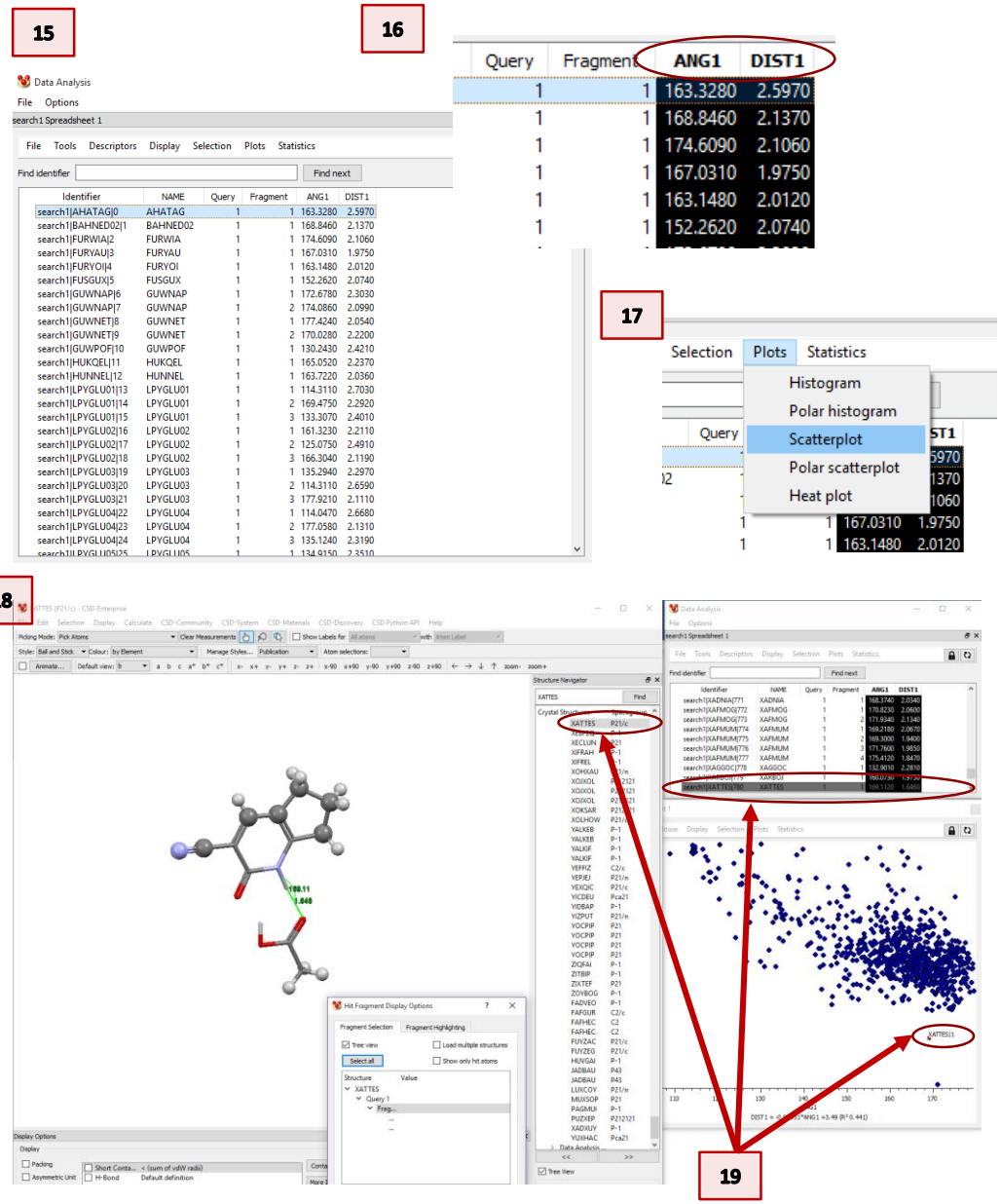
19. Clicking on any point in the scatterplot will highlight this entry in the **Data Analysis** window and display the structure in the Mercury app.

20. Clicking on a refcode in Mercury will highlight the entry in the **Data Analysis** window and display the refcode next to the corresponding point in the plot window.

Mining the data from the CSD

Now that you have performed the search and obtained a set of distance and angle parameters for interactions between amides and carboxylic acid groups, you will want to be able to see how the structure of interest, XECLUN from above, compares to the values from your search. Since this structure is already in the CSD, it will be returned by the search you just performed.

We want to know if the H-O distance and N-H---O angle in XECLUN are typical



20. Find XECLUN in the list of refcodes in the **Data Analysis** window, and click the line to highlight it. Note the values for the H---O distance (2.223Å) and the N-H---O angle (136.515°).

21. To put these values in context, click Statistics and then *Descriptive Statistics* from the drop-down menu.

22. This will produce a spreadsheet of statistics in the bottom half of the **Data Analysis** window. You can click the top bar of the spreadsheet window and drag it outside the **Data Analysis** window to make it easier to view the data.

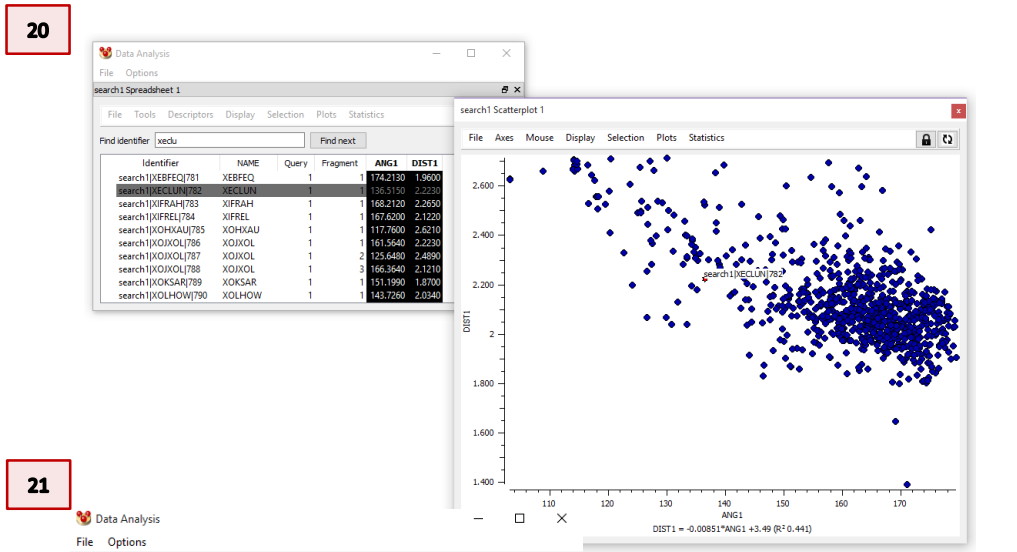
23. Look at the column labeled Mean. Note the difference between the average values for the entire data set and those for your molecule of interest. The H-bond distance in XECLUN is 0.097Å longer and the N-H---O angle is 24.371° more acute. While the distance is within one standard deviation, the angle is not. This is an atypical angle for this sort of interaction.

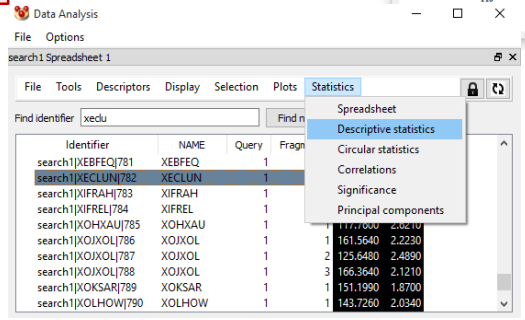
24. You can highlight a set of refcodes by selecting the corresponding points in the scatterplot. To do this, simply click and drag out a region in the plot.

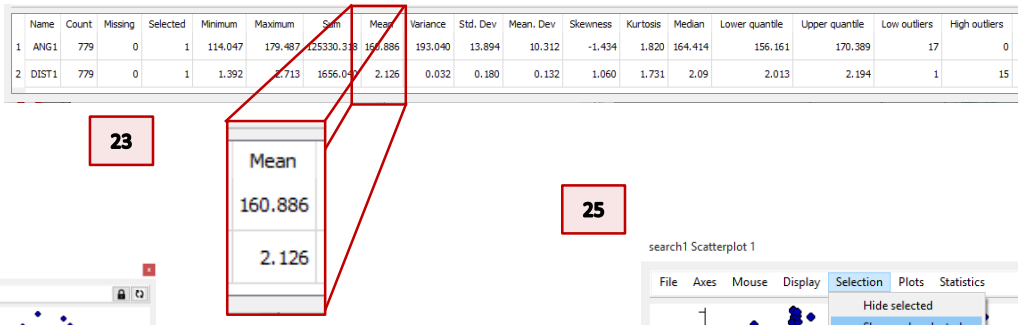
25. If you want to view only the selected entries, click Selection and then *Show only selected* from the drop-down menu. You can do this from either the **Data Analysis** window, or from the plot window.

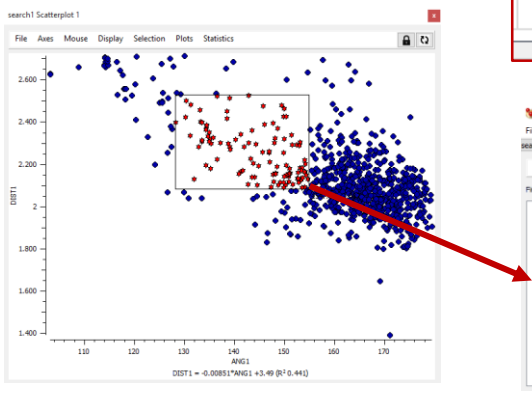
26. Use the commands in the File menu of the plot window to save or export your work.

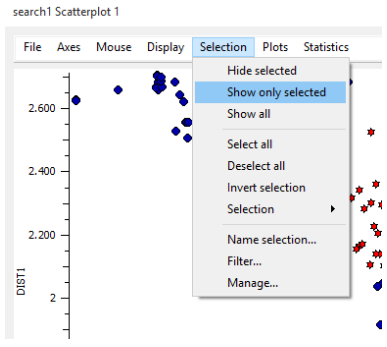
27. Close Mercury and the **Data Analysis** window when you are done with your work. Leave the ConQuest window with your search open for the next steps.











Advanced Data Mining

Suppose you were interested in looking at the interplay between the N-H distance and the strength of the hydrogen bonding interaction. To do this, we need to add the N-H distance to our list of geometric paramters.

1. Return to the *Build Queries* tab of ConQuest and click **Edit…** next to Query 1.

2. In the **Draw** window, click **ADD 3D**. Using the procedure from steps 2 and 3 above. ConQuest recognizes that this is a bonded contact so simply click **Done** in the *Geometric Parameters* dialogue box to add this distance to the list of 3D parameters.

3. Click **Search** and then **Yes** to overwrite the Query. Click **Start Search** after ensuring the boxes for the "3D coordinates determined" and "Only Organics" filters are set.

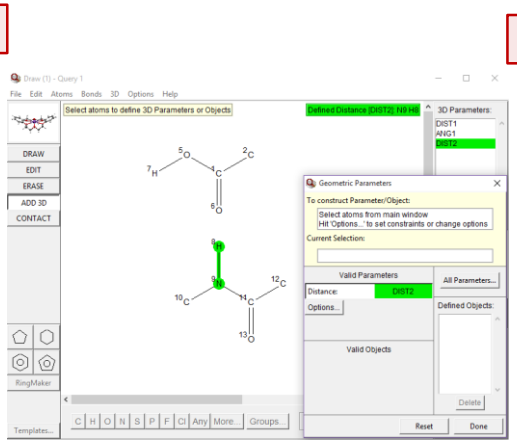
4. This search should again return 599 hits; you did nothing to change the fragments, only requested additional data to be returned.

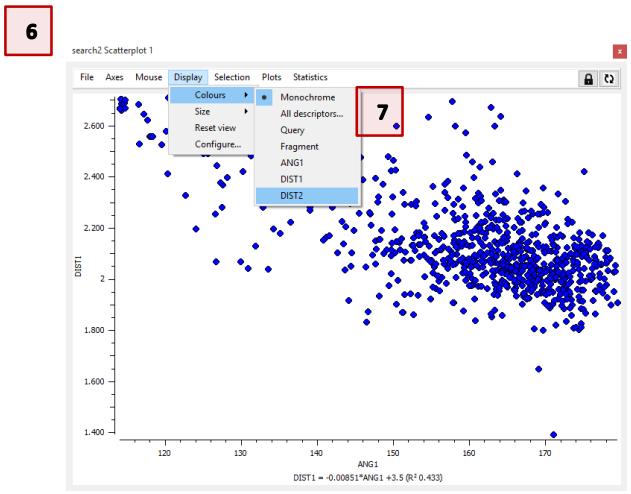
5. Again, click **Analyse Hitlist** and choose *Analyse Data* from the drop-down menu. Click **Analyze in Mercury** in the subsequent dialogue box.

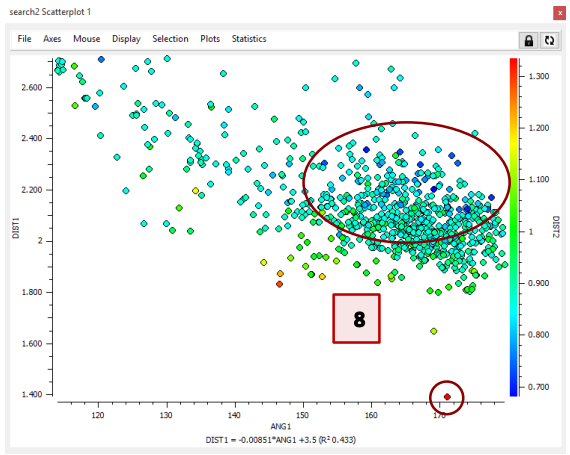
6. Following the instructions in steps 15-18 above, create a scatterplot of ANG1 and DIST1 from the **Data Analysis** window.

7. To see the correlation of these data with the N-H distance, choose Display from the plot window menu, then *Colors > DIST2* from the drop-down menu.

8. Here you can see that longer the N-H distances (warm colors) correlate with shorter C=O distances. However, the shortest N-H distances fall in the middle of the range of C=O distances, but correlate with larger N-H---O angles.



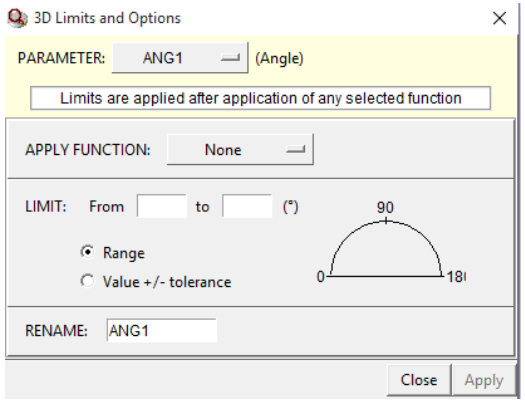




Conclusions

Now you have seen how to add 3D parameters to your query to mine the CSD for even more structural information. This tutorial has demonstrated a basic search using distance and angle information; however, you can also construct queries that include information about planes, vectors, centroids, dummy points and even atom characteristics such as atomic number and van der Waals radius.

The data analysis window allows you to plot your data in a variety of ways, determine statistics on your datasets, and to perform calculations on the values of the parameters returned by the search.



Further Exercises

- Choose a molecule or fragment from your own research and set up a query that will return 3D parameters of interest to you. Try plotting these in different ways.

- Use the Options associated with contacts, bonds or angles to filter the results returned by your search. For instance, in the search described in this tutorial, we could have used the Options for ANG1 to limit our search to angles between 150° and 180°.

- In the **Data Analysis** window, use the Tools menu to explore the *Calculator* functionality. You can use this to create new columns (descriptors) of calculated values.

- Again, use the Tools menu to explore the *CSD data…* functionality. This will allow you to add CSD entry data to your spreadsheet.