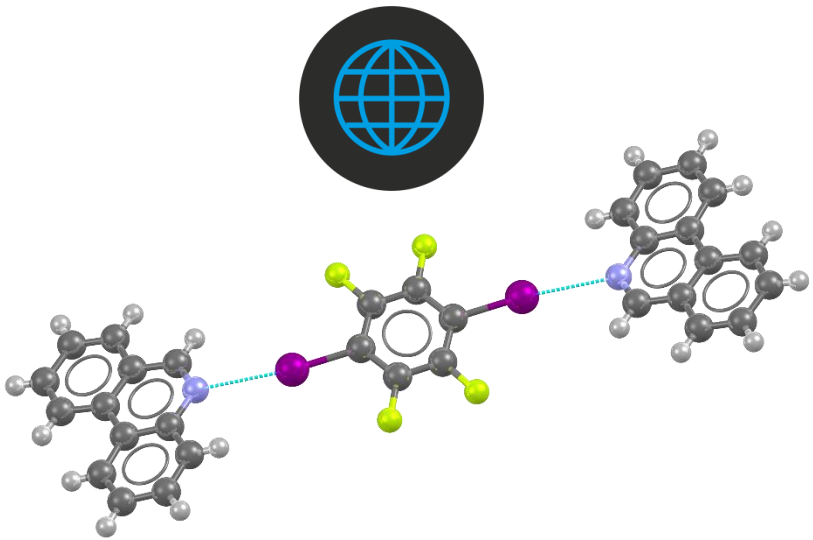
Searching the CSD online with WebCSD (WCSD-002)

Developed using WebCSD version 1.9.32 (Feb 2025)



CCDC
advancing structural science

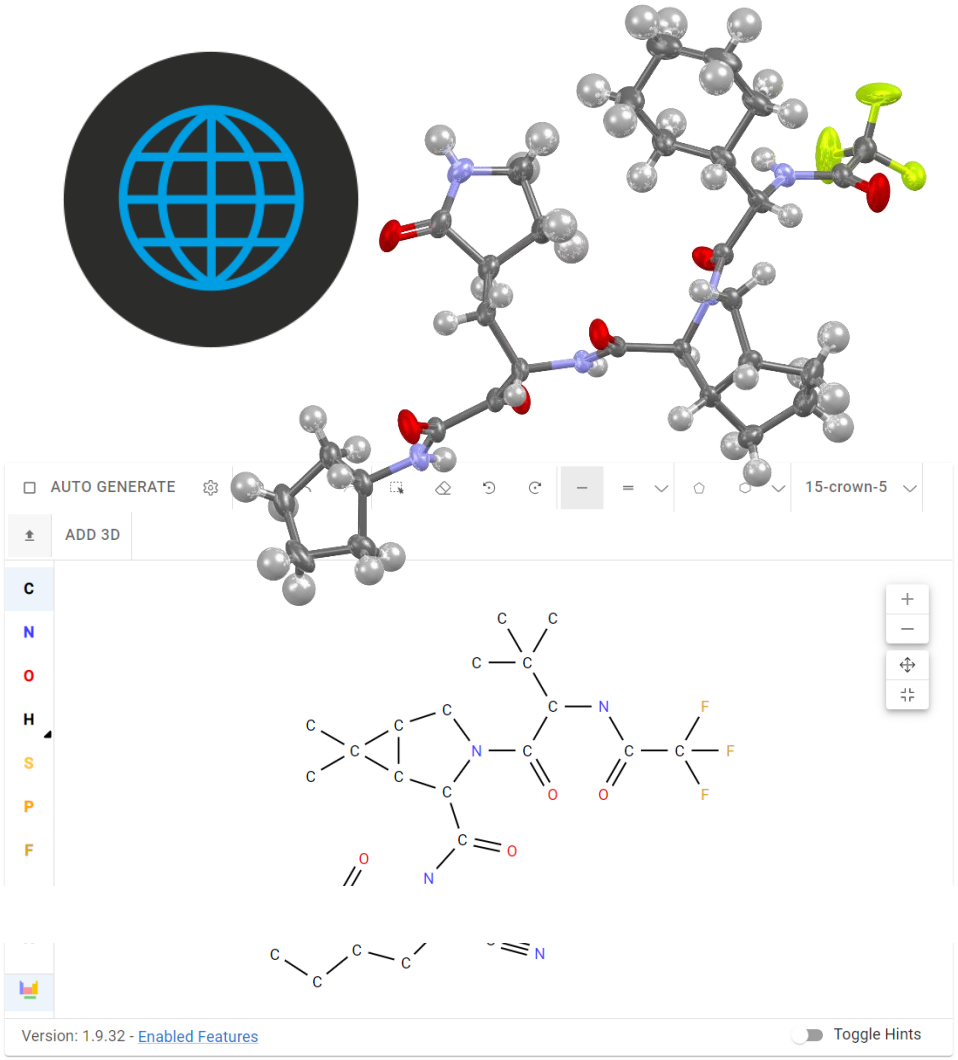
Table of Contents

Introduction

This workshop will introduce you to advanced searching using WebCSD.

WebCSD allows you to search for a specific structure and to find the context of the structure in relation to its originality and/or comparison to other known molecules. WebCSD supports text/numeric, structure, unit cell, and formula searching across the CSD, with the benefit of returning the most up-to-date possible results as structures are added to the CSD. Original deposited data can be obtained and exported for further analysis.

Before beginning this workshop, please ensure that you are signed in to WebCSD with a user account connected to a valid CSD license (see here or this FAQ).



Learning Outcomes

In this workshop you will explore the tools available for searching the CSD online using WebCSD. After completing this workshop, you will be able to:

- Create substructure searches and include defined 2D and 3D parameters.
- Search for compounds based on their chemical formula.
- Use the unit cell search function to compare structures.

This workshop will take approximately ***45*** minutes to complete. The words in *Blue Italic* in the text are reported in the Glossary at the end of this handout.

Pre-required Skills

There are no pre-required skills for this workshop.

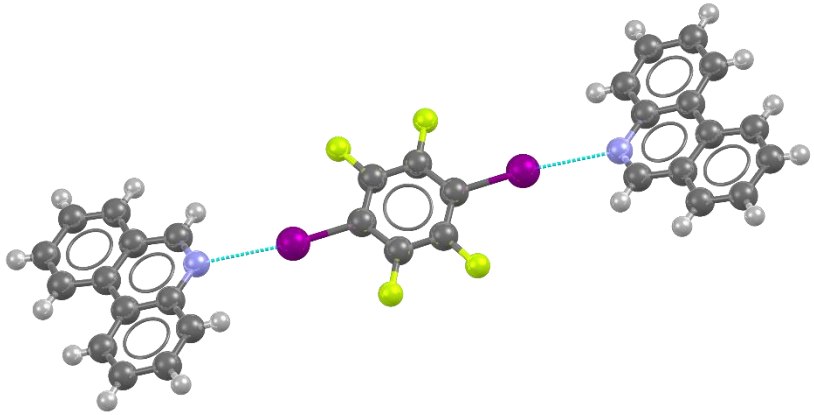
Materials

No additional materials are required for this workshop.

Example 1. Investigating N···I interactions in the CSD using a substructure search.

The CSD can be used to help us understand what non-bonded contacts exist and what geometrical features these contacts have in the solid form. In this use case we will be using WebCSD to explore a particular type of halogen bond to help us learn about how strong the interaction is and what geometrical features are observed in these interactions.

A *halogen bond* occurs when there is evidence of a net attractive interaction between an electrophilic region associated with a halogen atom in a molecular entity and a nucleophilic region in another, or the same, molecular entity.[1] We will extract geometrical information about *intermolecular contact* distances and angles associated with halogen bonds in the solid state, using a *substructure* search of the CSD. Specifically, we will search for N···I interactions taking R—I to be the halogen bond donor, where R is an aromatic group, and taking sp^2 hybridised nitrogen to be the halogen bond acceptor.



CSD entry TOKFEG displaying halogen bonds between the iodine atoms of 2,3,5,6-tetrafluoro-1,4-diiodobenzene and the nitrogen atoms of the phenanthridine molecules.

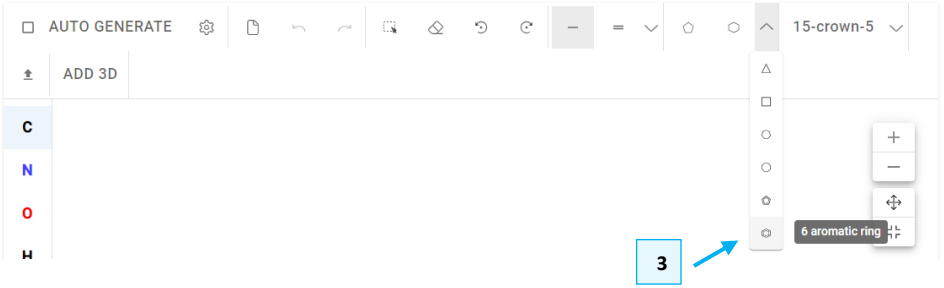
We will begin by defining our donor and acceptor groups.

1. Open a web browser and navigate to the Access Structures webpage (https://www.ccdc.cam.ac.uk/structures/).

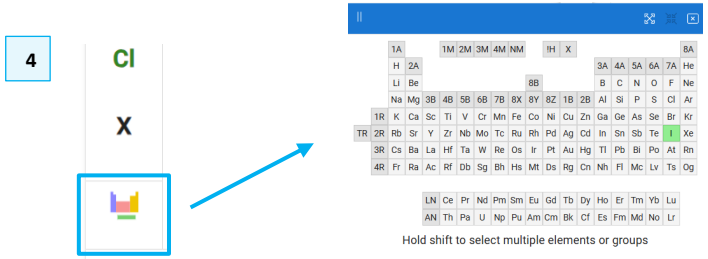
2. Click on the **Structure Search** tab to bring up the sketcher. If the periodic table is visible, you can drag it to reposition it in a convenient area or close it for the time being.

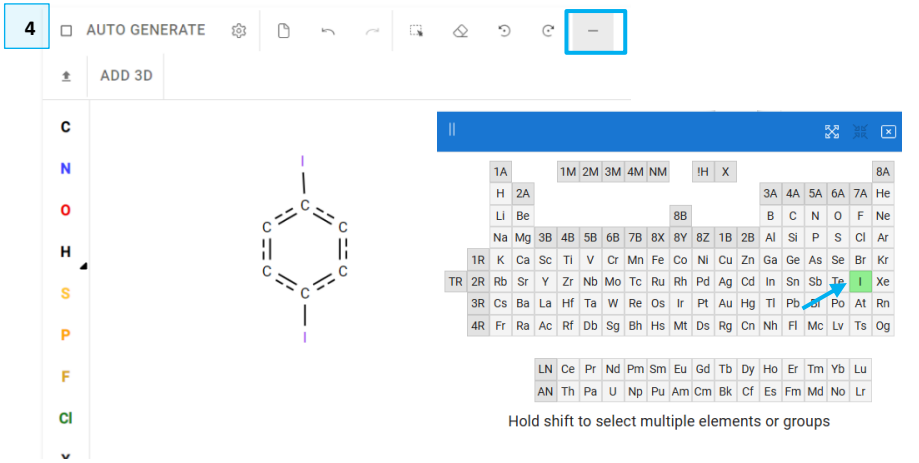
[1] G. R. Desiraju, P. S. Ho, L. Kloo, A. C. Legon, R. Marquardt, P. Metrangolo, P. Politzer, G. Resnati and K. Rissanen, *Pure Appl. Chem.*, 2013, **85**, 1711-1713.

3. From the drop-down menu next to the cyclohexane ring in the tools bar, select the *6 aromatic ring*  then **click** in the sketcher area to add an aromatic group.

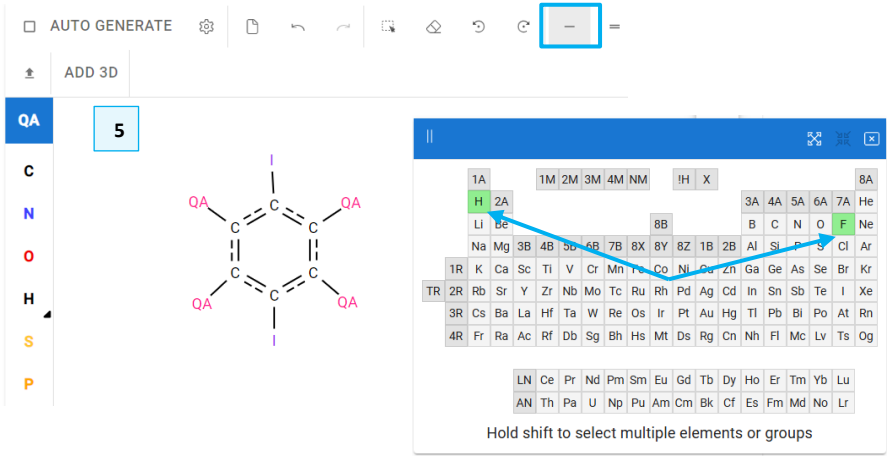


4. As the halogen bond donor molecule, we will use 1,4-diiodobenzene and fluorinated analogues, $C_6H_nF_{4-n}I_2$. If the periodic table is not visible, click the icon in the left-hand side toolbar. Click on the **single bond** symbol in the toolbar, then select iodine by clicking on the **I** symbol on the periodic table. Left-click on a carbon atom and drag to add a C—I bond. Repeat this procedure for the carbon atom in the 4-position.

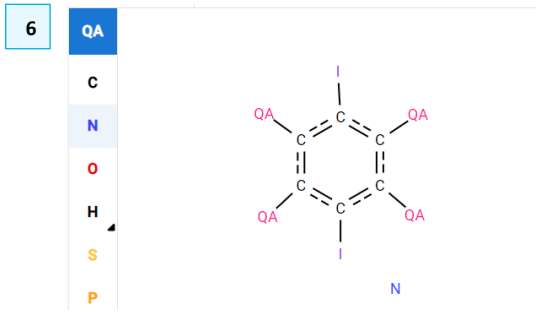




5. For the remaining substituents, we need to allow either hydrogen or fluorine. In the periodic table, click **H** then shift+left-click on **F**. Click and drag from the remaining unsubstituted carbon atoms, which will add “QA” substituents.

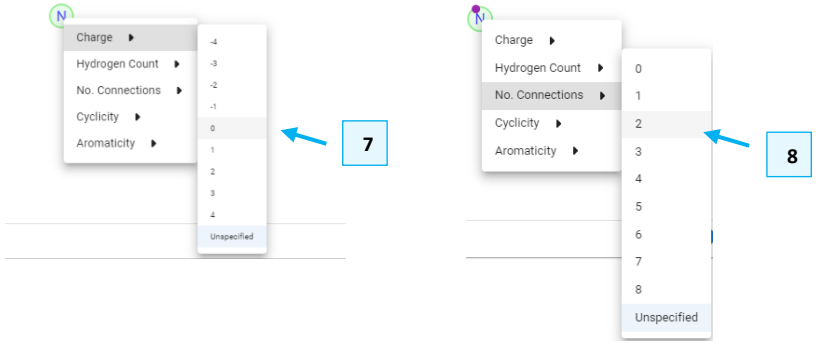


6. For the acceptor group we will add a neutral *sp*^{2} hybridised nitrogen atom. Click **N** in the left-hand toolbar and then click near to an iodine atom in the sketcher.



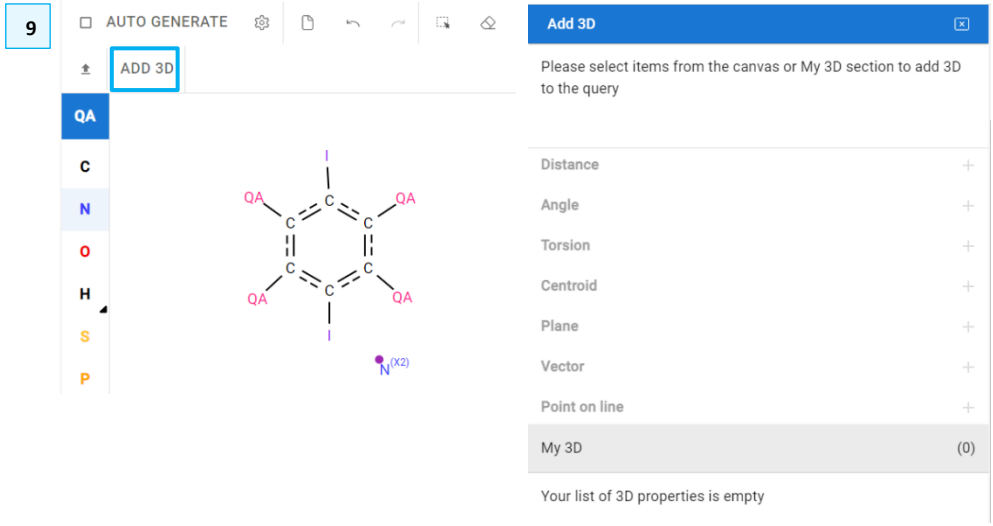
7. Right-click on **N** in the sketcher to bring up the drop-down menu. Select *Charge* and then choose *0*.

8. Right-click on **N** again and select *No. Connections* then choose *2*. The nitrogen atom should now appear as N(X2).

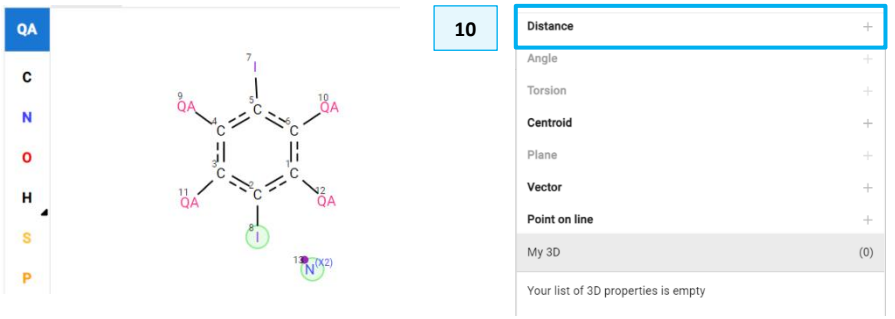


We will now define the geometric data to be extracted from the structures during the search. The three parameters that we will define are: the I···N distance, the C—I···N angle and the perpendicular distance of the N atom from the *mean plane* defined by the six carbon atoms of the aromatic ring.

9. Click the **ADD 3D** button from the top toolbar to bring up the 3D parameter options. The options will be greyed out until we have selected enough atoms to define a valid geometric parameter.

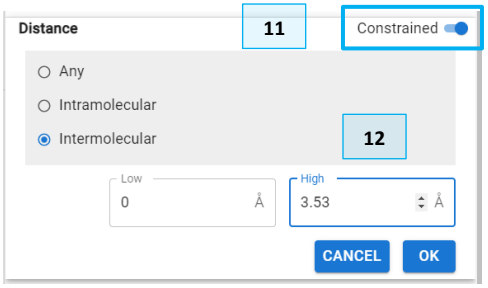


10. Shift+left-click on the **N** and **I** atoms (they will become highlighted in green); this will enable options to define a distance, centroid, vector or a point on a line. Click the **+** button next to *Distance*.

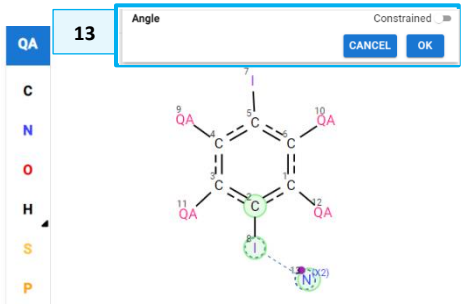


11. Move the *Constrained* slider to **on** and click the **Intermolecular** radio button.

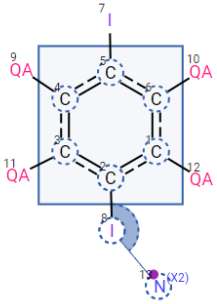
12. Leave the *low* limit as "0" and enter "3.53" in the *high* limit box. This is the sum of the *van der Waals radii* for nitrogen and iodine.[2] This will ensure only results that show definite evidence of halogen bonding are returned. Press **OK**.

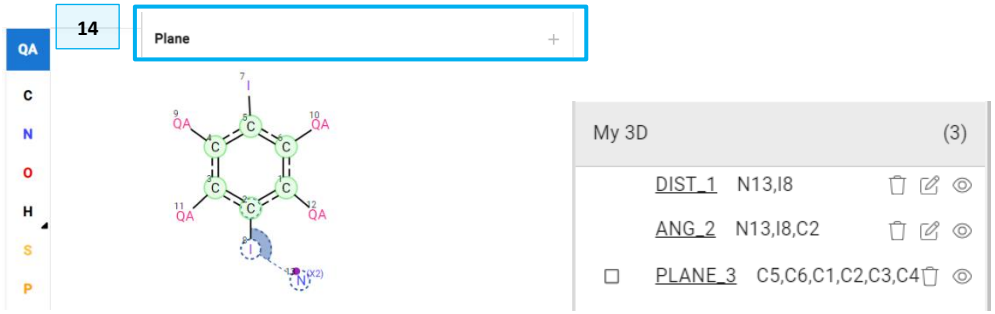


13. Shift+left-click on **N**, then **I**, then the **C** bonded to I, in that order. This will enable the option to define an angle. Click **+** next to *Angle*, then click **OK**. Ensure that the *Constrained* option is not enabled.



14. Select all six atoms of the aromatic ring using Shift+left-click, then click **+** next to *Plane*. The newly defined plane will be highlighted as a square in the sketcher window.



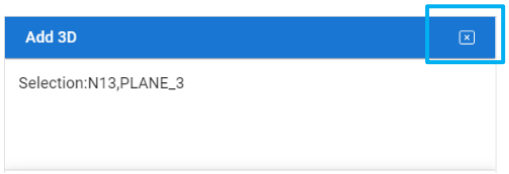


[2] M. Mantina, A. C. Chamberlain, R. Valero, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. A.*, 2009, **113**, 5806—5812.

15. Click on **N** in the sketcher and then tick the box next to *PLANE_3* in the **Add 3D** window. This will then enable the option to define the perpendicular distance between the N atom and plane of the aromatic ring.

16. Click **+** next to Distance and ensure that the *Constrained* option is NOT enabled by sliding it to the off position if necessary. Click **OK**.

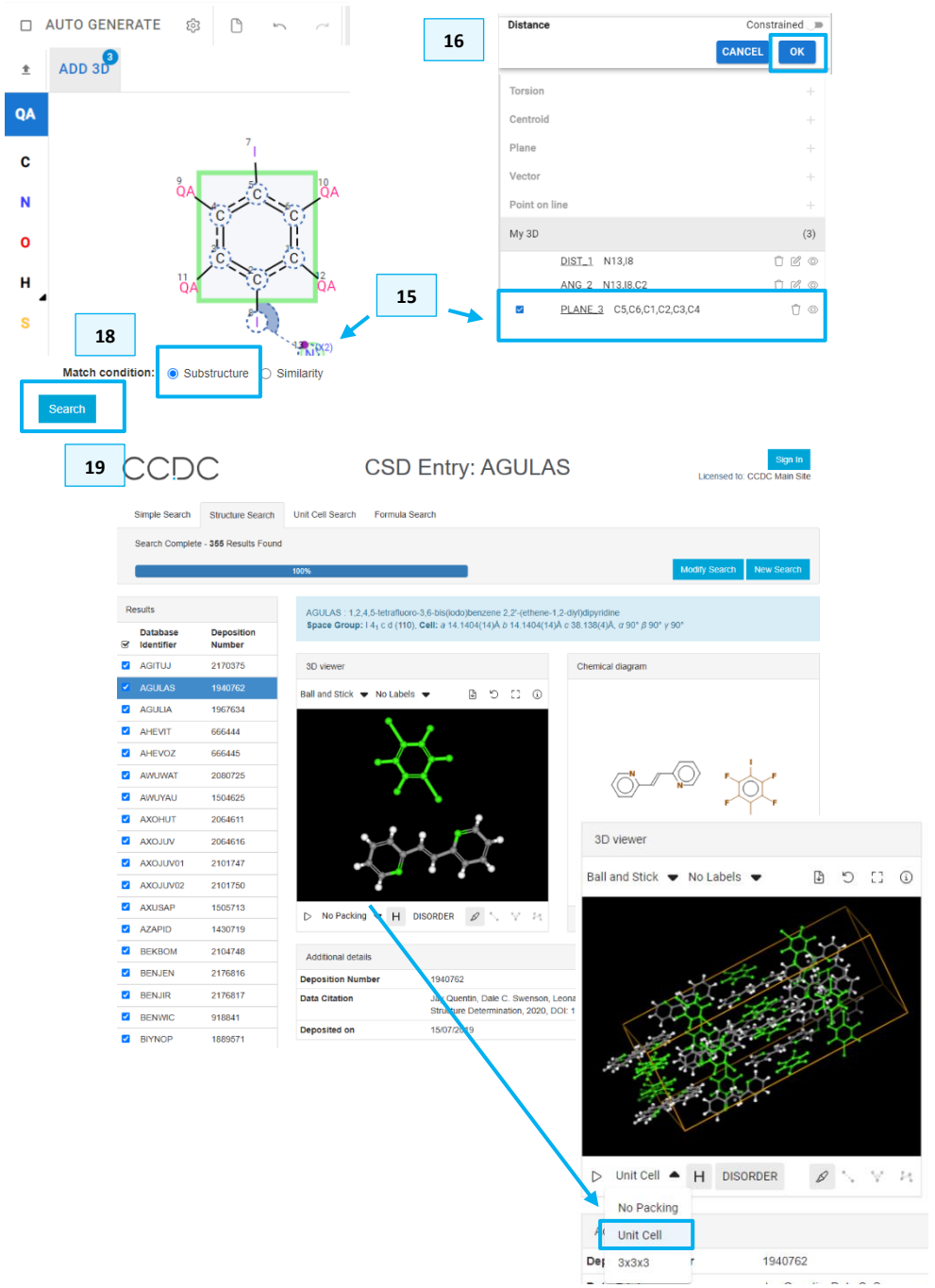
17. You should now have four defined parameters. Close the **Add 3D** window.



18. Underneath the molecule sketcher, ensure that the **Substructure** radio button is selected, then click **Search**.

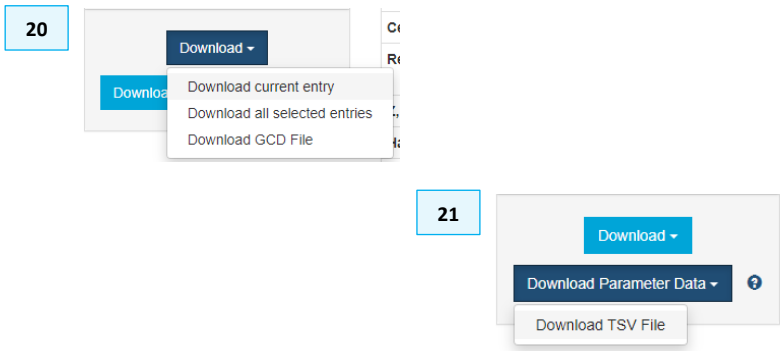
19. A new page will load showing the results, listed by refcode on the left-hand side, and information about the currently selected entry. There should be at least 355hits (please note that your value may be much larger, due to WebCSD being updated every minute). In the 3D viewer, the hit fragments are highlighted in green, while in the 2D diagram they are highlighted in brown. Scroll through the results and notice how the information and viewer are updated as you do so. As we are studying an intermolecular interaction, you may need to turn of the packing display to visualise all interactions. To do so, from the dropdown menu for the display options, select *Unit Cell.*

Tips and tricks: you can find more tips on how to get the most out of the 3D viewer here https://www.ccdc.cam.ac.uk/community/education-and-outreach/outreach/home-learning-activities/3d-visualizer-tips/



20. Click **Download** at the bottom of the results list to download entries as *CIF* files or as a *GCD* file, which allows you to load the CSD entries for the hitlist directly in Mercury. *Tip: very recent additions to the CSD may not be available through desktop applications until the next CSD data update.*

21. Click **Download Parameter Data** to download the 3D information gathered during the search as a TSV file. This can be imported into a suitable program, such as Microsoft Excel, for further investigation. *Note: 3D parameters are reported to two decimal places, and a maximum of 10 3D parameters per structure is permitted.*



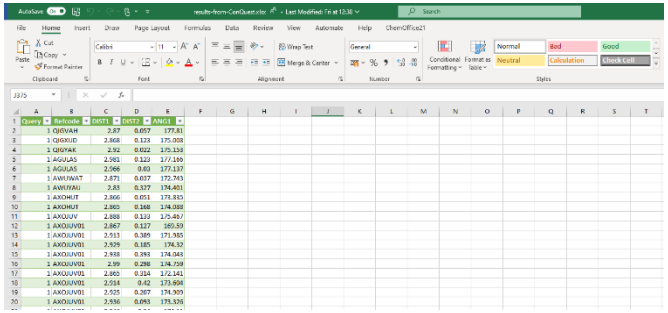
Conclusion

We have seen how to construct a substructure query using WebCSD, how to add 2D constraints to define aspects of the chemistry, and how to search for 3D data so that it can be extracted for further analysis. In this example, we have explored structural evidence for halogen bonding between aromatic iodides and non-bonded sp^2 nitrogen atoms.

Exercises

1) Import the downloaded data into a suitable program, such as Microsoft Excel. Note that because there may be multiple measurement of interactions within the same structure, the number of items in the TSV file is expected to exceed the number in the list displayed in WebCSD.

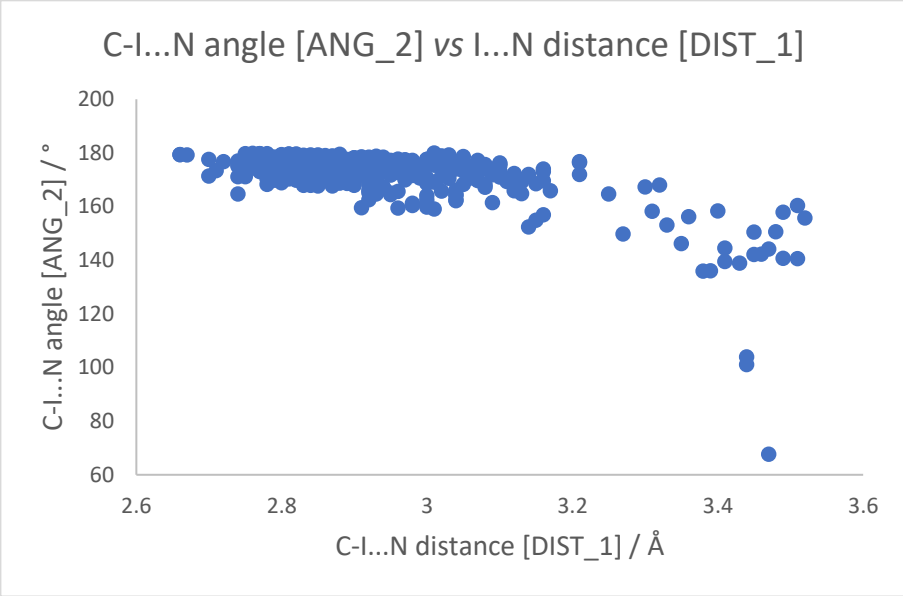
1

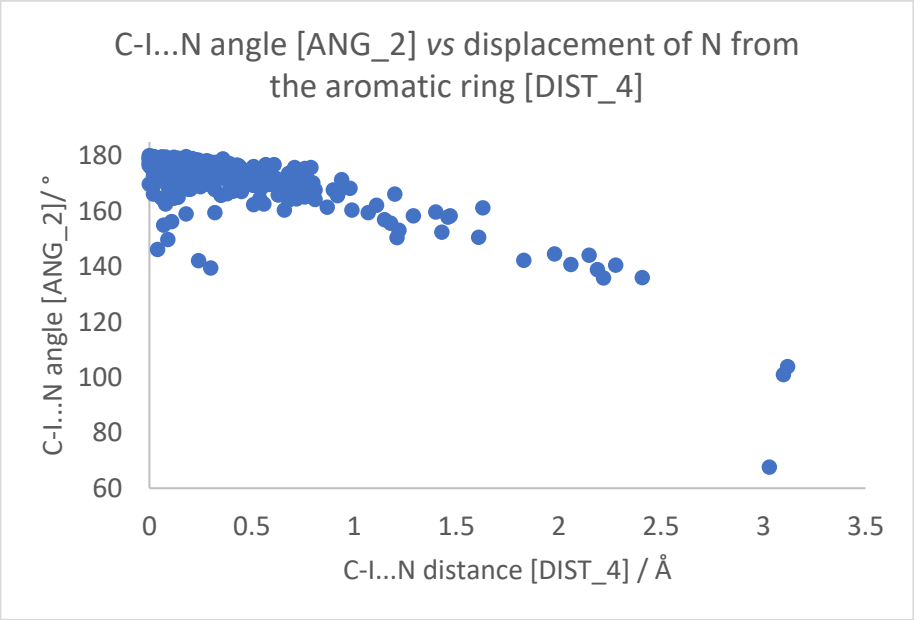


2) Investigate correlations within the data, for example by plotting the non-bonded N···I distance against the C—I···N angle. Also investigate any correlation between the displacement of N from the mean plane of the aromatic ring with the C—I···N angle. What can you conclude about the geometry of the N···I interactions?

Note: your results may look slightly different as the number of entries in the CSD accessible through WebCSD changes daily.

2

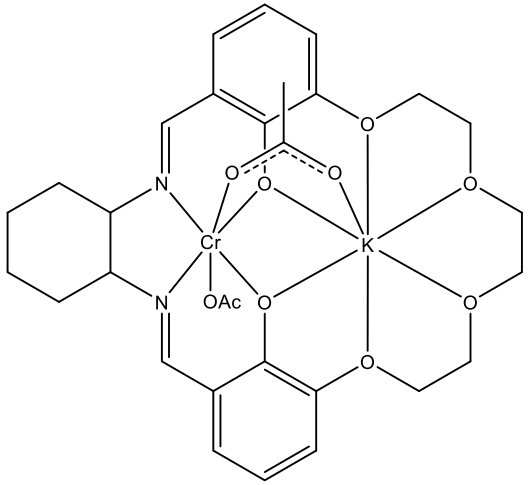




Example 2. Discovering catalysts using a formula search

A WebCSD formula search offers a quick way to find compounds with a specific chemical composition, as well as to discover related compounds containing specific elements. The latter is particularly useful for transition metal complexes where related structures may have different ligands despite containing common metal atoms or ions.

A *bimetallic* catalyst was reported to selectively deliver poly(dithiocarbonates) from CS_2 and oxetanes.[3] However, the crystal structure of the catalyst was not determined. We will search for potential structural motifs for the catalyst in the CSD based on formula.



2D structure of the polymerisation catalyst reported in reference 5.

We will search for bimetallic compounds which contain at least one chromium and one potassium atom.

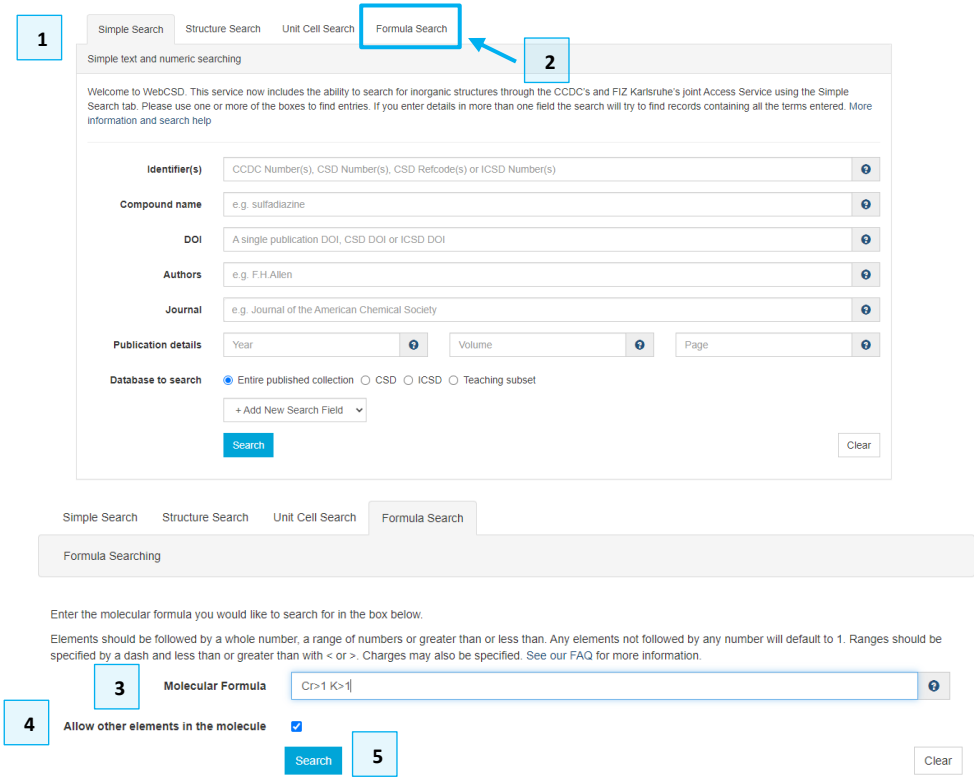
1. Open a web browser and navigate to the Access Structures webpage (https://www.ccdc.cam.ac.uk/structures/) if you do not already have WebCSD open. If you have already run a search, click **New Search** at the top.

2. Select the **Formula Search** tab.

3. In the *Molecular formula* box type "Cr>1 K>1". This specifies that there will be at least two chromium atoms and two potassium atoms in the curated molecular *formula unit*. Make sure there are no spaces between the element symbol, inequality sign and the number.

4. Tick *Allow other elements in the molecule*. This allows for any of the remaining elements that may be present in the ligand(s).

5. Click **Search**.



[3] C. Fornacon-Wood, B. R. Manjunatha, M. R. Stühler, C. Gallizioli, C. Müller, P. Pröhm and A. J. Plajer, *Nat. Commun.*, 2023, **14**, No. 4525.

6. A new page will load showing the results (the refcodes are listed on the left-hand side) and information about the currently selected entry. There should be at least 27 records returned (your value may be significantly higher as WebCSD is updated every minute). Scroll through the results and notice how the information and viewer is updated as you do so. You should notice that there are at least four entries with similar structures to the complex on the previous page.

7. You can explore these results further by following the DOI link in the *Associated publications* section. For example, CSD RIPCUS is an earlier structure reported by the same research group.

8. Click **Modify Search** or **New Search**.

9. In the *Molecular Formula* box, enter "Cr1 K1-2". This tells WebCSD to look for complexes which contain one chromium and one to two potassium atoms per formula unit.

10. Click **Search**.

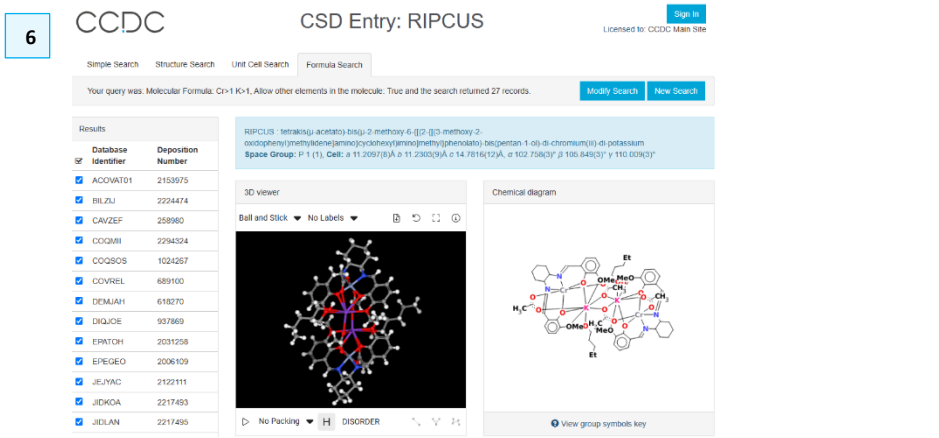
11. You should find at least 90 search records returned. Browse through these to compare them with previous results.

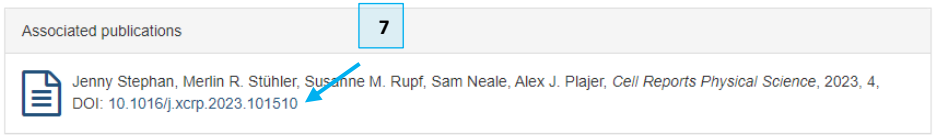
Conclusion

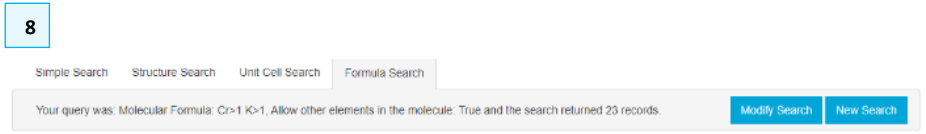
In this example, we have seen how to make use of several options for formula searching in WebCSD to uncover bimetallic complexes which may be of interest for catalytic applications by specifying elements that must be present and placing constraints on their occurrence in the molecular formula.

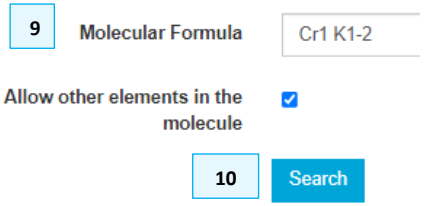
Exercise

Use the formula search to explore other bimetallic complexes where potassium is switched for other alkali metals.





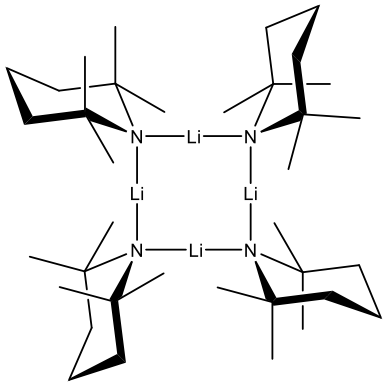




Example 3. Unit cell searching with WebCSD

In an X-ray diffraction experiment, a *unit cell* search is often a convenient way of checking whether a data set has been published before the full experiment is undertaken, potentially saving time and resources. In this example, we will see how the WebCSD Unit Cell search feature can be used to confirm that a lithium-containing structure has already been published in the CSD. Lithium-containing by-products are common in many reactions; therefore, it is usually advisable to screen for these before devoting resources to a full experiment.

Lattice centring may not always be reliably determined prior to data collection therefore it is advisable to search for a *reduced cell*.



Putative amidolithium compound, CSD entry BUXNUD.

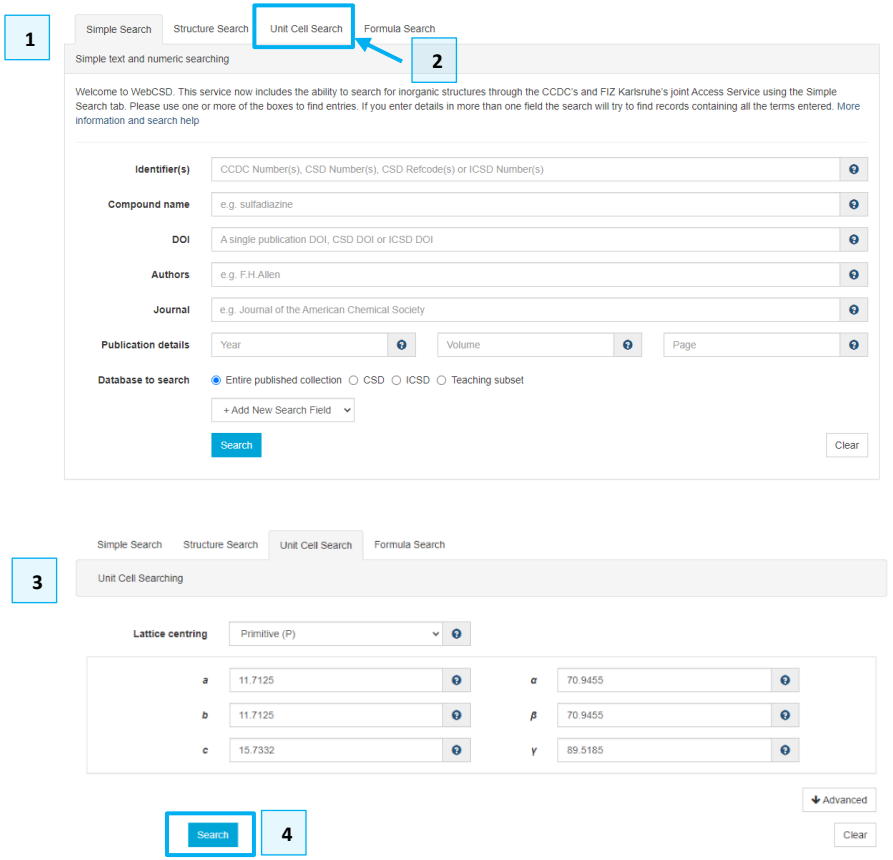
---

1) Open a web browser and navigate to the Access Structures webpage (https://www.ccdc.cam.ac.uk/structures/) if you do not already have WebCSD open. If you have already run a search, click **New Search** at the top.

2) Navigate to the **Unit Cell Search** tab.

3) The reduced cell for the putative lithium compound is *a* = 11.7125, *b* = 11.7125, *c* = 15.7332; α = 70.9455, β = 70.9455, γ = 89.5185. Enter these values in the relevant boxes and select the lattice centring as primitive (P).

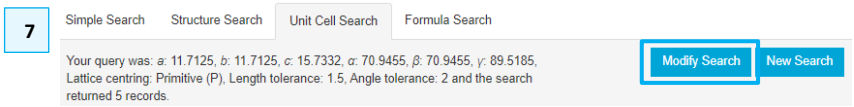
4) Press **Search**.



5) A new page will load showing the results (refcode listed on the left-hand side) and information about the currently selected entry. There should be at least 6 records returned (your value may be significantly higher as WebCSD is updated every minute). Scroll through the results and notice how the information and viewer is updated as you do so. The refcode family BUXNUD corresponds to the putative structure. Notice that CSD entry JUJCOI is an isostructural compound with copper in place of lithium.

6) Scroll down to the *Chemical details* and *Crystal details* to confirm that the physical and chemical properties match those expected for the target compound.

7) We can refine the results further by decreasing *Length tolerance* and *Angle tolerance* in the search. Click **Modify search** to return to the search setup.



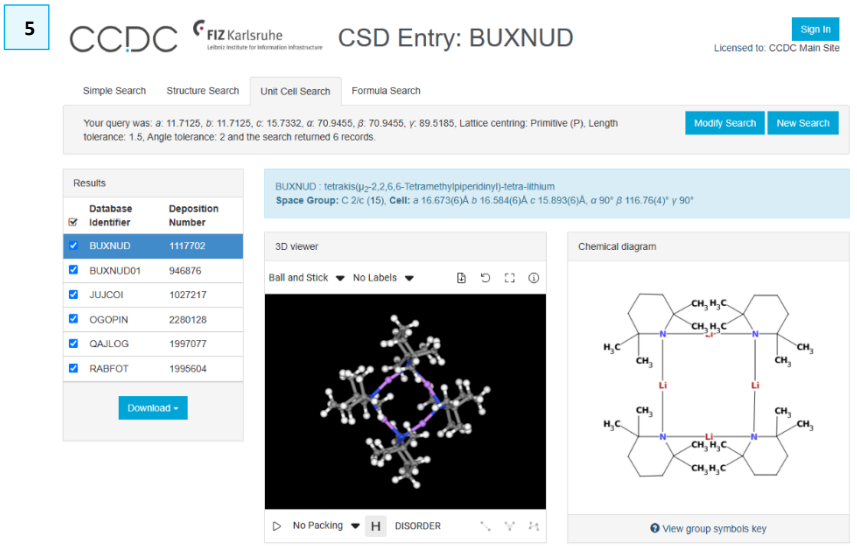
8) Click on the **Advanced** button to bring up the *Tolerances* options. Reduce the *Length tolerance* to "0.5" and the *Angle tolerance* to "1". Hover over the symbol to reveal the units.

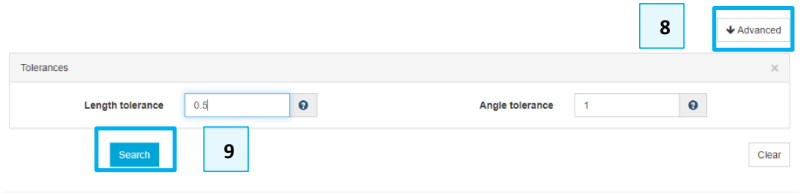
9) Click **Search**.

10) This time, a single result, only BUXNUD01 is returned. You may wish to experiment with the *Tolerances* to find the optimum number of results.

Conclusion

This example demonstrated that the unit cell searching feature in WebCSD is a fast and convenient way to retrieve crystallographic information, and particularly to screen a putative crystal structure which may already be published. The search can be customised to allow for the precision of the determined unit cell.





Summary

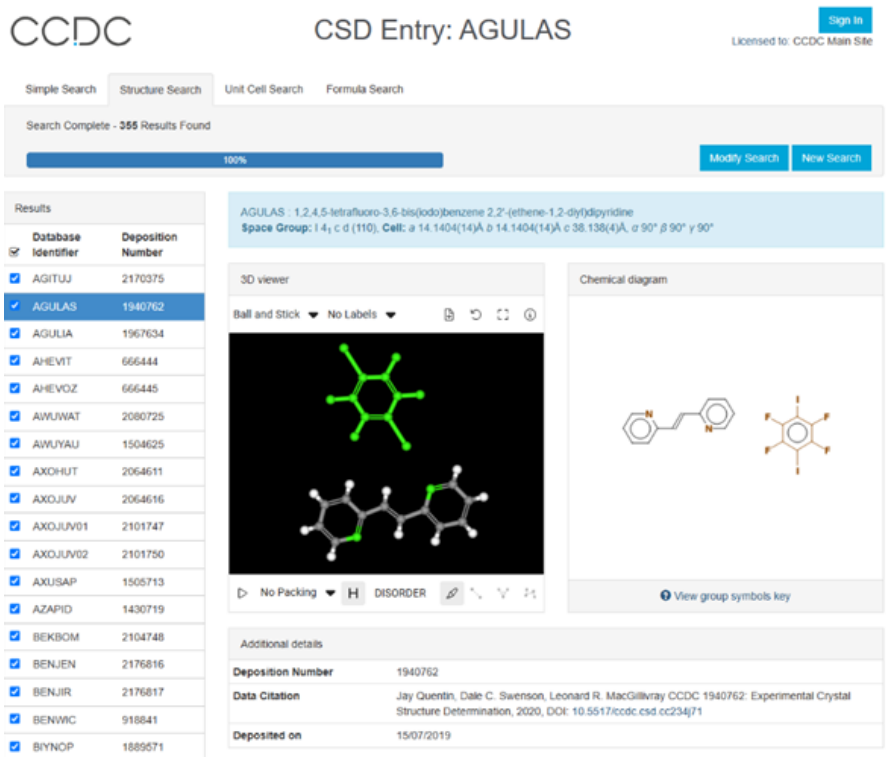
In this workshop you have explored the search options that are available with WebCSD. You should now be confident in:

- Creating substructure searches.
- Defining 2D and 3D search parameters.
- Searching for compounds by chemical formula, including defining variable atom counts.
- Conducting a unit cell search, including customising the tolerances of the metric parameters.

For further information, and tips on how to make the most of WebCSD, see the FAQs relating to WebCSD on our website.

Next Steps

You might like to explore some of the search results further using the structure visualisation software Mercury. You can find self-guided workshops on Mercury here and an on-demand training course here.



Feedback

We hope this workshop improved your understanding of WebCSD and you found it useful for your work. As we aim to continuously improve our training materials, we would love to hear your feedback. Follow the link on the workshop homepage and insert the workshop code, which for this self-guided workshop is *WCSD-002*. It will only take 5 minutes and your feedback is anonymous. Thank you!

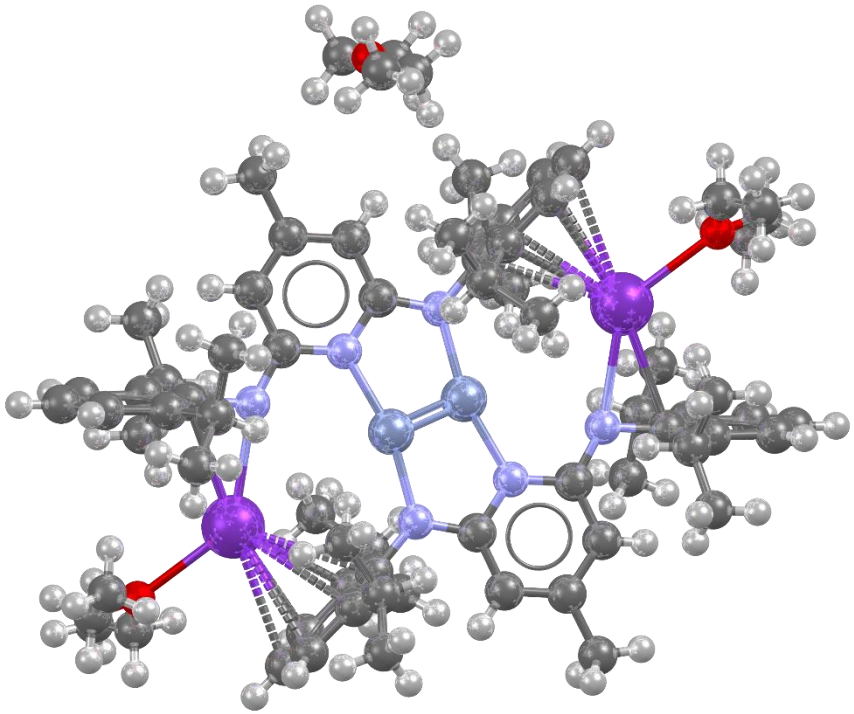
Glossary

Bimetallic
A bimetallic complex contains two metal centres which may be of the same type or different (heterobimetallic).

CIF
A Crystallographic Information File. This is the standard file format for crystallographic data. Further information about this format can be obtained from the IUCr website here (https://www.iucr.org/resources/cif) and from the CCDC website here (https://www.ccdc.cam.ac.uk/community/access-deposit-structures/deposit-a-structure/guide-to-cifs/).

Formula unit
In WebCSD, a formula unit refers to an individual molecule or ion within a structure.

GCD
A GCD file contains a refcode list which can be read into programs such as Mercury and ConQuest.



A bimetallic (or strictly, heterobimetallic) species from CSD Entry TIFPAB.

Halogen bond

A halogen bond occurs when there is evidence of a net attractive interaction between an electrophilic region associated with a halogen atom in a molecular entity and a nucleophilic region in another, or the same, molecular entity. [IUPAC definition, *Pure Appl. Chem*., 2013, **85**, 1711–1713].

Schematically, the halogen bond can be represented as R—X···Y, where in the entity R—X, X is any halogen atom with an electrophilic region, and is designated the halogen bond donor, whilst Y, is the entity with the nucleophilic groups, and is called the halogen bond acceptor.

In a typical halogen-bonded complex, R—X···Y, the interatomic distance X···Y tends to be less than the sum of the van der Waals radii, the R—X bond is usually extended relative to unbonded R—X, and the R—X···Y angle tends to be close to 180°.

Intermolecular contact

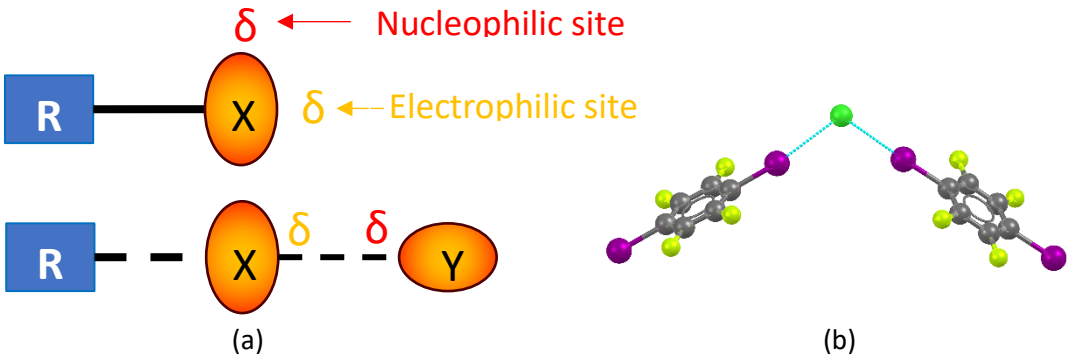
An intermolecular contact is simply a contact between two different molecules. Usually, it connotes non-covalent interations, such as intermolecular hydrogen or halogen bonding interactions.

Lattice centring

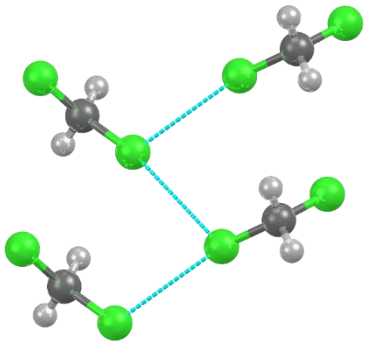
When the unit cell does not reflect the symmetry of the lattice, it is usual to refer to a 'conventional', non-primitive basis instead of a primitive basis. This is done by adding lattice nodes at the centre of the unit cell or one of the three faces. The conventional lattice centrings are primitive (*P*), A-face centred (*A)*, B-face centred (*B*), C-face centred (*C*), body centred (*I*), all-face centred (*F*) and rhobohedrally centred (*R*).

Mean plane

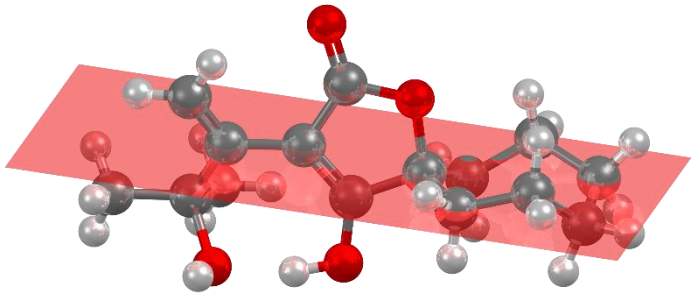
A plane in which the average squared radial distance of the atoms to the plane is minimised.



(a) Schematic illustration of a halogen bond involving donor group R—X and acceptor group Y. (b) Halogen bond between an aromatic C-I group and Cl^- in CSD Entry OHOVUD.



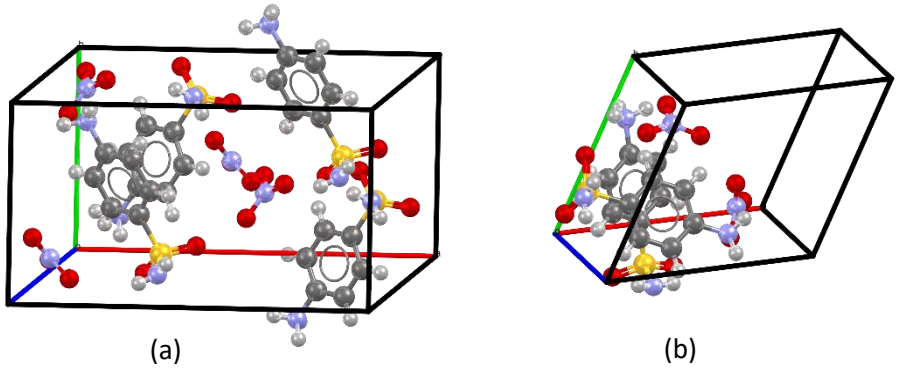
The solid-state structure of dichloromethane (CSD Entry DCLMET10) displaying intermolecular Cl···Cl contacts.



The mean plane defined by the seven carbon atoms of the cycloheptane ring in CSD Entry VODDUN.

Reduced cell

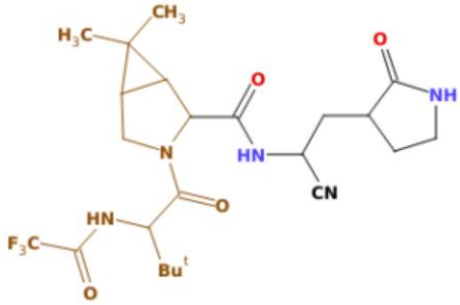
A parallelepiped whose sides are the reduced basis vectors (**a**, **b** and **c**). It is a primitive cell; there can be no lattice vectors shorter than **a**; of those not directed along **a**, none is shorter than **b**; and of those not lying in the ab plane, none is shorter than **c**. The angles between the vectors are either all acute or all obtuse. See https://dictionary.iucr.org/Reduced_cell for further information.



(a) The conventional C-centred monoclinic unit cell of CSD Entry ABADOY and (b) the reduced unit cell of the same structure.

Substructure

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



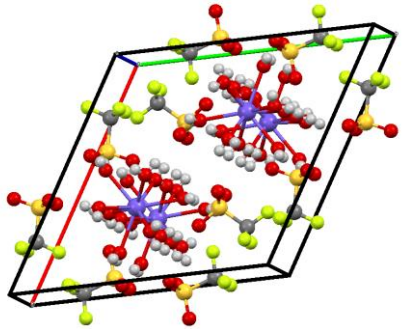
Nirmatrelvir, CSD Entry ZIVMAE, with a substructure highlighted in brown.

Unit cell

The unit cell is the basic building block of a crystal, repeated infinitely in three dimensions.

Van der Waals radius

The van der Waals radius of an element is one half of the distance between the closest approach of two non-bonded atoms of a given element.



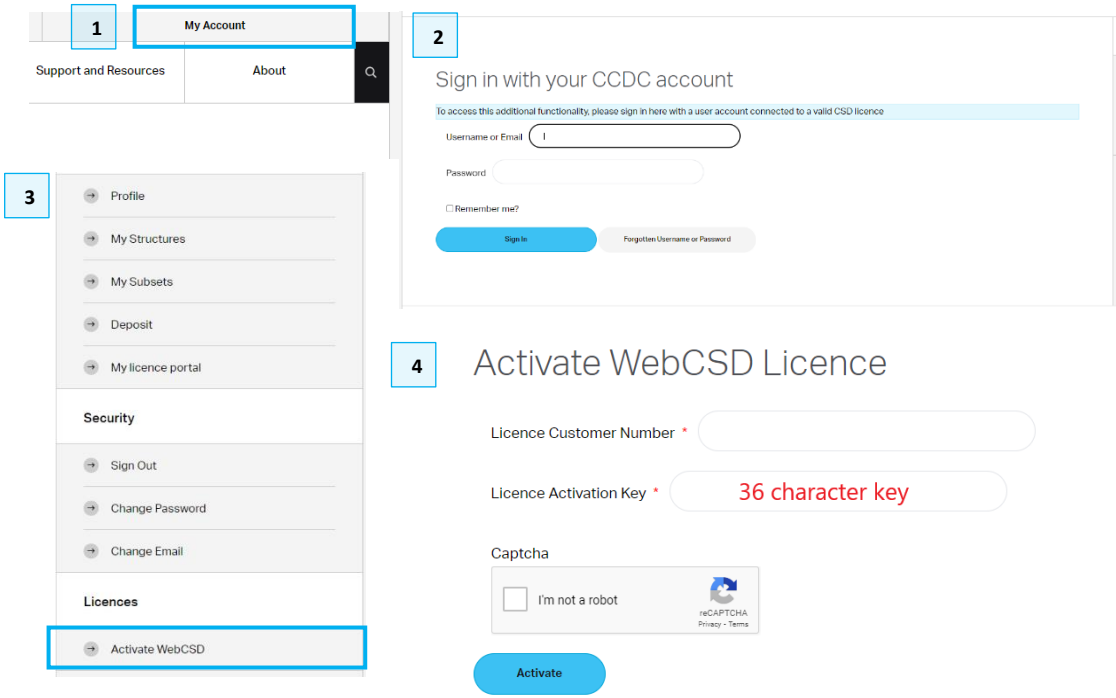
The unit cell of CSD entry YEVROH (with contents).

Activating WebCSD

To access the advanced WebCSD searching options, such as structure search and unit cell search, within the CSD web interface you will need to sign in to the site and connect a valid CSD licence. You can connect a CSD licence by either of the following methods:

- Accessing the CSD web interface from an IPv4 address registered to a CSD campus licence (we are unable to support IPv6).
- By entering a valid CCDC customer number and activation key once signed in. To do this you will need to:
  1. Go to My Account.
  2. Sign In or Register if necessary.
  3. Select "Activate WebCSD" under Licences on the right-hand side of the page.
  4. Enter your CCDC customer number and activation key.

The campus licence holder at your institution should be able to provide you with the CCDC customer number and activation key for your institution or they can provide us with the relevant IP addresses. If you are unsure of who this is or you are unsure if your institution has a licence then please contact us using our [enquiry pages (https://www.ccdc.cam.ac.uk/contact-us/)](https://www.ccdc.cam.ac.uk/contact-us/).



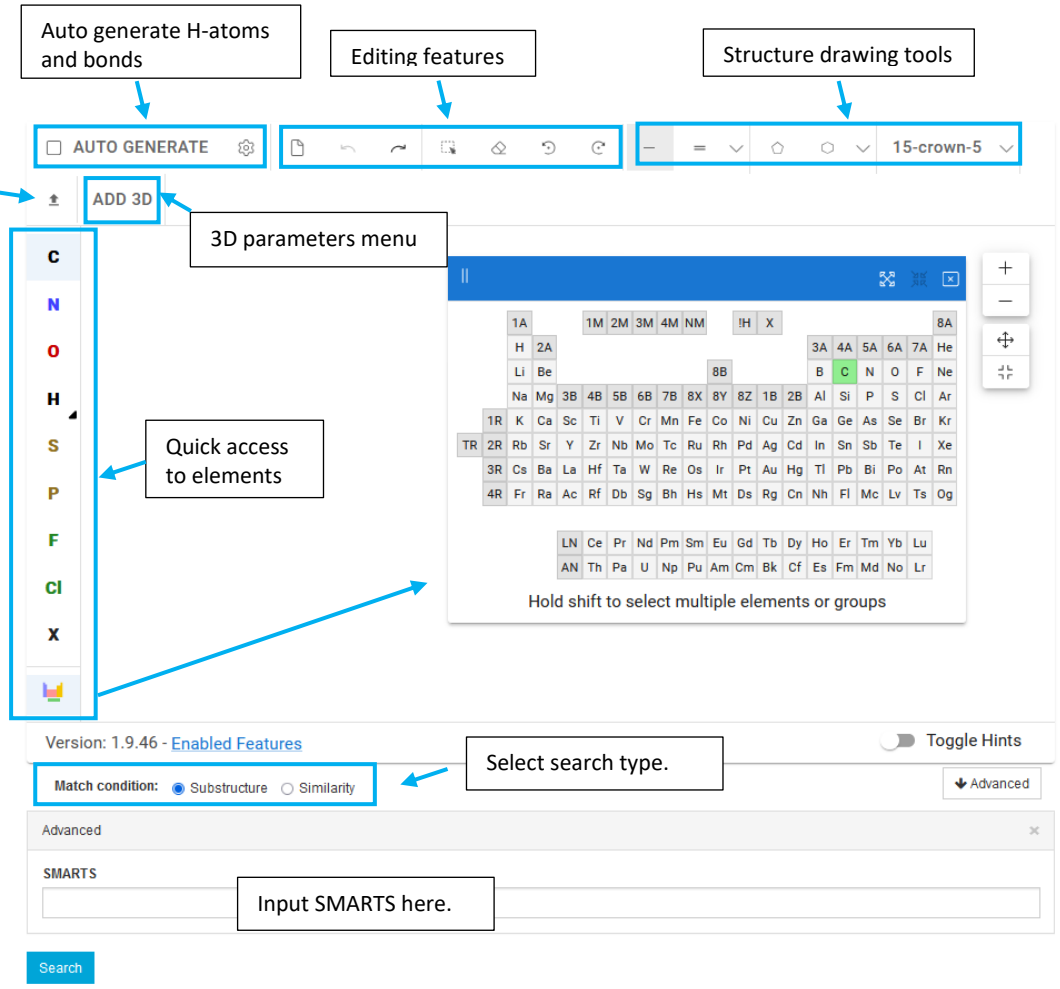
Basics of the WebCSD Sketcher

The WebCSD sketcher is the interface for creating substructure and similarity searches in WebCSD.

In the following we will see some of the basics of using the sketcher to create search queries.

In the WebCSD sketcher, we find:

1.
   - **Top left**: tools for automatic hydrogen and connection generation.
   - **Top middle**: tools to create new sketches, to edit and select parts of existing sketches and options to undo/redo operations.
   - **Top right**: tools to select bond types, simple ring drawing templates, and advanced templates.
2.
   - **Second line from the top**: button to upload a structure from a MOL file directly into the sketcher, and button to access the 3D parameters menu.
3.
   - **Left hand side toolbar**: quick access to elements and options for hydrogen placement, button to show/hide periodic the table. *Note: the periodic table is floating and can be repositioned using* **||**
4.
   - **Main window**: click here to add structure components (atoms and bonds) and click on atoms to select them to define 3D parameters.
5.
   - **Below sketcher**: options to select substructure or similarity search.
   - **Below sketcher**: options: advanced options – input SMARTS here.



Explanation of 3D parameters

3D parameter options are accessed from **ADD 3D**. To define a 3D parameter, you must select atoms. Shift + left mouse click to select multiple atoms.

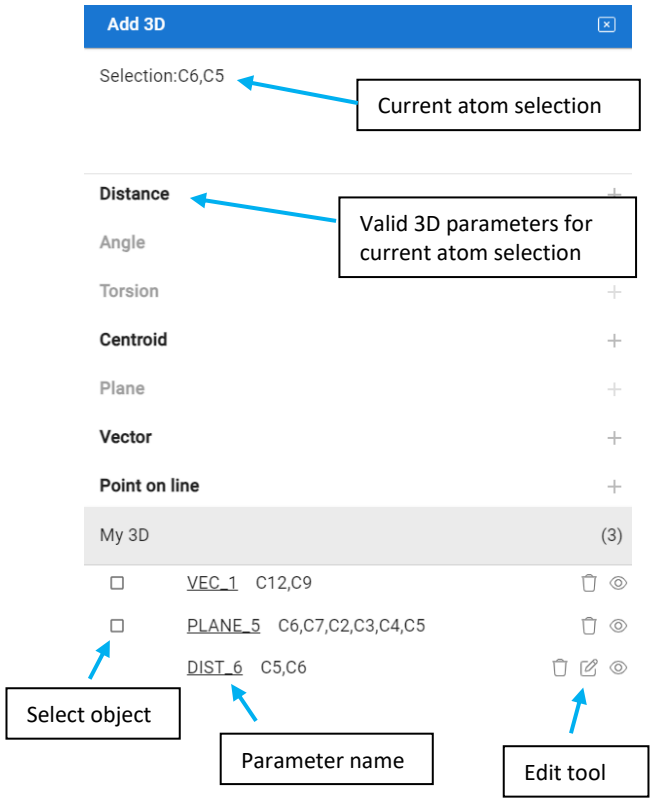
- Distances require exactly two atoms.
- Angles require exactly three atoms, in the correct sequence.
- Torsions require exactly four atoms, in the correct sequence.
- Planes require a minimum of three atoms (the sequence does not matter).
- Centroids require a minimum of two atoms.
- Vectors and points on lines require two atoms (and the point on the line requires a distance of extension from the tip of the vector defining the line).

It is possible to define 3D parameters involving other defined geometric objects. For example, with a vector and a plane defined, the angle between them may also be defined.

The Add 3D menu consists of:

1. Currently selected atoms.
2. Choice of 3D parameters to define. Valid parameters for the atoms selected are in black; non-valid ones are greyed out.
3. List of defined parameters and geometric objects, which may be edited by selecting the pen tool (geometric objects may be selected using the tick box).

To add a parameter, click the **+** icon next to a valid option.



Certain 3D parameters may be constrained:

- Distances may be constrained to be intramolecular or intermolecular, or any type, within a specified range.
- Angles and torsions may be constrained within a specified range.

To add a constraint, move the slider to constrained and enter the desired values.