

Using 3D Information in Searches in ConQuest (CQ-002)

Developed using
2022.2 CSD Release
CSD version 5.43 (November 2021) + 2 data updates

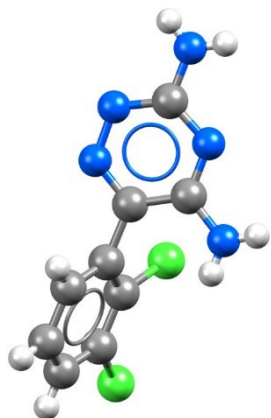


Table of Contents	
Introduction.....	2
Objectives	2
Pre-required skills.....	2
Materials.....	2
Using 3D Information and Subsets in Searches.....	3
Gathering 3D data for comparison.....	3
Searching the CSD with distance and angle constraints	7
Conclusion	9
Further Exercises	9
Conclusions.....	9
Next steps.....	10
Feedback	10
Glossary	10
Review. ConQuest Interface.....	11
Review. Draw Window	12
ConQuest sketching conventions.....	12

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 references to name a few. Chemical substructure search in ConQuest also includes the ability to define chemical constraints such as charge, hybridization state and cyclicity.

Before beginning this tutorial, ensure that you have a registered copy of CSD-Core 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 using ConQuest for [substructure](#) searching. At the end of this tutorial, you should be able to:

- Use 3D information in searches.
- Search with geometric constraints.
- Use [subsets](#) to restrict search results.

This workshop will take approximately **1** hour to be completed.

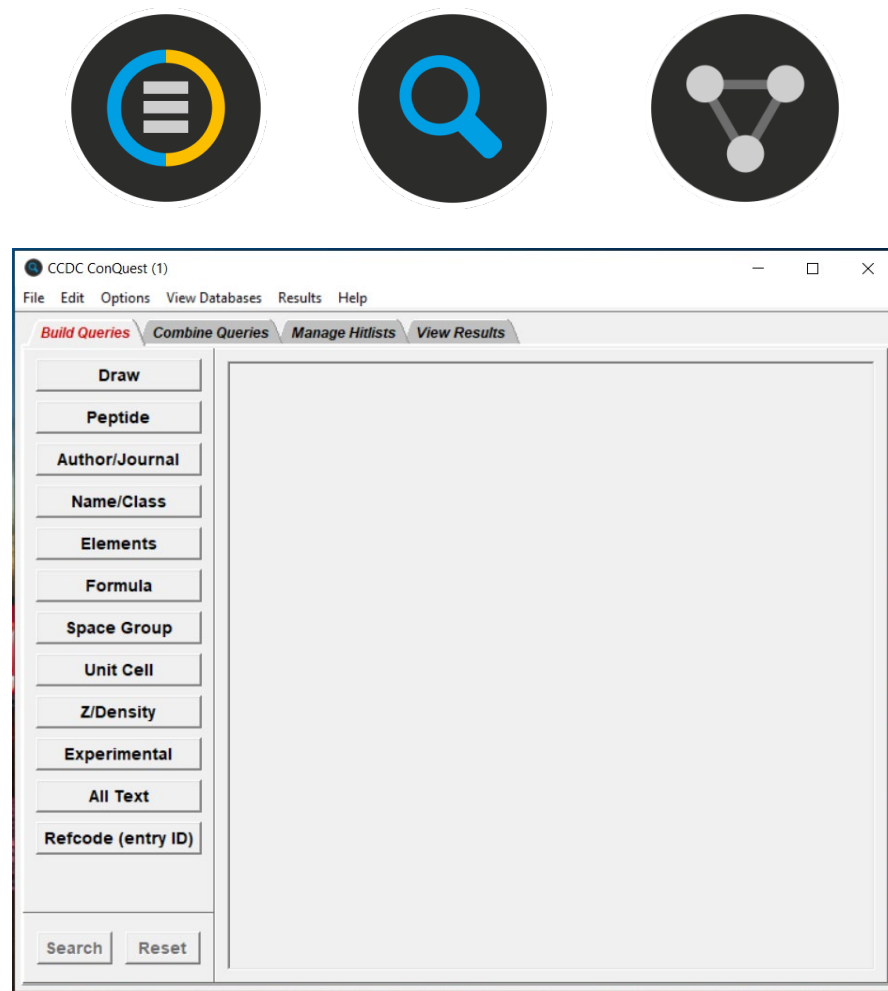
Note: The words in *Blue Italics* are reported in the [Glossary](#) at the end of this handout. A [review](#) of the ConQuest interface is also at the end of this handout.

Pre-required skills

For this tutorial, we recommend running through the introduction to ConQuest and visualizations in Mercury workshop which can be found [here](#).

Materials

There are no additional materials required for this workshop.

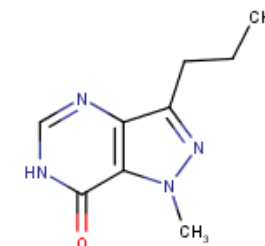
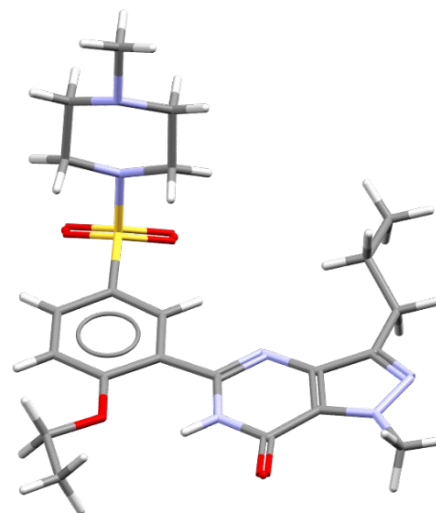


Using 3D Information and Subsets in Searches

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. The ConQuest *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 specific spatial arrangements of chemical moieties. This can be extremely helpful when doing structure-function or structure-property research in fields ranging 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 prior knowledge of the *Draw* function in ConQuest.



The 3D structure of sildenafil (QEQTUT, left) and a 2D sketch of the fragment used for this exercise (right).

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**. At least 40¹ 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.

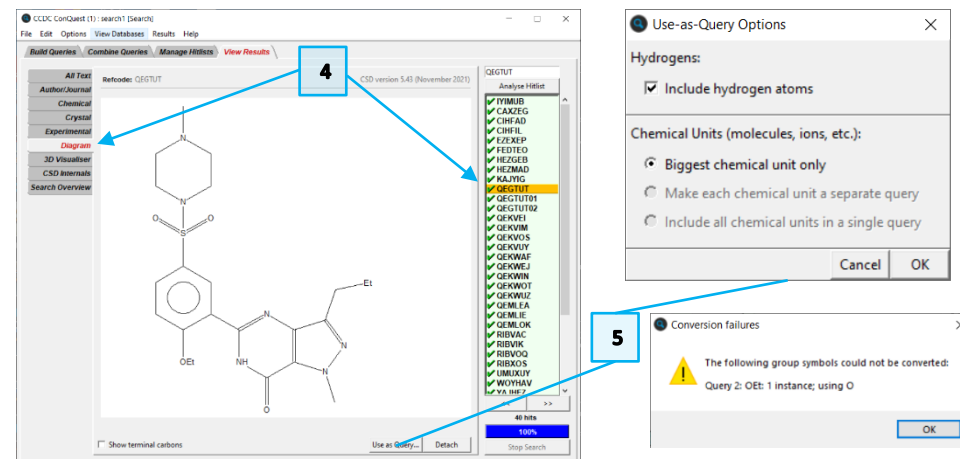
The figure displays three screenshots of the ConQuest software interface, illustrating the search process:

- 1. Name/Class (1) - New:** The 'Compound Name' field contains 'sildenafil'. The 'Contains:' field is empty. The 'Search' button is highlighted.
- 2. Search Setup:** The 'Search Name' is 'search1'. The 'Available Databases' section shows 'CSD version 5.43 (November 2021) + 2 updates'. The 'Filters' tab is selected, and the '3D coordinates determined' checkbox is checked. The 'Start Search' button is highlighted.
- 3. View Results:** The 'Chemical' tab is selected. The search results are displayed, showing the chemical structure of sildenafil (QEQTUT) and its synonyms. The search term 'sildenafil' is highlighted in yellow.

¹ As of version 2022.2 + 2 data updates of the CSD.

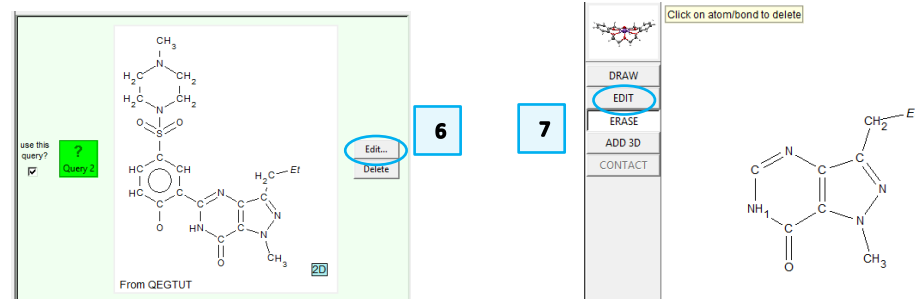
To assess the relationship of the propyl group to the aromatic plane of the pyrazole pyrimidine moiety, we will need to conduct a 3D geometry search. Click on the [refcode](#) QEGTUT in your [hitlist](#). This is the structure of neutral sildenafil, with no solvents or other molecules in the lattice. Click the **Diagram** tab on the left to show the 2D drawing.

- 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 that should be ticked is "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**.

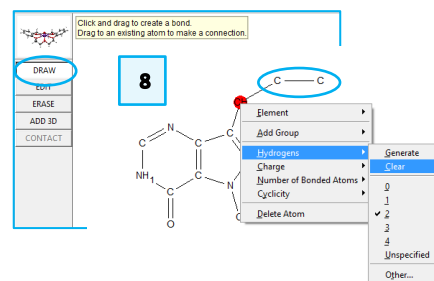


- ConQuest now shows the structure from QEGTUT as Query 2. We will need to edit this to remove some features of the structure and to add our 3D parameters. Click **Edit...** to edit the query. This will launch the **Draw** window.

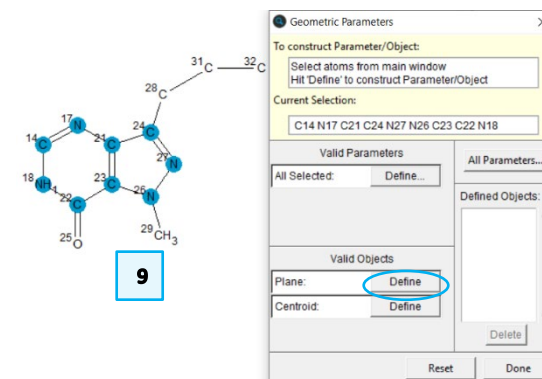
- First, delete all atoms but the pyrazole pyrimidine and propyl moieties (see figure 7). To do this, click the **Erase** button on the left of the window. This will change your cursor to a square; click on each atom to be deleted.



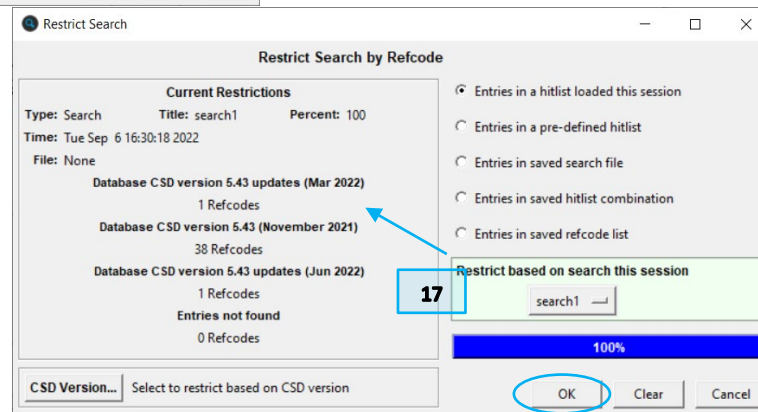
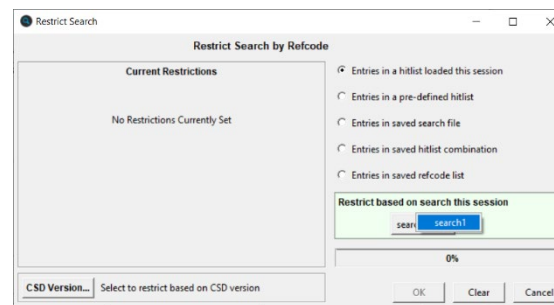
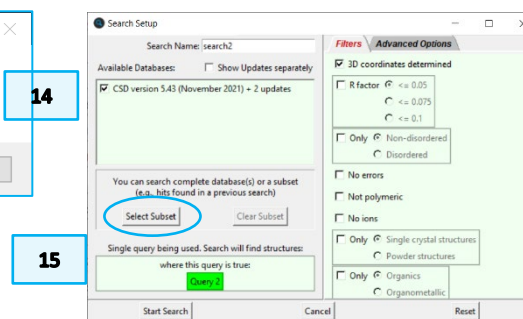
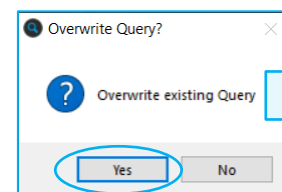
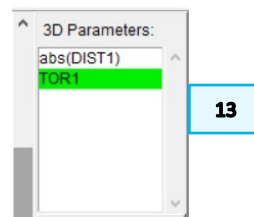
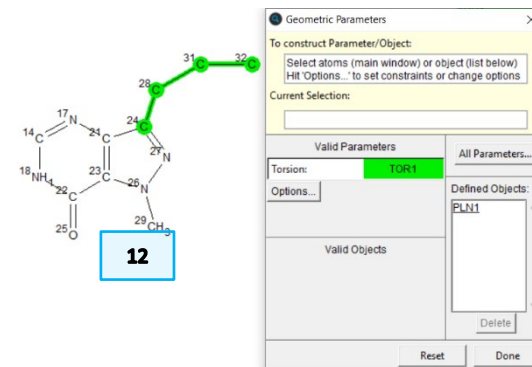
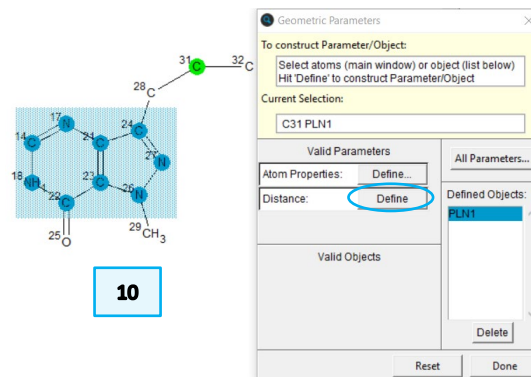
- Because we want to search the [torsion](#) angle of the propyl ring, we will need to draw out all the carbon atoms in the chain. Using the procedure in Step 7, delete the "Et" group. Then click **Draw** and add 2 carbon atoms to complete the propyl group. Delete the hydrogen atoms on the carbon atom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



- 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" to define the plane.



9. 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.
10. 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.
11. To specify the torsion angle of the propyl chain, start with the terminal carbon (C32) and click the three carbon atoms of the chain and the carbon atom in the pyrazole ring (C24). *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.
12. 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.
13. Click **Search** to launch the searching window. You will be prompted to "Overwrite existing Query." Click **Yes** to overwrite and save this as Query 2.
14. 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* dialog box.
15. 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**.
16. 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.



17. 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 40 refcodes (as of the 2022.2 release). Make sure both boxes are ticked and click **Start Search** to begin the search.

18. 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**.

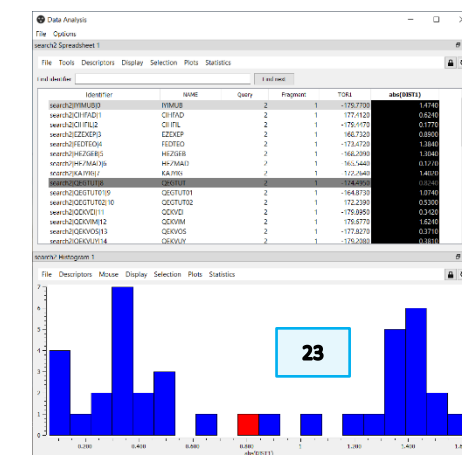
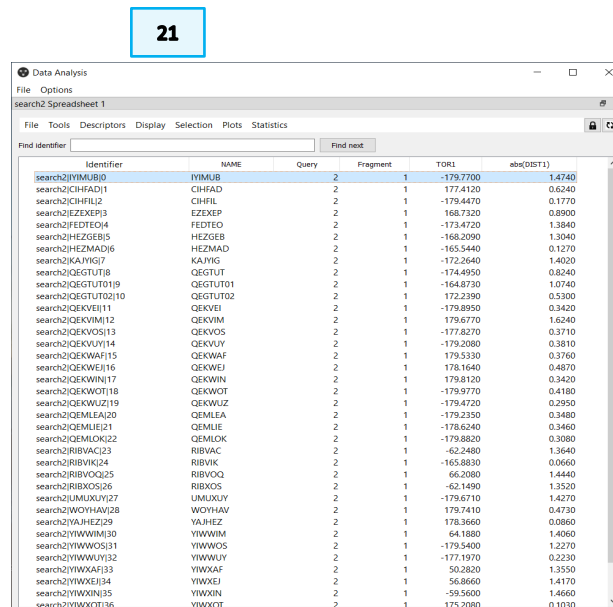
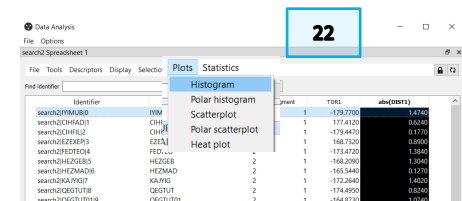
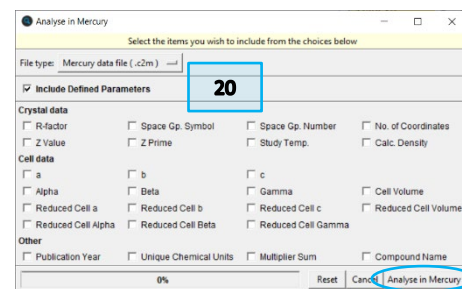
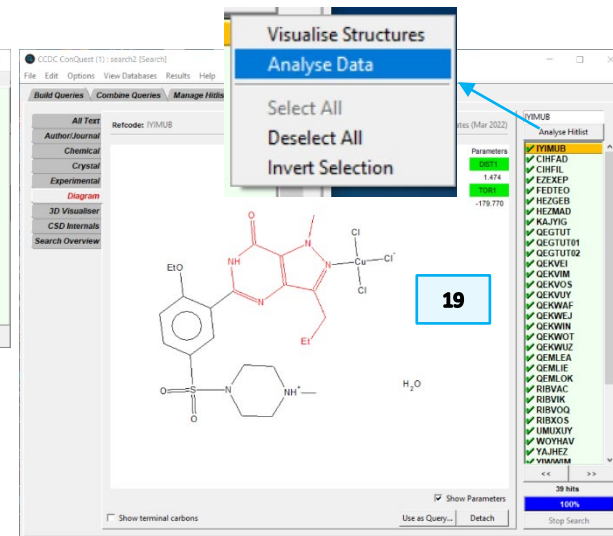
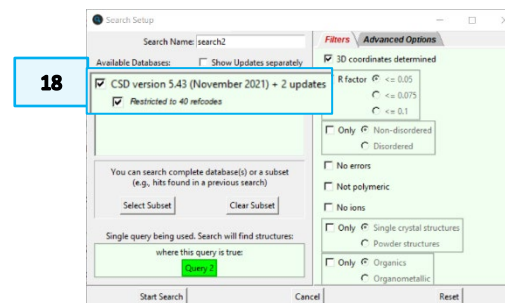
19. In the *Analyse in Mercury* dialogue box, make sure the box is ticked to "Include Defined Parameters." We are not interested in any other parameters, so simply click the **Analyse in Mercury** button to view the data.

20. 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.

21. 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**.

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

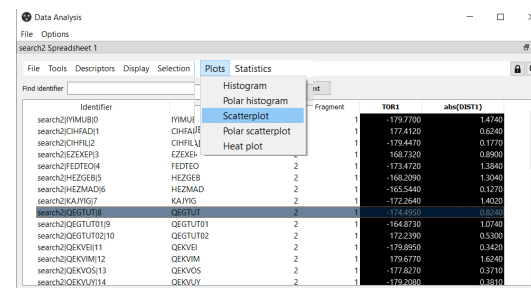
23. You will see QEGTUT highlighted in the spreadsheet. This is the structure of sildenafil with no solvents or other molecules in the lattice. You can click on this entry in the spreadsheet to see the structure in Mercury. Click to select other bars in the histogram and the corresponding entries to see the other hits.



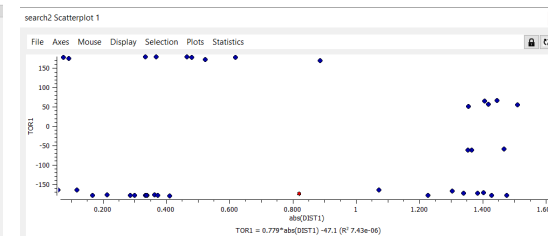
24. To see how the torsion angle of the propyl group relates to the distance above the plane you can create a scatterplot. To do this, click the *TOR1* header to select the torsion data column in addition to the distance data. With both columns selected, click *Plots* from the top menu and then select *Scatterplot*.

25. 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. Close ConQuest as you will be drawing a new structure in the next section.

25



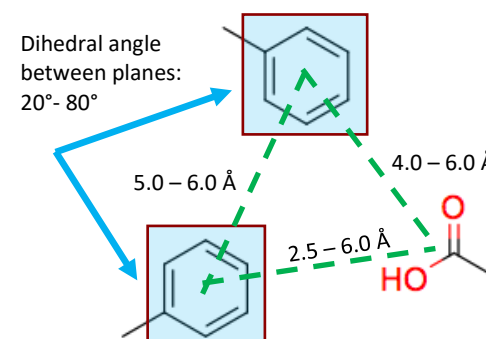
26



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.



The sample pharmacophore, known for CIC-K Cl⁻ channel activation (DOI: 10.1073/pnas.0708977105)

1. Open ConQuest and click **Draw** to start a new substructure query. Draw the carboxylic acid group, and the two toluene groups (see figure 1). 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.

- To ensure the search returns structures with these groups in the correct geometry, various distances and angles must be restrained.
- 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.
- 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 Å for “to”. Click **OK** to set the range limits. You can choose to rename the parameter here, but we will leave it as DIST1.
- 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.
- 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°.
- You should now have four entries in the **3D Parameters** box in the **Draw** window. Click **Done** to close the *Geometric Parameters* dialogue box. You can click on each defined parameter to review the defined objects.
- 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”.

4

5

6

7

8

9

10. Click **Start Search** to begin the search. The search will take a couple of minutes and should return at least 281 hits (*2022.2 release*), 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*. In the *Analyse in Mercury* dialogue box, make sure the box is ticked to “Include Defined Parameters.” then click the **Analyse in Mercury** button to view the 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 kinds 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 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 molecules, ligands or moieties that may exhibit useful properties such as increased bioactivity or catalysis.

Further Exercises

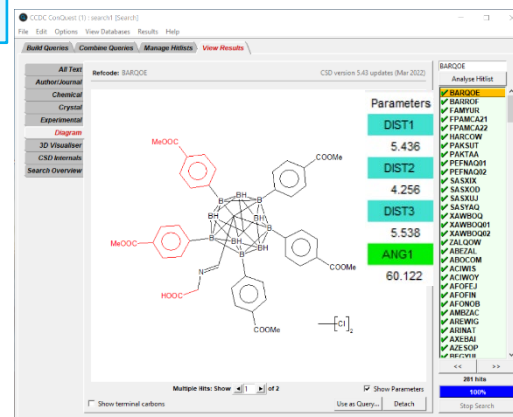
- Repeat the exercise but change the limits on the distances and angles to make them broader. What do you find?

Conclusions

This workshop introduced using 3D parameters while searching in ConQuest and how to utilize geometric constraints. You should now be familiar with:

- Using the compound name search function.
- Using previous result as a new query.
- Defining 3D parameters in your search.
- Using subsets in your search.
- Defining geometric constraints in your search.
- Exporting and visualizing structures in Mercury.
- Data analysis in Mercury.

10



11

The screenshot shows the Data Analysis window with a table of search results. The table has columns for Identifier, NAME, Query, Fragment, ANG1, DIST1, DIST2, and DIST3. The results are sorted by ANG1 in descending order.

Identifier	NAME	Query	Fragment	ANG1	DIST1	DIST2	DIST3
search1SASXOD18	SASXOD	1	1	53.0850	5.7170	2.8550	4.9180
search1SASXOD19	SASXOD	1	2	53.0850	5.7170	2.8550	4.9180
search1SASXU120	SASXU1	1	1	54.2940	5.8070	2.8730	4.8580
search1SASXU121	SASXU1	1	2	54.2940	5.8070	2.8730	4.8580
search1SASYAQ22	SASYAQ	1	1	56.4900	5.8510	2.8640	4.9150
search1SASYAQ23	SASYAQ	1	2	56.4900	5.8510	2.8640	4.9150
search1XAWBOQ24	XAWBOQ	1	1	56.5960	5.0310	2.8700	5.6470
search1XAWBOQ125	XAWBOQ1	1	1	49.2890	5.0710	2.8600	5.6610
search1XAWBOQ226	XAWBOQ2	1	1	44.8830	5.0720	2.8660	5.6290
search1XAWBOQ227	XAWBOQ2	1	2	31.3360	5.2020	2.8600	5.6210
search1ZALQOW28	ZALQOW	1	1	72.2550	5.9490	2.8980	5.0680
search1ZALQOW29	ZALQOW	1	2	72.2550	5.9750	2.8980	5.0030
search1ZALQOW30	ZALQOW	1	3	72.2550	5.9490	2.8980	5.0680
search1ZALQOW31	ZALQOW	1	4	72.2550	5.9750	2.8980	5.0030
search1ABEZAL32	ABEZAL	1	1	74.8970	5.0750	5.4280	5.0700
search1ABEZAL33	ABEZAL	1	2	74.8970	5.0750	5.0700	5.4280
search1ABOCOM34	ABOCOM	1	1	57.9690	5.9520	4.2370	5.0470
search1ABOCOM35	ABOCOM	1	2	57.9690	5.9520	5.0470	4.2370
search1ACWIS36	ACWIS	1	1	64.5860	5.9280	2.8790	4.5410
search1ACWIS37	ACWIS	1	1	75.9260	5.9920	2.8810	4.4700
search1AFOFE138	AFOFE1	1	1	53.2490	5.1680	4.3960	4.3780
search1AFOFE139	AFOFE1	1	2	75.9360	5.0930	4.4720	4.6470
search1AFOFE140	AFOFE1	1	1	70.3380	5.0030	4.2690	4.4450
search1AFOFE141	AFOFE1	1	2	51.8990	5.1310	4.2660	4.4450
search1AFOFE142	AFOFE1	1	3	72.9840	5.2670	4.4330	4.4430
search1AFOFE143	AFOFE1	1	1	65.6500	5.0410	2.8530	5.6670
search1AMBEZAC44	AMBEZAC	1	1	70.5870	5.0540	2.8760	5.1020
search1AREWIG45	AREWIG	1	1	38.5400	5.8710	4.9760	4.9040
search1AREWIG46	AREWIG	1	2	38.5400	5.8710	4.6170	5.9780
search1AREWIG47	AREWIG	1	3	38.5400	5.8710	4.9040	4.9760
search1AREWIG48	AREWIG	1	4	38.5400	5.8710	5.9780	4.6170
search1ARINAT49	ARINAT	1	1	64.0420	5.5570	3.7520	5.3730
search1AXEBAI50	AXEBAI	1	1	54.4970	5.2130	5.7420	5.4470
search1AXEBAI51	AXEBAI	1	2	77.8630	5.1990	2.8630	5.7420
search1AXEBAI52	AXEBAI	1	3	54.4970	5.2130	5.4470	5.7420

Next steps

You can explore other workshop materials for ConQuest from the [CSD-Core workshops](#) section. The ConQuest user guide and other available ConQuest documents can be found [here](#).

Feedback

We hope this workshop improved your understanding of ConQuest and you found it useful for your work. As we aim to continuously improve our training materials, we would love to get your feedback. Click on [this link](#) to a survey, it will take less than 5 minutes to complete. The feedback is anonymous. You will be asked to insert the workshop code, which for this self-guided workshop is CQ-002. Thank you!

Glossary

Centroid - In Mercury, the centroid is the geometric centre of a group of selected atoms.

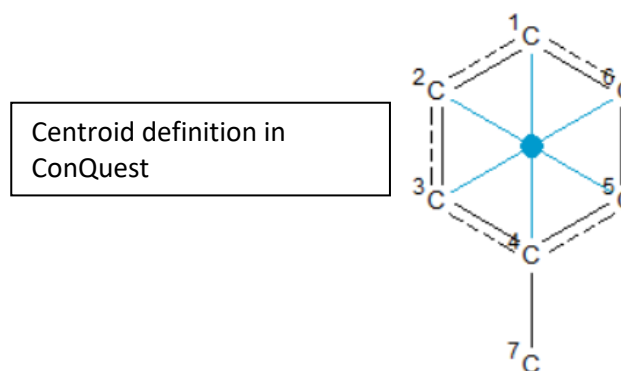
CSD Refcode - A database reference code, containing 6-8 characters and used to identify entries in the CSD. Learn more about CSD Refcodes in the blog: <https://www.ccdc.cam.ac.uk/Community/blog/post-20/>.

Hitlist - A hitlist is a subset of CSD entries which can include search results, refcode lists, or the results of combining these.


Subset - Subsets break the database down into more manageable subcategories and is an efficient way to search a certain category of compounds.

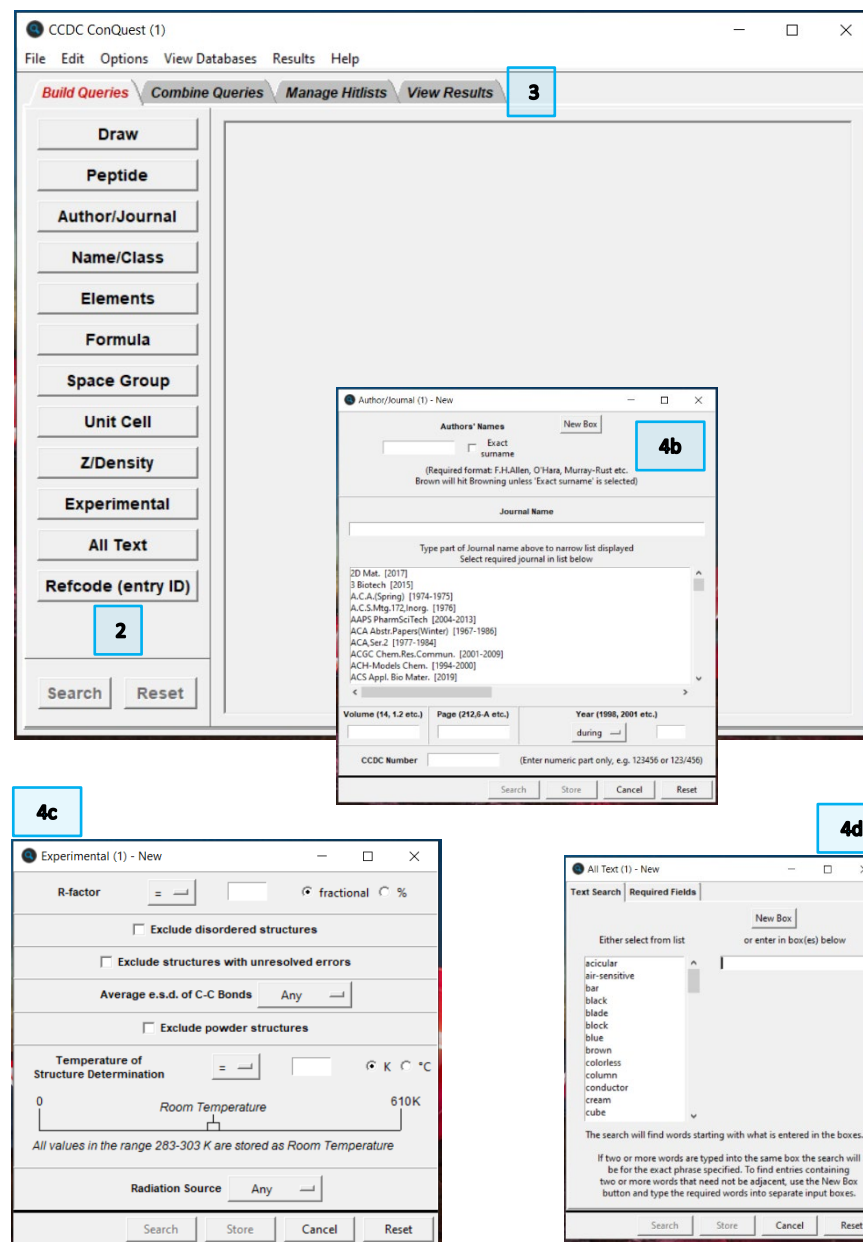
Substructure – A substructure is a part or section of a whole molecule.

Torsion Angles - Torsion angles are used to describe conformations around rotatable bonds. The torsion angle between 4 atoms A-B-C-D is the angle by which the vector A-B must



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).

