Intermolecular Interactions Searching in ConQuest
(CQ-003)

Developed using
2022.2.0 CSD Release
CSD version 5.43 (November 2021) + 2 data updates
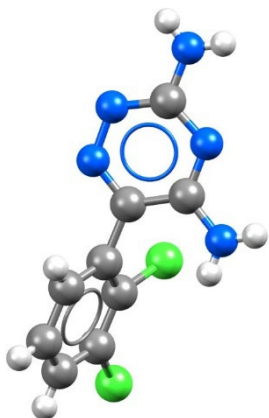
CCDC
advancing structural science

Table of Contents

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 constrains 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:

- Search for intermolecular interactions.
- Analyse data in Mercury.

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, advanced searching in ConQuest and visualizations in Mercury workshop which can be found here.

Materials

There are no additional materials required for this workshop.
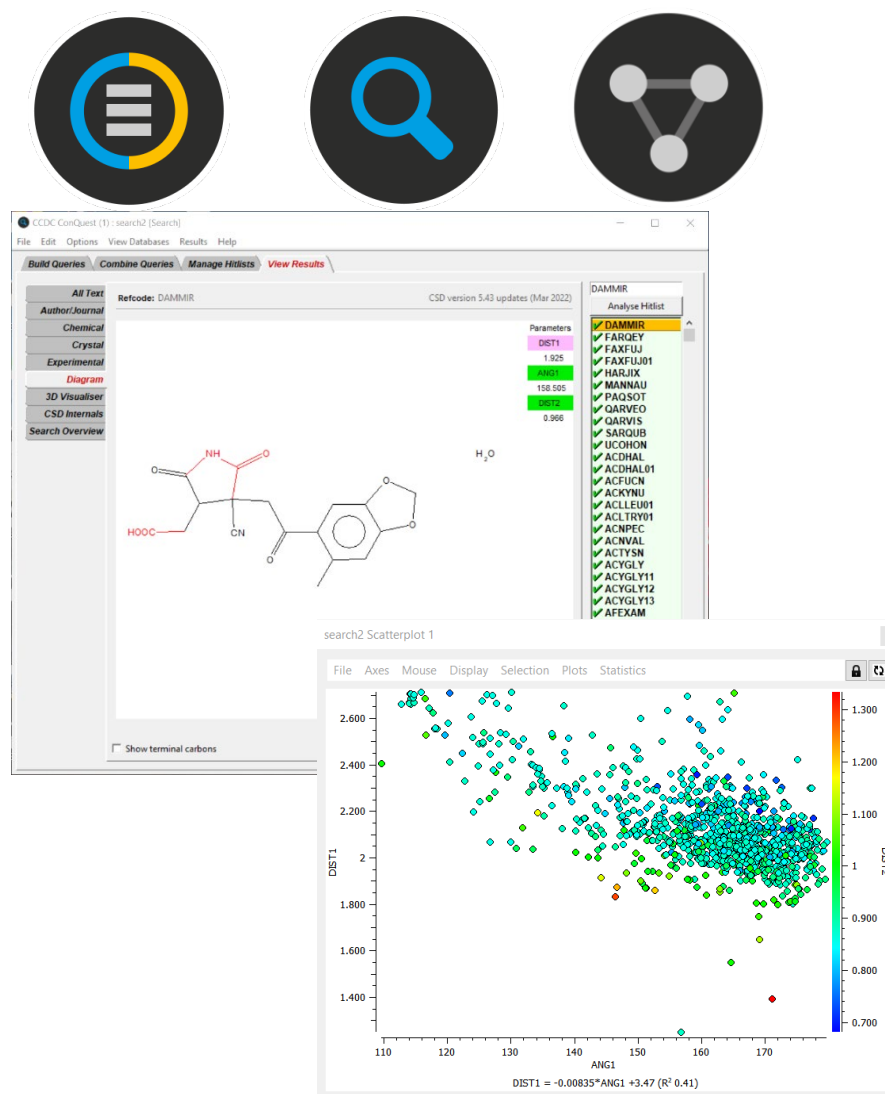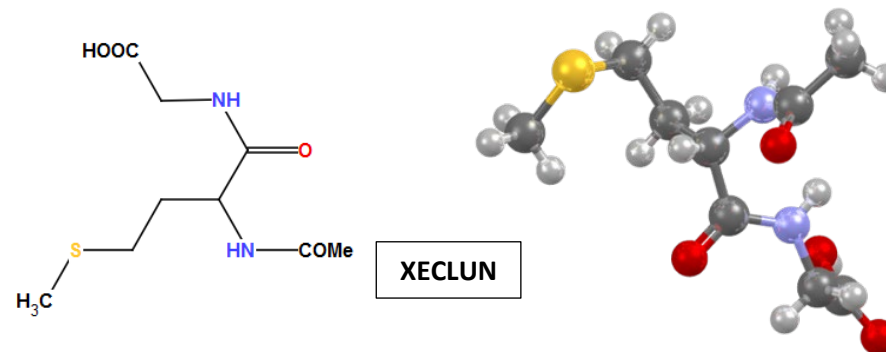
Intermolecular Interactions with ConQuest

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

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

1. Start out by drawing a generic carboxylic acid and amide group using the ConQuest **Draw** window. For ease of setup, 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 the Geometric Parameters 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.

XECLUN

6. Now we want to define the N-H---O angle. In the **Draw** window 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 do not 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 at least 820[1] 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.

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

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.

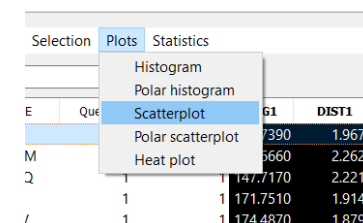We want to know if the H-O distance and N-H---O angle in XECLUN are typical.
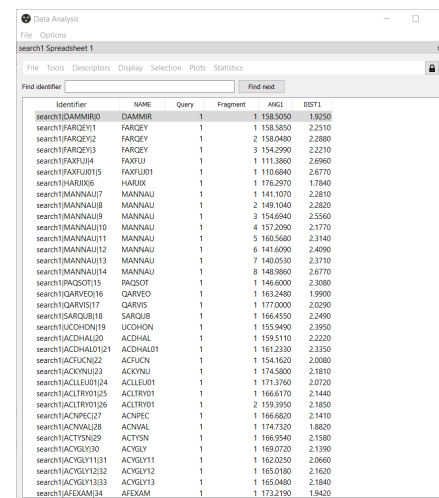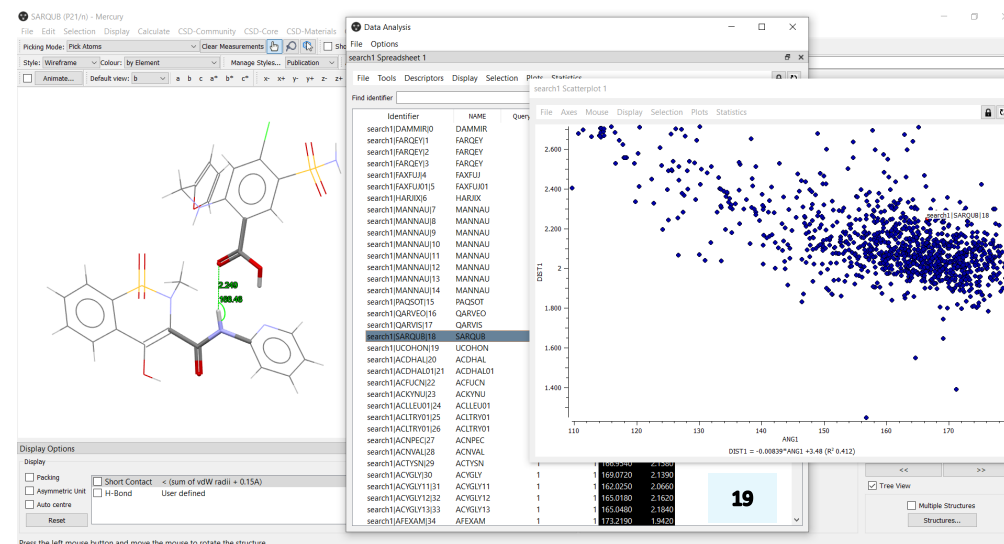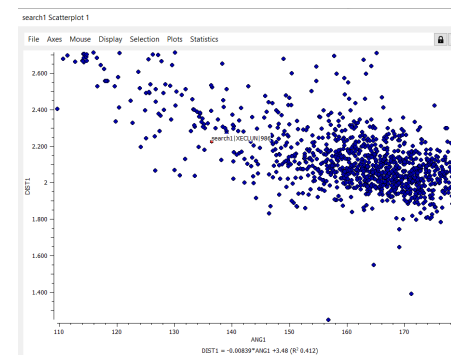
21. Find XECLUN in the list of refcodes in the **Data Analysis** window and click the line to highlight it. Note the values for the H---O distance (2.223Å) and the N-H---O angle (136.515°).

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---O bond distance in XECLUN is 0.095 Å longer and the N-H---O angle is 24.067° 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 at least 820 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 longer H---O distances. However, the shortest N-H distances fall in the middle of the range of H---O distances but correlate with larger N-H---O angles.
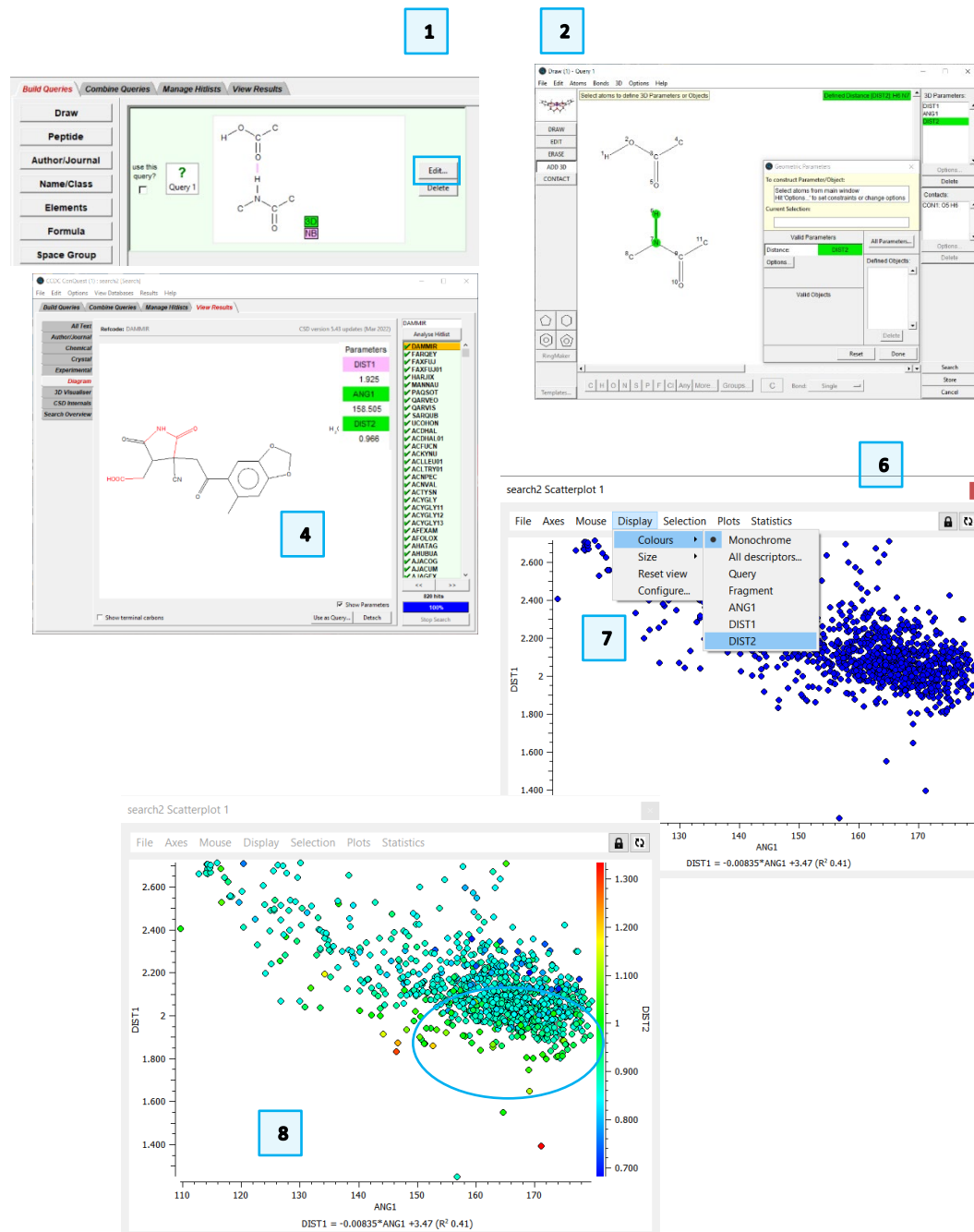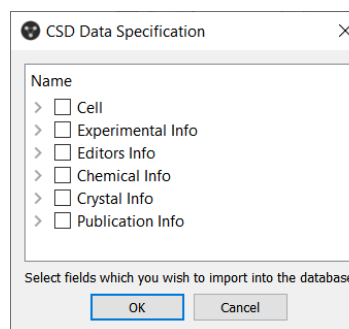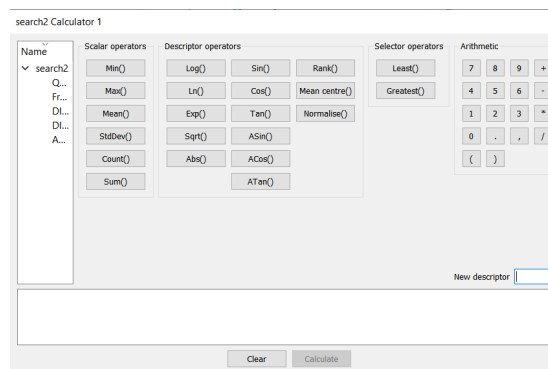
Conclusions

Now you have seen how to add 3D parameters to your query to mine the CSD for even more structural information. This workshop has demonstrated a basic search using distance and angle information. You can also define 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.
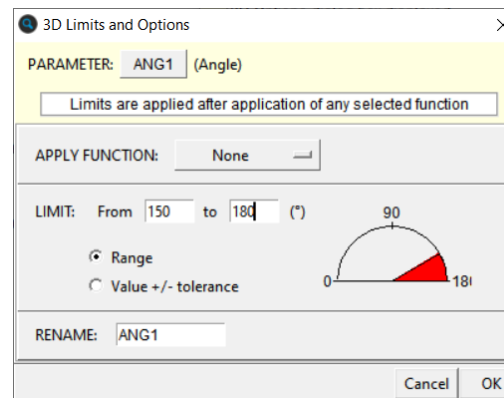
Conclusions

This workshop introduced how to define intra and intermolecular interaction in ConQuest. You should now be familiar with:

- Defining 3D parameters in your search.
- Searching for specific interactions.
- Exporting and visualizing structures in Mercury.
- Data analysis in Mercury.

Next steps

Advanced ConQuest workshops can be found [here](https://www.ccdc.cam.ac.uk/Community/educationalresources/workshop-materials/csd-core-workshops/) (https://www.ccdc.cam.ac.uk/Community/educationalresources/workshop-materials/csd-core-workshops/). 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-003. Thank you!

Glossary

Nonbonded contact – This type of search can be used for finding intermolecular interactions such as hydrogen bonds, halogen---halogen interactions, intramolecular non-covalent interactions and generating tables of geometries for nonbonded interactions. Nonbonded contacts between atoms are defined in terms of distance limits.

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.
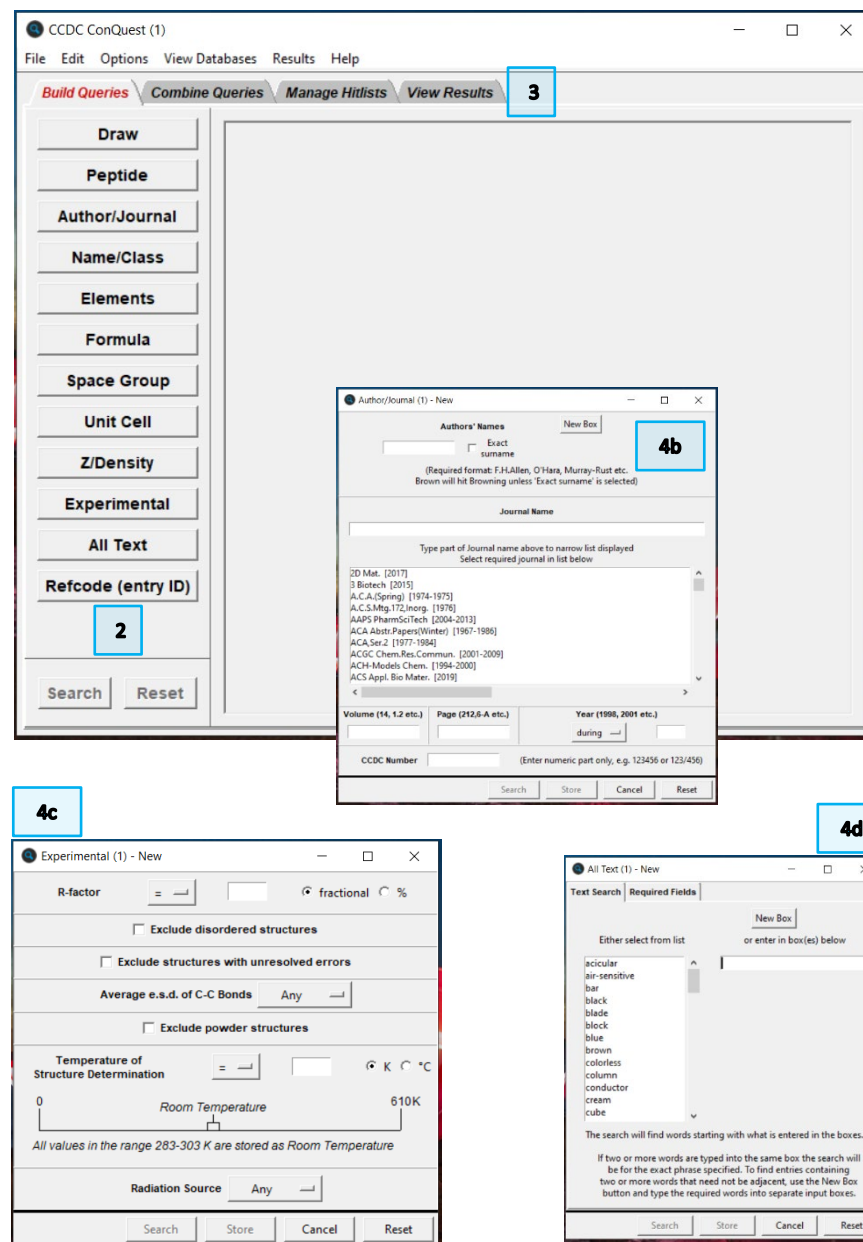
Review. Draw Window

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).

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. adamantane

Choice of specific or general atom types/functional groups

Choice of bond types