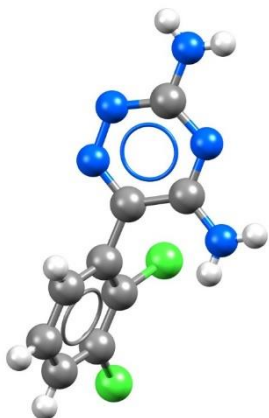# Searching and Analysing Metal-Organic Structures using ConQuest and Mercury

**Developed using
2020.3.0 CSD Release
+ Feb21 & May21 Data Updates**

CCDC
advancing structural science

Table of Contents

Introduction

This tutorial will introduce you to advanced searching in ConQuest focused on *metal-organic complexes*, *metal-organic frameworks* subset and data analysis in Mercury.

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

Before beginning this workshop, 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

In this workshop, you will learn how to:

- Conduct searches of metal-organic complexes.
- Locate and load subsets.
- Add 3D parameters in your ConQuest search.
- Use the data analysis function in Mercury.

This workshop will take approximately **25** minutes to be completed.

Note: The words in *Blue Italic* in the text 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

The following exercises assume you have a working knowledge of the programs ConQuest and Mercury, namely, how to conduct basic structural and textual search in ConQuest and how to display and manipulate structures in Mercury.
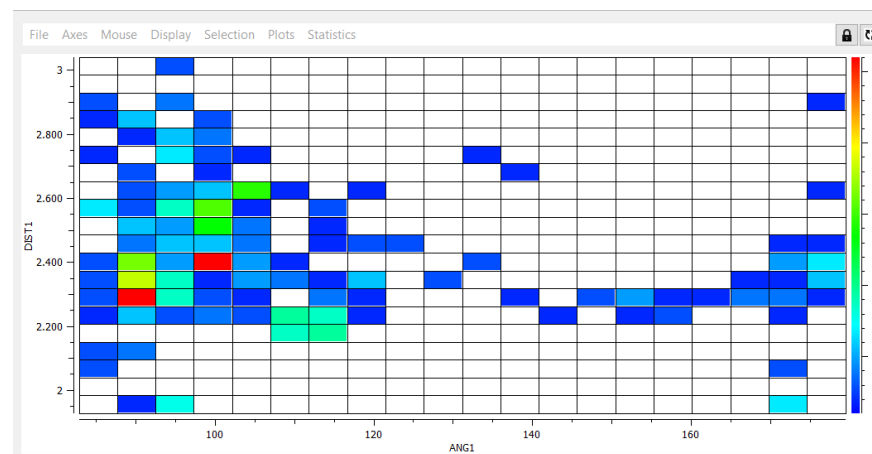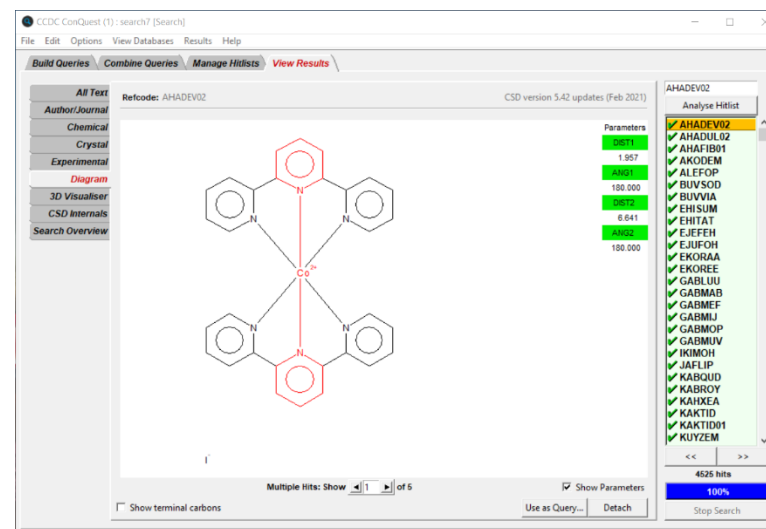
Example 1. Metal-Organic Complex Searching and Data Analysis.

Searching for structures in ConQuest is typically conducted by sketching a substructure and searching for hits in the database. But what if we have a large structure like a *metal-organic framework* (MOF) or an *organometallic* complex that is too complicated to sketch? How might we approach such a search? Apart from being able to build a search that returns relevant hits, what sort of insights can we gain about metal-organic structures in the CSD that might assist in our research? In this workshop, we will learn how to build complex queries for *Metal-organic complexes* by editing a MOF structure, how to use the MOF subset in our search to refine the hits returned and probe substituent effects on metal-organic structures using the data analysis module in Mercury.
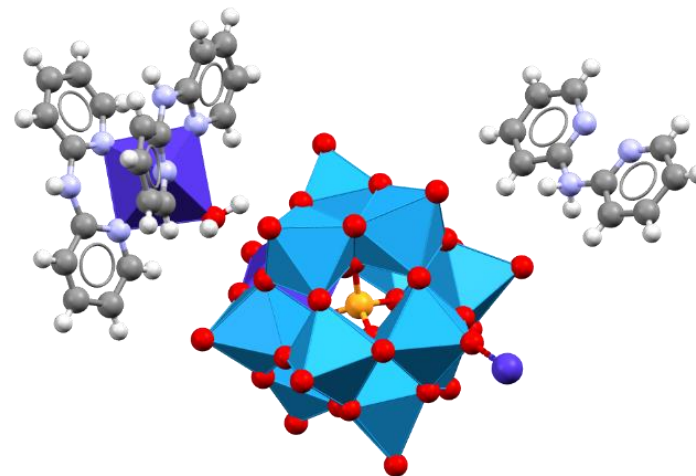
1. Launch ConQuest by clicking the ConQuest Icon 🔍 on your desktop or launching it from the Start or Applications menu. In the main ConQuest window, click on **View Databases** then click on *Lists in CSD version 5.42 (November 2020)*. A list of available subsets will be displayed, click on MOF subset to load the subset.

2. It will take a couple of seconds to load and there should be at least 109,839 structures in the subset. You can scroll through the hitlist to view the diagram of the entries. While the diagram representation can be useful for sketching a substructure for future searches, the 3D representation is helpful for viewing the full structure especially when the structure is too large to depict as a diagram. Click on **3D Visualiser** to switch to the 3D view.

3. Right-click in the 3D window to change the Display Style to your desired style. Capped Sticks style is shown in figure 3. From the hitlist, select refcode ABOHOO. There are 3 components to this structure: a tungsten-cobalt cluster, a Cobalt-*dipyridylamine* complex and a dipyridylammonium *ligand*. We will use this structure to build our query. Return to the diagram view.

4. Back in the diagram view, you can see the three components of this structure. We are interested in metal-dipyridylamine complexes and how various substituents, for example halogens on the metal can affect distances and

angles. Click on **Use as Query…** in the bottom right corner. The Use-as-Query Options window will open. Although we want the cobalt-dipyridylamine complex as our query, this is not the biggest chemical unit, therefore we are going to select *Make each chemical unit a separate query*. Untick the *Include hydrogen atoms* box then click OK to proceed to the Build Queries tab.

5. There should be three separate queries in the **Build Queries** tab. If we were interested in cobalt-dipyridylamine structures, we could proceed directly to searching, but we want to broaden the search to include other first row transition metals and add some 3D parameters, for which we need to edit the substructure. Click **Edit**… next to *Query 2*.

6. We want to replace the Cobalt atom with any first row transition metal and reduce the number of *dipyridyl*amine groups. To select the elements, click on **More…** > **Other Elements** to bring up the periodic table. Click on **1R** and the selected atoms will be highlighted in green, then click **OK**.
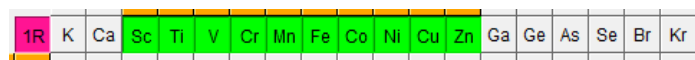
7. In the Draw window, click on the Co atom and it will be replaced with 1R. Right-click on 1R, then click on *Charge* and select *Unspecified* to remove the +2 charge.

8. We are going to remove one of the dipyridylamine groups and replace the Oxygen atoms with any halide and any atom. Click on **Erase** and remove one of the *dipyridylamine* groups, by clicking on all atoms belonging to the group. Go back to drawing mode clicking on **Draw**. Click on **More…** and select **Any Halogen**. Click one of the Oxygen atoms to replace it with 7A. Select **Any** from the bottom row of elements then click on the remaining Oxygen atom to replace it with X.
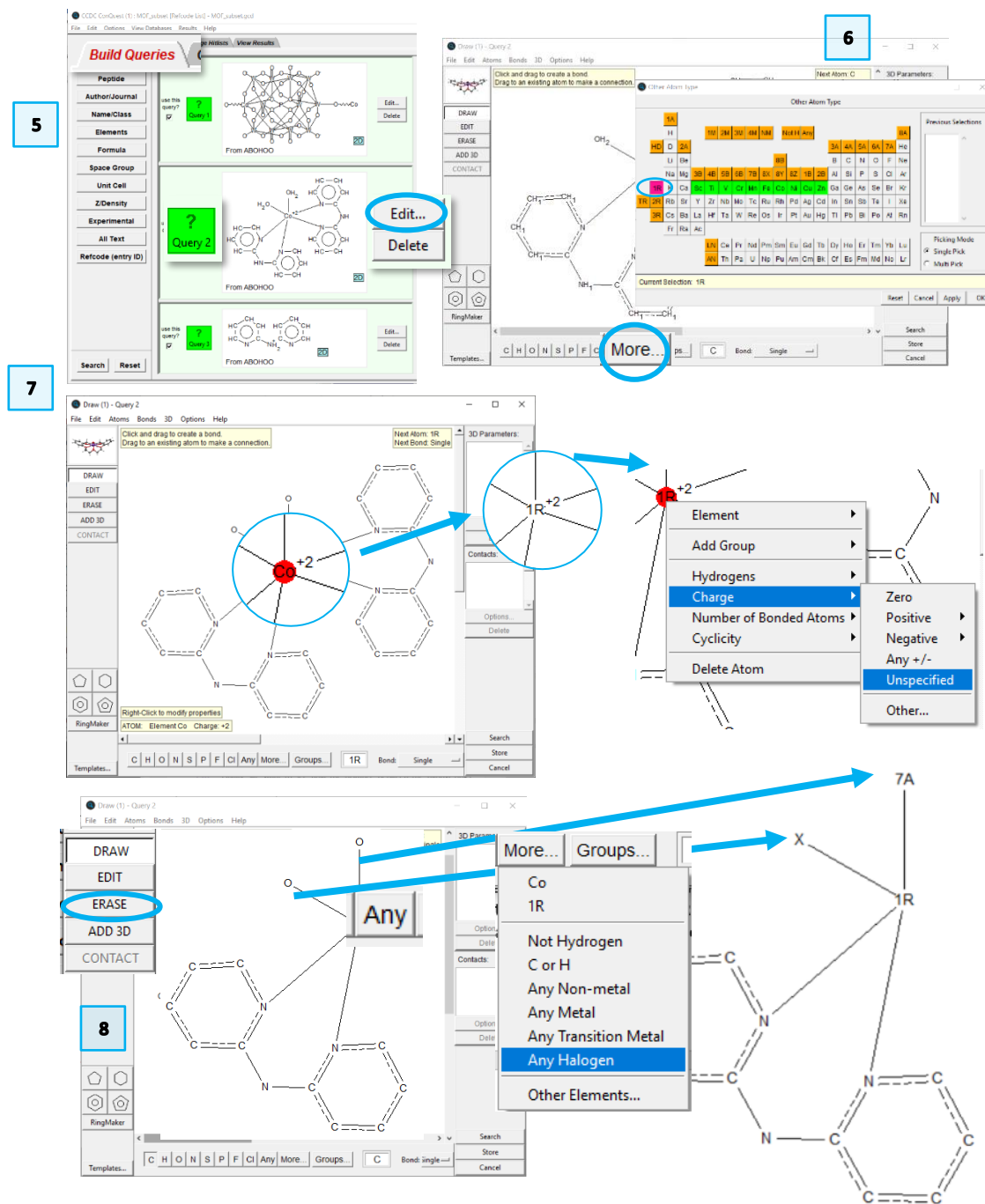
9. We will now define the 3D parameters of interest. Click on **ADD 3D** to open the *Geometric Parameters* window. First, let us define our metal-halide and metal-nitrogen distances. Click on **1R** and **7A** then select **Define** next to distance. You should now see DIST1 in the 3D Parameters box. Repeat the previous step but select **1R** and **N2** instead. To define an angle, select **7A**, **1R** and **N2** and select **Define** next to angle.

10. Finally, we want to see how the distance between the dipyridylamine rings changes. We can do this by first defining centroids. Select every atom in the ring starting with N2, then under the valid objects section, click **Define** next to *Centroid*. Repeat this step on the second ring to define the second centroid. In the *Defined Objects* box select *CENT1* and *CENT2* and click **Define** next to distance. Click **Done** to close the *Geometric Parameters* menu.

11. There should be 4 parameters in the 3D Parameters box: DIST1, DIST2, ANG1, and DIST3. You can add other parameters of interest or continue. Click **Search** and select no in the *Overwrite Query?* window. This will bring up the *Search Setup* window.

12. We want to search the entire database for structures that match our sketched substructure. In the *Search Setup* window, make sure to tick the boxes for "3D coordinates determined". We could specify "Only Organometallic" but since we have included first row transition metals in our substructure query, only organometallic structures will be returned. Click **Start Search** to begin the search.
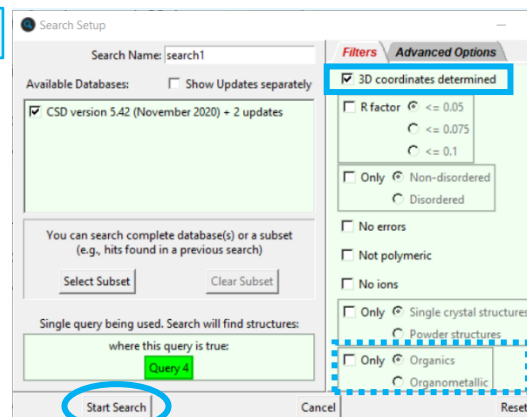

13. There should be at least 155 structures returned. Note that the values for the distance and angle parameters you defined are shown in the upper right corner of the *View Results* tab. You can use the arrow keys to scroll through the refcodes to investigate the hits returned by this query.

14. 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.
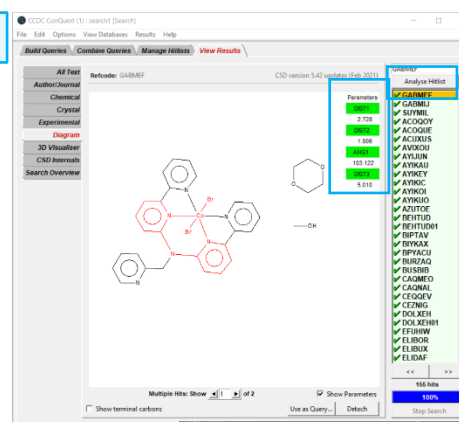

15. If you are interested in analysing other parameters, you can tick them off in the dialogue box that appears. We will skip this step for now. Simply click **Analyse in Mercury.**

16. 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.
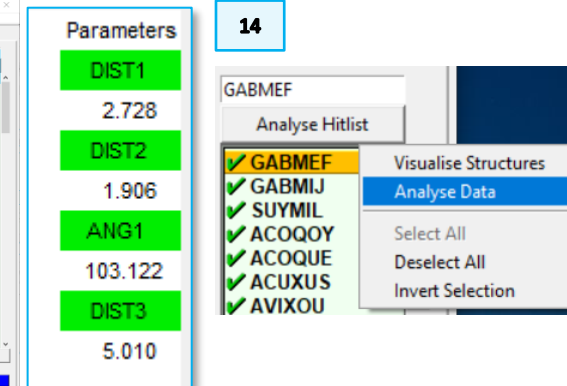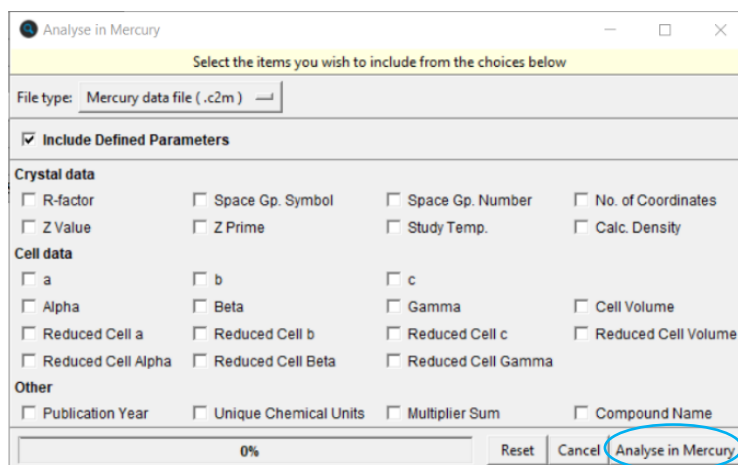
Data Analysis in Mercury.

Now that you have performed the search and obtained a set of distance and angle parameters, we can use the data analysis module in Mercury to visualize the values.

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

2. To view the distribution of angles in this hitlist, click the ***ANG1*** header in the column to highlight the column, then click *Plots* from the top menu and then select *Histogram.*

3. The angles seem to cluster in two regions, one centred around 100 degrees and a smaller one centred around 170 degrees. Click on the bar of the histogram around 170 to highlight the corresponding entries in the spreadsheet.

4. It will be easier to scroll through the refcodes in that region by sorting the spreadsheet by ANG1. Right-click on the **ANG1** heading and click *Sort*.

5. You should now see the highlighted refcodes in the spreadsheet that correspond to the angles clustered around 170 degrees. Clicking on each entry in the spreadsheet will place the corresponding structure in the Mercury display.

6. As you scroll through the highlighted refcodes, can you observe a trend in the structures clustered here? The metal centre and halide changes but most of the structures have a metal-oxygen bond (ROHCUN shown here as an example in *polyhedral display*). Can you generate other plots, for example, a scatterplot of DIST1 and ANG1? Are there any insights there?
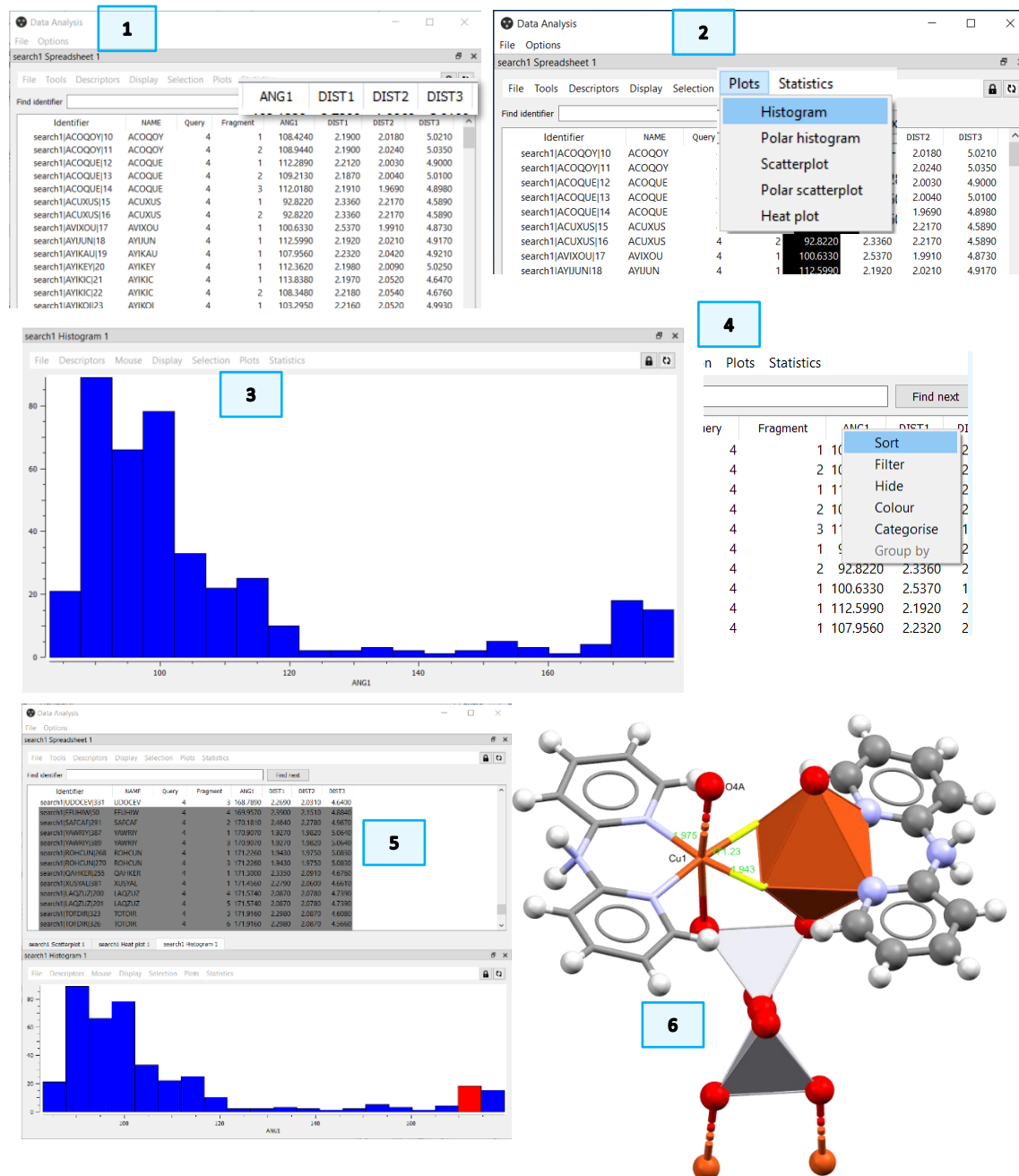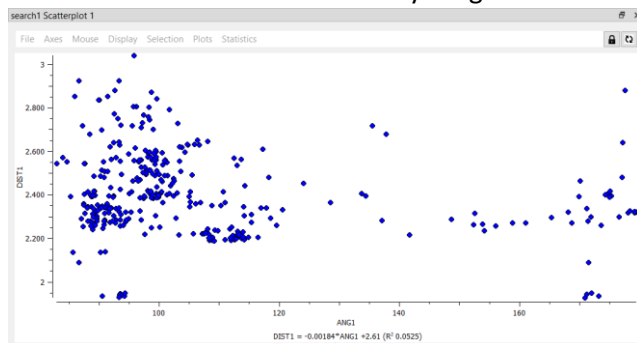
Example 2. Searching within the MOF subset

In the previous example, we chose a structure from the MOF subset, edited to a substructure that was then searched. How can we search for substructures or specific elements of interest that is a MOF? We will go over ways to search for MOF structures in the CSD.

1. Return to the main ConQuest window. If you have previously closed ConQuest, reload the MOF subset. We will first conduct a name search that should return most of the structures in the MOF subset.

2. Click on the **Build Queries** tab and select **Name/Class**. In the *Compound Name* box, type *catena* click **Add** and then click **Search**.

3. In the *Search Setup* window, make sure to tick the boxes for "3D coordinates determined" and specify "Only Organometallic" and click **Start Search**. At least 119813 structures should be returned. This is higher than the structures in the MOF subset.

4. Since catena is not specific enough, we could focus instead on a specific metal and oxidation state.

5. Return to the **Build Queries** tab and select **Name/Class**. In the *Compound Name* box, type (ii). It is important to include the parenthesis to capture complexes with exactly that oxidation state. Click **Add** then click **Store**.

6. In the **Build Queries** tab, click **Elements**. In the new window, click on the **Select from Table** box then Click 2R and select *Done*, leave the other default options and click **Store**.

7. You should be returned to the **Build Queries** tab. Make sure the last 2 queries, i.e., the Name/Class and the Elements Query, are selected and no other queries are, then click **Search** at the bottom of the *Build Queries* tab.

8. We want to limit this search just to the MOF subset. In the *Search Setup* window, click **Select Subset**. In the *Restrict Search* window, select "*Entries in*

a pre-defined hitlist" and then select the MOF subset. Once the subset is loaded, click **OK** to return to the *Search Setup* window.

9. You should now see that your search will be restricted to the structures within the MOF subset. Make sure to tick the box for "3D coordinates determined". Although part of our search criteria is second row transition metals, we could specify "Only Organometallic" and click **Start Search**.

10. At least 7380 structures should be returned. In the **View Results** window, click on **Chemical**. You should see the name match *(ii)* highlighted in yellow. You can further refine you search by searching for a specific element or a different oxidation state.

Further Exercises

- Return to ConQuest and edit Query 4 by replacing the X (any group) with Oxygen and choose a specific first row transition metal, Cu for example. Add the metal-oxygen distance as a 3 D parameter and conduct a new search.

- Can you export your results to Mercury and use the data analysis feature to generate descriptive statistic to gain more insight on your data?

Conclusions

This workshop introduced how to search for metal-organic complexes in ConQuest and analyse data using Mercury.

In the two examples 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 way to search using distance and angle information and how to use a previously found structure to build your query. You can also define dummy points and use atom characteristics such as atomic number and van der Waals radius in searches.

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.

You should now be familiar with:
- Editing a structure to use in a query.
- Loading pre-defined subsets.
- Adding 3D parameters in your search.
- Restricting search to a subset.
- Conducting data analysis in Mercury.

Next steps

After this workshop, you can explore more exercises in the self-guided workshops available in the CSD-Core (ConQuest) and CSD-Materials (Mercury) workshops area on our website.

https://www.ccdc.cam.ac.uk/Community/educationalresources/workshop-materials/

Please see the ConQuest documentation for more details.
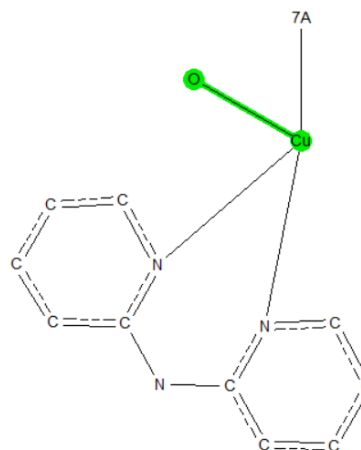
Images representative of the further exercises suggested in example 2.

Glossary

Catena

Polymeric organometallic structures in the CSD (purely organic polymers are prefixed by Poly instead)

Centroid

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

Dipyridylamine

Two aromatic rings with one of the carbon atoms replaced with a Nitrogen atom in each ring and the rings are connected by a secondary amine group.

Ligand

A Ligand is an ion or molecule that bonds to a central metal atom.

Metal-Organic Complex

A chemical compound that contains a metal and organic ligand.

Metal-Organic Frameworks (MOFs)

A MOF, or metal organic framework, is a material formed of metal clusters or nodes, linked by organic molecules.

Only Organometallic

Selecting this filter restricts the search to structures containing at least one transition metal, lanthanide, actinide, or any of Al, Ga, In, Tl, Ge, Sn, Pb, Sb, Bi, Po.

Pi bonds

Pi-bonds in the CSD refer to bonds between a metal and a ligand. For a cyclopentadienyl ring with no substituents or just acyclic substituents, a pi-bond is recorded between the metal atom and each of the ring C atoms.
Polyhedral display

Polyhedral display

Visualise metal-organic structures using polyhedral representations for the metal-organic coordination centres.

Figure of a metal-organic complex, BEHTUD01 (Dibromo-(di-2-pyridylamine-N,N')-copper(ii)), with the dipyridylamine group highlighted by the box and the arrow points to a defined centroid on the right aromatic ring.
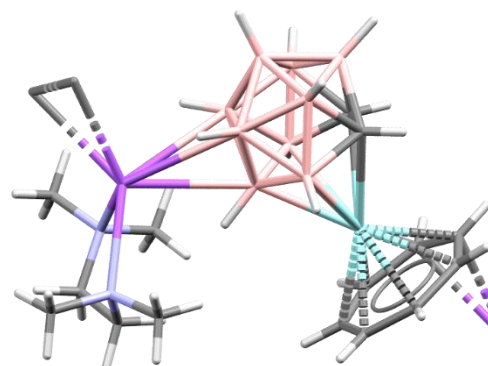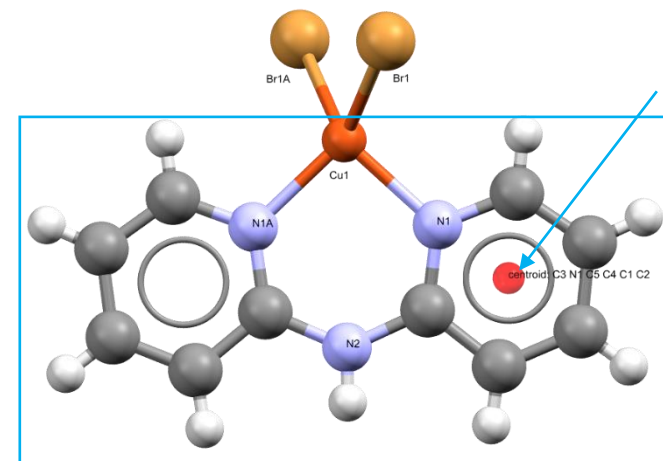

Figure of a MOF, ABUREX.
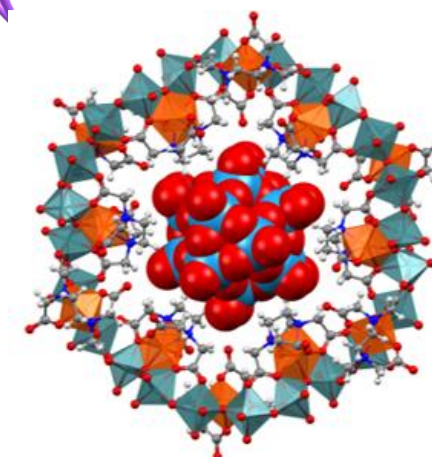
Polyhedral display

Central Elements - atoms around which polyhedra are centred.

Ligand Elements - neighbouring atoms which create the corners of the polyhedron.

Polymer Expansion by

- **Whole unit** expansion will add an additional repeat of the initial crystal chemical unit to the structure. All polymeric structures in the CSD will include a full coordination sphere of all symmetry-independent atoms of the structure.

- **Sub unit** expansion adds all atoms and bonds up to the next polymeric atom or bond in the structure.

Example of Polymer Exansion
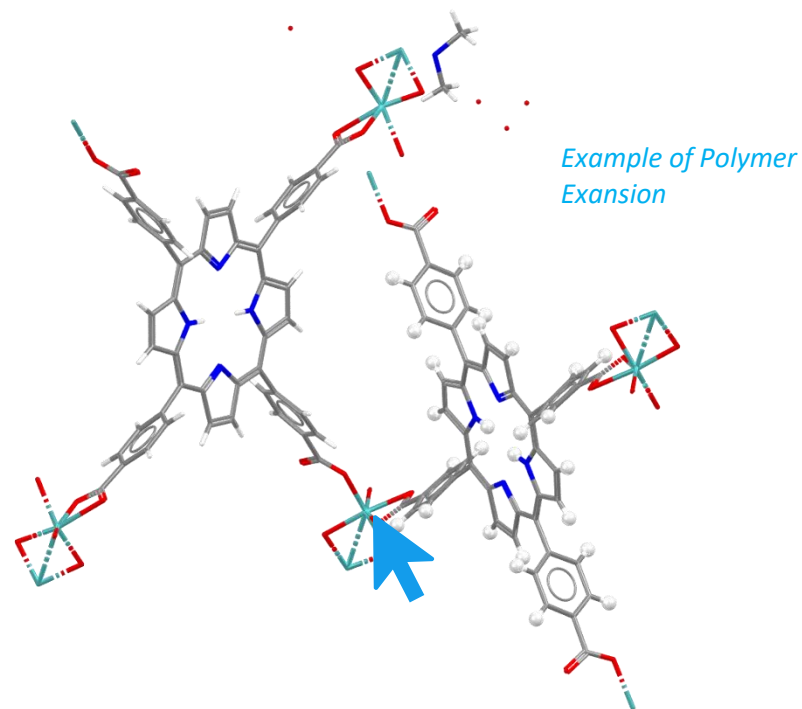
Review. ConQuest Interface

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
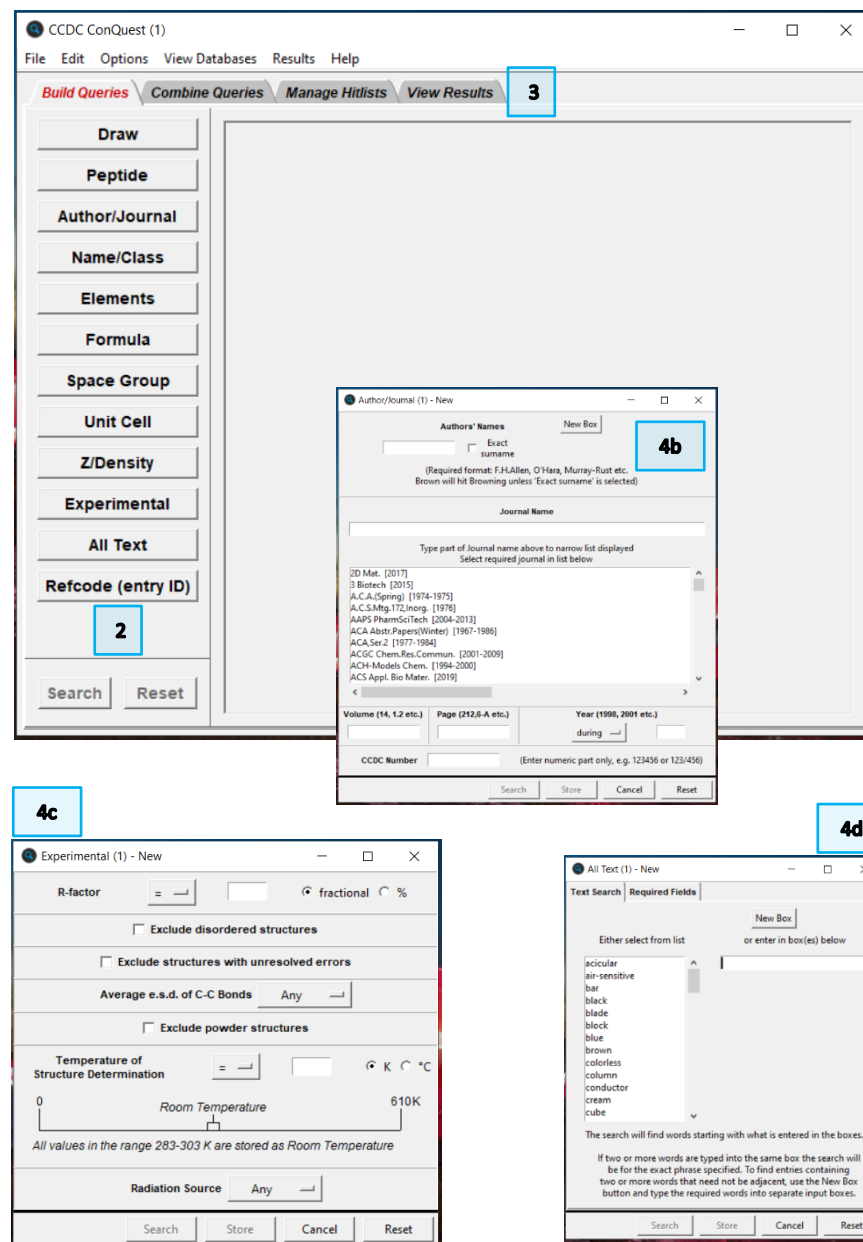   a. Draw – substructure and 3D information searching
    b. Author/Journal – bibliographic searching
    c. Experimental – experimental set up searching
    d. All Text – generic text-based searching
5. The majority of the searching we will do in these tutorials will be substructure searching, so we will 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