# Introduction to ConQuest

Version 1.6a – November 2017 CSD v 5.39

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table of Contents</td>
<td>2</td>
</tr>
<tr>
<td>Introduction</td>
<td>2</td>
</tr>
<tr>
<td>Overview of ConQuest</td>
<td>2</td>
</tr>
<tr>
<td>Introduction to the Draw Window</td>
<td>3</td>
</tr>
<tr>
<td>ConQuest sketching conventions</td>
<td>3</td>
</tr>
<tr>
<td>Example 1. Searching for Similar Molecules</td>
<td>4</td>
</tr>
<tr>
<td>Viewing search results</td>
<td>6</td>
</tr>
<tr>
<td>Conclusions</td>
<td>7</td>
</tr>
<tr>
<td>Further Exercises</td>
<td>7</td>
</tr>
<tr>
<td>Example 2. More Searching in ConQuest</td>
<td>8</td>
</tr>
<tr>
<td>Combining Searches</td>
<td>12</td>
</tr>
<tr>
<td>Exporting Results</td>
<td>13</td>
</tr>
<tr>
<td>Conclusions</td>
<td>13</td>
</tr>
<tr>
<td>Further Exercises</td>
<td>13</td>
</tr>
<tr>
<td>Example 3. Using 3D Information in Searches</td>
<td>14</td>
</tr>
<tr>
<td>Gathering 3D data for comparison</td>
<td>14</td>
</tr>
<tr>
<td>Searching the CSD with distance and angle constraints</td>
<td>18</td>
</tr>
<tr>
<td>Conclusion</td>
<td>20</td>
</tr>
<tr>
<td>Further Exercises</td>
<td>20</td>
</tr>
<tr>
<td>Example 4. Intermolecular Interactions with ConQuest</td>
<td>21</td>
</tr>
<tr>
<td>Mining the Data from the CSD</td>
<td>23</td>
</tr>
<tr>
<td>Advanced Data Mining</td>
<td>25</td>
</tr>
<tr>
<td>Conclusions</td>
<td>26</td>
</tr>
<tr>
<td>Further Exercises</td>
<td>26</td>
</tr>
</tbody>
</table>

The Cambridge Crystallographic Data Centre
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 v5.39, ConQuest v1.20 and Mercury v3.10.

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 (e.g. C or N or O).

Define bonds, angles or torsions to be monitored during the search, or define geometric objects e.g. planes, centroids that can be used in computing geometric parameters.

Ring template selector or builder

List of templates for challenging substructures e.g. fullerene

Choice of specific or general atom types/functional groups

Choice of bond types
Example 1. Searching for Similar Molecules

Sildenafil citrate or “Viagra” is a high profile 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.

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 drop-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 drop-down menu, then left-clicking on the bond.

10. Modify the bond types shown. Click the Bond type drop-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.
Viewing search results

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

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

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

18. To inspect crystal packing, select **Analyse Hitlist > Visualise Structures**.

19. The **All 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.
Conclusions

There are 375 CSD entries 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), however, Vardenafil is not in the CSD. Two closely related compounds, BITZOT and PEYGIL, are. Many other hits from this search 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.

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 13,000 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.

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

(a) primary alcohols
(b) cyclic and acyclic secondary alcohols
(c) cyclic secondary alcohols only
(d) unsaturated cyclic and acyclic alcohols and carboxylic acids
(e) unsaturated cyclic alcohols only e.g. phenols
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 24,000 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 (fragment C above).

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

18. This time you will get over 16,000 hits. Again, scroll through the refcode list to explore your results. Notice that refcode ALOMEV appears in both searches.
19. Again repeat steps 7 and 8 above to return to the Draw window showing the fragment from Query 2.

20. Right-click the C atom and choose Hydrogens and 0 from the drop-down menu. Right-click the C atom again and choose Number of Bonded Atoms, and then 3. This will give you a fragment that matches unsaturated alcohols (fragment “D” 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 46,000 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 “E”). 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 24,000 hits for this Query. Scroll through the refcode list to explore your results. Note that ALOLEU is included in this search as well as the previous search.
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 labelled “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 850 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 (or click the Reset button). 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 7816 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 70,000 hits!
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. Then 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 (7816) 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 example, 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. Using 3D Information in Searches

Gathering 3D data for comparison

Queries based on 3D geometric constraints involving atoms, centroids or planes give important data which help relate the structure of a molecule to its function. ConQuest’s Draw feature allows the incorporation of 3D queries or filters, which can be used to extract metric data from the molecules in the CSD or to limit searches to particular spatial arrangements of chemical moieties. This can be extremely helpful when doing structure-function or structure-property research in fields from drug discovery to materials engineering.

The following example relates to a pharmacophore used for drug discovery, but can easily be adapted to organometallic or organic-materials cases. This example looks at the relationship between the propyl group and the pyrazole pyrimidine aromatic plane of sildenafil (Viagra), shown at right. This tutorial assumes you have knowledge of the Draw function in ConQuest.

1. Open ConQuest and click the Name/Class button to start a text-based search for a compound name. Type the word sildenafil in the Compound Name box. Click Add to include the word sildenafil in the “Contains:” box. Click Search to begin the search.

2. In the Search Setup dialog box, tick the box for “3D coordinates determined” and then click Start Search. About 25 hits will be returned. These hits all contain the word “sildenafil” in the compound name or synonym field.

3. Click the Chemical tab on the left of the View Results display. This shows the chemical information for each hit including the Name (typically IUPAC) and any synonyms supplied by the depositors. The text that matches your search term will be highlighted in yellow. Scroll through your hits to see what matches were found.
4. In order to assess the relationship of the propyl group to the aromatic plane of the pyrazole pyrimidine moiety, we will need to do a 3D geometry search. Click on the refcode QEGTUT in your hitlist. This is the structure of neutral sildenafil, with no solvents or other molecules in the lattice. Click the Diagram tab on the left to show the 2D drawing.

5. You can use this 2D diagram as the basis of the 3D geometry search. Click the Use as Query… button. Make sure the box is ticked for “Include hydrogen atoms” in the Use-as-Query Options dialog box. The radio button should be ticked by “Biggest chemical unit only”; the other options are greyed-out because there is only one molecule in the structure. Click OK to continue. You will get a warning that the OEt group cannot be converted and instead ConQuest is using O. We do not need to worry about this group, so simply click OK.

6. ConQuest now shows the structure from QEGTUT as Query 2. We will need to edit this to remove some features of the structure and to add our 3D parameters. Click Edit… to edit the query. This will launch the Draw window.

7. First, delete all atoms but the pyrazole pyrimidine and propyl moieties. To do this, click the Erase button on the left of the window. This will change your cursor to a square; click on each atom to be deleted.

8. Because we want to search the torsion angle of the propyl ring, we will need to draw out all the carbon atoms in the chain. Using the procedure in Step 7, delete the “Et” group. Then click Draw and add 2 carbon atoms to complete the propyl group. Delete the hydrogen atoms on the carbon atoms directly bonded to the ring by, right-clicking on the carbon atoms with the added hydrogens, select Hydrogens and then select Clear.

9. Now we need to add in the search for 3D parameters. Click the ADD 3D button to launch the Geometric Parameters dialog box. This is the window for defining all 3D constraints and objects. Click each atom in the rings to select them. They will be highlighted in blue. Click the Define button next to “Plane” in order to define the plane.
10. Now we need to define the distance of the propyl group to the plane we have just defined. Click the middle carbon of the propyl chain to select it, then select PLN1 in the Defined Objects section. Then click the Define button next to “Distance:” in the Valid Parameters section.

11. There is now a distance constraint of unspecified length. For this search, we do not want to restrict it as we merely want to gather data.

12. To specify the torsion angle of the propyl chain, start with the terminal carbon and click the three carbon atoms of the chain and the carbon atom in the pyrazole ring. It is important to click these atoms in bonding order. Click the Define button next to “Torsion:” in the Valid Parameters section as you did for the distance in Step 10. Again, there are no limits on this angle.

13. You will now see two entries, abs(DIST 1) and (TOR1), in the 3D Parameters box in the upper right hand of the Draw window. Click the Done button in the Geometric Parameters dialogue box.

14. Click Search to launch the searching window. You will be prompted to “Overwrite existing Query.” Click No to save this as a separate query.

15. Rather than searching through the entire database, we want to search only the structures in the sildenafil hitlist we just produced. Click the Select Subset button in the Search Setup dialogue box.

16. In the Restrict Search dialogue box, tick the radio button to select Entries in a hitlist loaded this session. A box will appear with the text “Restrict based on search this session” and a dropdown box labelled search1.

17. Click the search1 box to select it. The hit list information will be shown in the Current Restrictions section. Click OK to close the window.
18. 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 26 refcodes. Make sure both boxes are ticked and click Start Search to begin the search.

19. In the View Results tab, notice that now the results have the matching fragment highlighted, as well as both parameters measured. The data is measured, now it needs to be analysed. Click Analyse Hitlist and select Analyse Data.

20. In the Analyse in Mercury dialogue box, make sure the box is ticked to “Include Defined Parameters.” We are not interested in any other parameters, so simply click the Analyse in Mercury button to view the data.

21. This launches Mercury and the Data Analysis window, which shows all the data from your query in a spreadsheet, sorted by refcode. NOTE: The Data Analysis window may be behind the main Mercury display.

22. To view the distribution of distances in this hitlist, click the abs(DIST 1) header in the column to highlight the column, then click Plots from the top menu and then select Histogram.

23. The distances seem to cluster in two regions, one centred on 0.3Å and one centred on 1.4Å, with one outlier in the middle near 0.8Å. Click on this bar of the histogram to highlight the corresponding entry in the spreadsheet.

24. You will see QEGTUT highlighted in the spreadsheet. This is the structure of sildenafil with no solvents or other molecules in the lattice. You can click on this entry in the spreadsheet to see the structure in Mercury. Click to select other bars in the histogram and the corresponding entries to see the other hits.
25. To see how the torsion angle of the propyl group relates to the distance above the plane you can create a scatterplot. To do this, click the TOR1 header to select the torsion data column in addition to the distance data. With both columns selected, click Plots from the top menu and then select Scatterplot.

26. The scatterplot shows a similar clustering to the histogram, which is what we would expect. The data on the left of the plot, i.e. smaller out-of-plane distances, show larger torsion angles. The data on the right hand side of the plot, i.e. larger out-of-plane distances correspond to smaller torsion angles.

Searching the CSD with distance and angle constraints

A major component of structure-property or structure-function research involves searching for a specific geometric arrangement of atoms. For instance, scaffold hopping in drug discovery involves searching for known compounds with the same chemical moieties arranged around a variable molecular core. Using ConQuest to perform this type of search can return a variety of known crystal structures of potentially bioactive molecules.

The following example, based on the pharmacophore shown at right, shows how to set up a geometry based search with specific metric restraints.

1. Open ConQuest and click Draw to start a new substructure query. Draw the carboxylic acid group, and the two toluene groups. Make sure they are not bonded to each other. Make sure the acid hydrogen is included by right-clicking the oxygen atom, choosing Hydrogens and then Generate.

2. Following the procedure above (Step 9, etc.) click Add 3D to launch the Geometric Parameters dialogue box. Click each of the six C atoms in one ring to select them, then click the Define button next to “Centroid:” to define the centroid. Repeat the centroid definition step again for the other aromatic ring. Then repeat each selection, but click Define next to “Plane:” in order to define the planes, one each for each aromatic ring. In the end you should have four objects in the Defined Objects box: two centroids and two planes.
3. To ensure the search returns structures with these groups in the correct geometry, various distances and angles must be restrained.

4. To restrain the distance between the two ring centroids, click on CENT1 and CENT2 from the Defined Objects section to select that distance, and then click the Define button next to Distance. Choose “Intramolecular Distance” from the Distance Type dialogue box to ensure these features are both in the same molecule and click OK to exit the dialogue box.

5. Back in the Geometric Parameters dialogue box, click Options… to define the limits on the centroid-centroid distance. In the “LIMIT:” boxes, type 5.0 for “From” and 6.0 for “to”. Click OK to set the range limits.

6. Repeat this step for the distances between each centroid and the central carbon of the carboxylic acid: Click CENT1 in the Defined Objects: section and then click the central carbon in the carboxylic acid. Click the Define… button next to “Distance:” and make sure the button is ticked next to “Intramolecular Distance” in the Distance Type dialogue box. Click the Options… button and set the range limits to 2.5 and 6.0. Repeat again for CENT2 but this time set the range limits to 4.0 and 6.0.

7. To define the dihedral angle between the planes, select both planes from the Defined Objects section. Then click the Define… button next to “Angle”. Click Options… and set the limit between 20° and 80°.

8. You should now have four entries in the 3D Parameters box in the Draw window. Click Done to close the Geometric Parameters dialogue box.

9. Click the Search button to launch the Search Setup dialog box. For this example, since we are looking for potential pharmaceuticals, we are only interested in organic compounds. In the Filters tab, tick the boxes for “3D coordinates determined” and “Only Organics”. 
10. Click **Start Search** to begin the search. The search will return about 50 hits, with the corresponding distances and angles from the 3D geometry queries that have been included.

11. Click the **Analyse Hitlist** button and choose **Analyse Data**. This will load all the hits into Mercury for visualisation as well as launch the **Data Analysis** window. Explore the structures visually to see what sorts of molecules have been identified. Use the **Data Analysis** feature to explore the relationship between various parameters. You will find a diverse set of molecules returned by this search. Adding more functional groups and changing the limits on the distances and angles will affect the hits returned by the query.

**Conclusion**

These two exercises show how to set up 3D parameters for searching metric data and for filtering the hits returned based on specific limits. These tools are extremely useful in creating advanced searches to identify particular molecules, ligands or moieties that may exhibit useful properties such as increased bioactivity or catalysis.

**Further Exercises**

- Repeat the first exercise but do not limit the search to only sildenafil molecules. Do you find any interesting results?

- Repeat the second exercise but change the limits on the distances and angles to make them broader. What do you find?
Example 4. 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 analyse 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 analyse 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 around 650 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 analyse 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 analysing 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 minimise 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 ANG1 and DIST1 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.
21. 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°).

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

23. 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.

24. Look at the column labelled 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 0.098Å longer and the N-H---O angle is 24.408° more acute. While the distance is within one standard deviation, the angle is not. This is an atypical angle for this sort of interaction.

25. 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.

26. 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.

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

28. 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 parameters.

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, define the N-H bond distance. ConQuest recognises 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 around 650 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 Analyse 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 Colours > DIST2 from the drop-down menu.

8. Here you can see that longer the N-H distances (warm colours) 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. You can also define dummy points and use atom characteristics such as atomic number and van der Waals radius in searches. Please see the ConQuest documentation for more details.

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.