Access Structures Workshop (WCSD-001)

2020.3 CSD Release

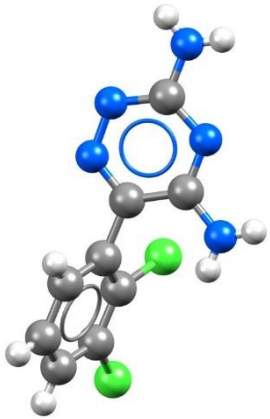


Table of Contents

Introduction

Access Structures is a web page that allows users to search the Cambridge Structural Database (CSD), the Inorganic Crystal Structure Database (ICSD) and the CCDC's Teaching Subset. The CSD is the world's repository of experimentally determined organic and metal-organic crystal structures and is compiled and distributed by the Cambridge Crystallographic Data Centre (CCDC). The ICSD is the world's largest database for completely determined inorganic crystal structures and is compiled and distributed by FIZ Karlsruhe – Leibniz Institute for Information Infrastructure. The CCDC's Teaching Subset is a subset of the CSD containing structures that have been carefully selected to aid the explanation and understanding of a variety of chemistry concepts, such as bonding, molecular geometry, stereochemistry, functional groups and symmetry. Access Structures supports a variety of different search parameters, as well as combinations of parameters.

Objectives

This is a self-guided workshop to lead you through the functionality of our Access Structures web-based service. At the end of this workshop, you should be able to:

- Conduct a search in Access Structures
- Visualize results in the browser
- Explore links to other databases
- Download structures

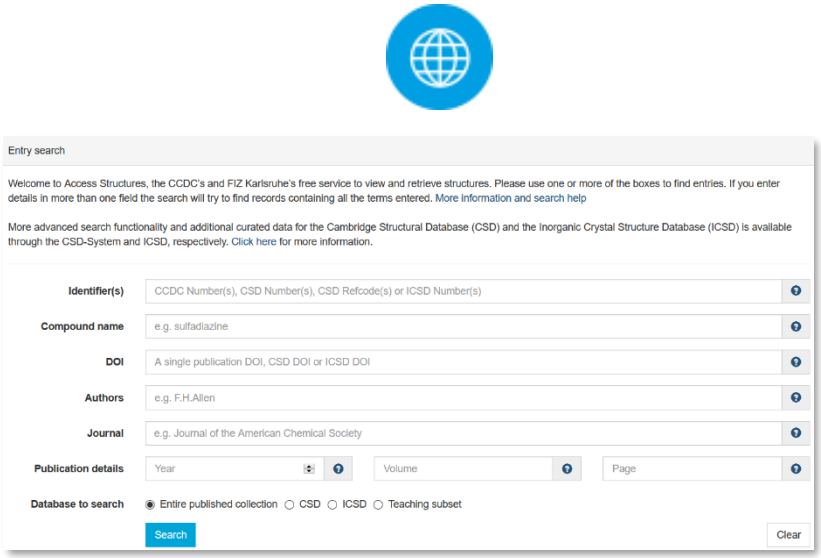
This workshop will take approximately 1.5 hours to complete

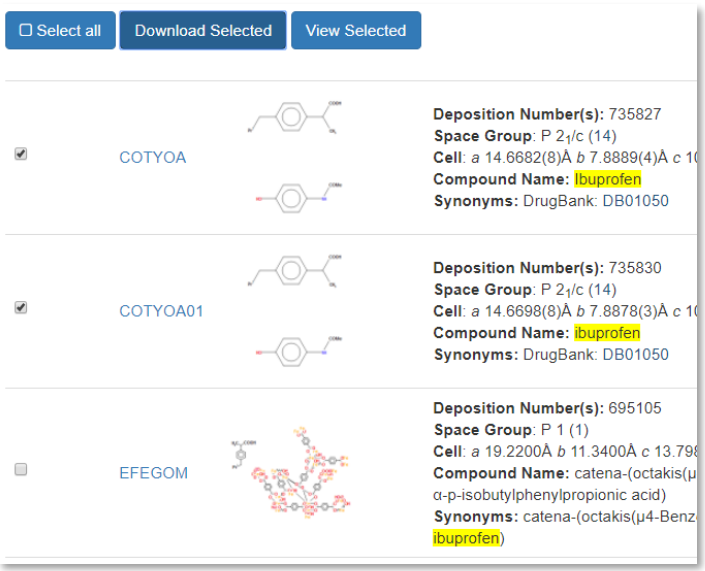
Pre-required skills

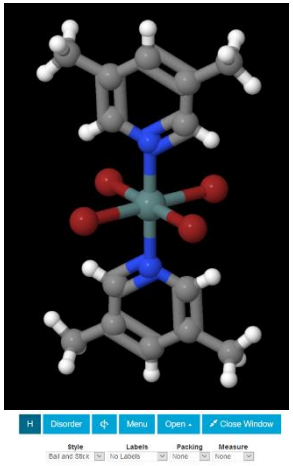
There are no pre-required skills for this workshop

Materials

There are no additional materials required for this workshop.







Example 1. Searching

Access Structures supports a variety of different search parameters, as well as combinations of search parameters.

Searching by *Compound name*

The Access Structures ***Compound name*** search field gives users the ability to enter full or partial chemical names of substances as well as some common names.

1. Open a web browser (such as Chrome) and navigate to www.ccdc.cam.ac.uk/structures

2. In the ***Compound name*** search field enter the name of a chemical you would like to view, in this example “caffeine” is used, and click **Search**.

3. Scroll through the first few results pages and notice the variety and differences between the hits, especially between the diagrams, *compound name* data fields and *synonym* data fields.

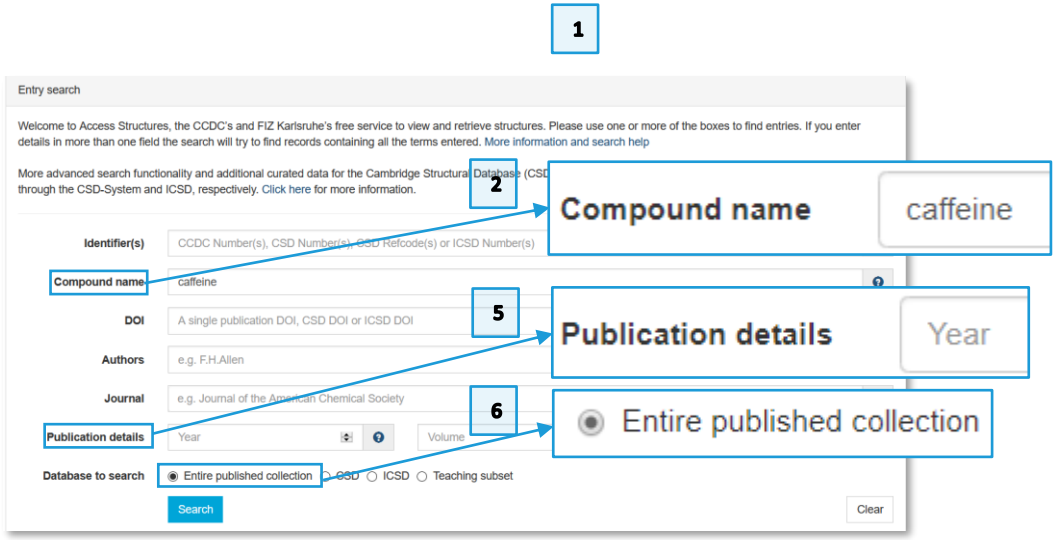
Searching by *Compound name* and *Publication details*

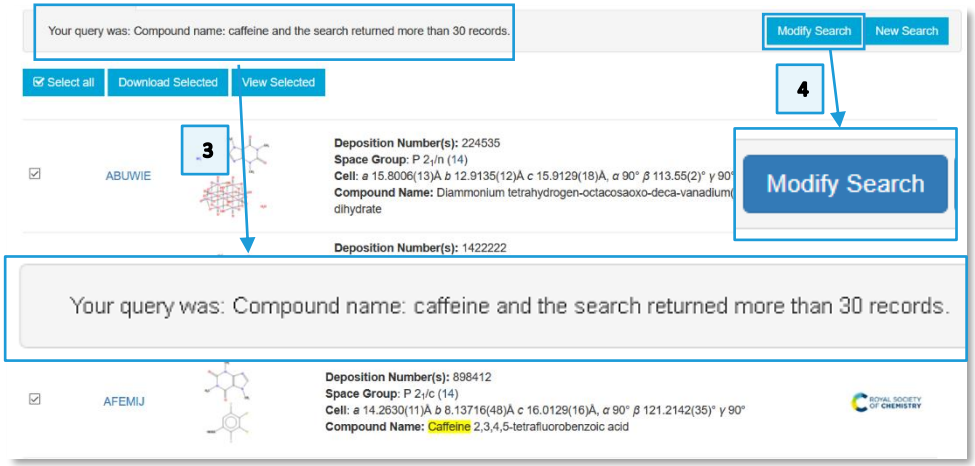
Search parameters can be combined to provide a more customisable search. The steps below describe a search for the structural determinations of caffeine that have been published in 2018, thereby narrowing the results from the previous search.

4. From the results page in 3, click the **Modify Search** button.

5. Add “2018” to the ***Publications details****, Year* search field.

6. Select the radio button for *Entire published collection* and click **Search**. There should be fewer results than after step 3.

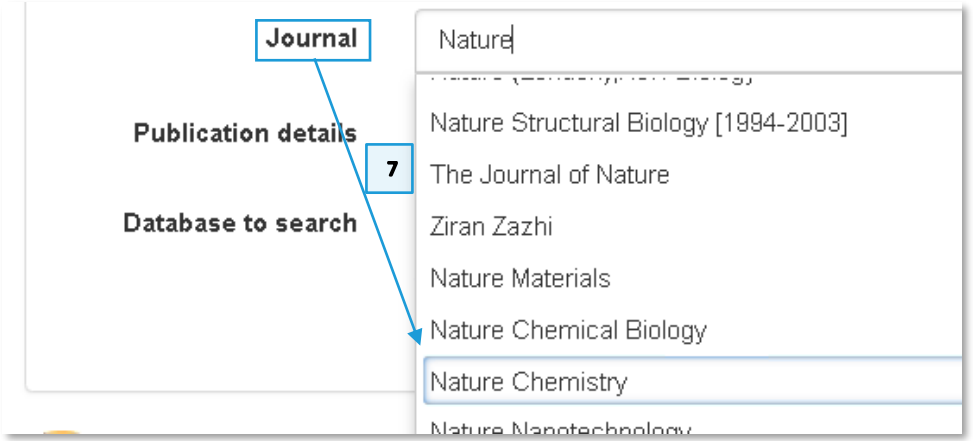




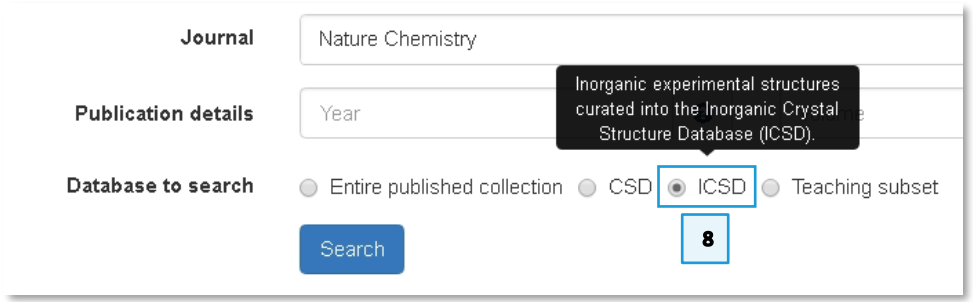
Searching by *Journal* and selecting a database

Another variation of customising the search is to complete the search parameters with your desired criteria and choose which database to search; either the *Entire published collection*, the *CSD*, the *ICSD* or the *Teaching subset*. For example, finding out what ICSD structures have been published in Nature Chemistry.

7. Click on **New Search**. In the *Journal* field enter "Nature". You will then see some pre-filled options to choose from. Scroll down the options until you find "Nature Chemistry" and select this.



8. Select the radio button for *ICSD* and click **Search**.

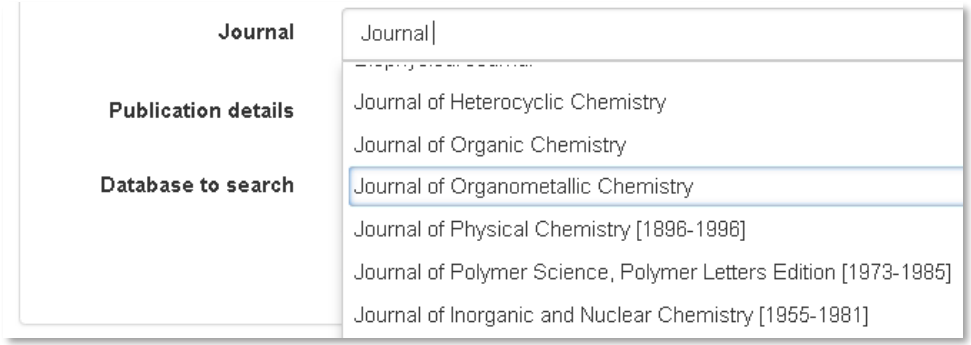


Conclusion

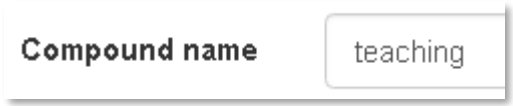
These exercises have shown some examples of different searches using Access Structures. The multiple search fields and 4 database options allow the user to customise their search parameters.

Further exercises

- Enter "Journal" into the *Journal* search field. You will see that this gives many options of different journals you can explore.



- Enter "Teaching" into the *Compound name* search field. This returns all the structures in the Teaching subset.



- Enter "DrugBank" into the *Compound name* search field. This returns the structures that have a link to the DrugBank website.

Example 2. Visualising

In “Example 1. Searching” you have seen that Access Structures is a powerful search tool that allows users to access the CSD, the ICSD and the CCDC’s Teaching Subset. In this example you will explore the detailed results pages provided by Access Structures.

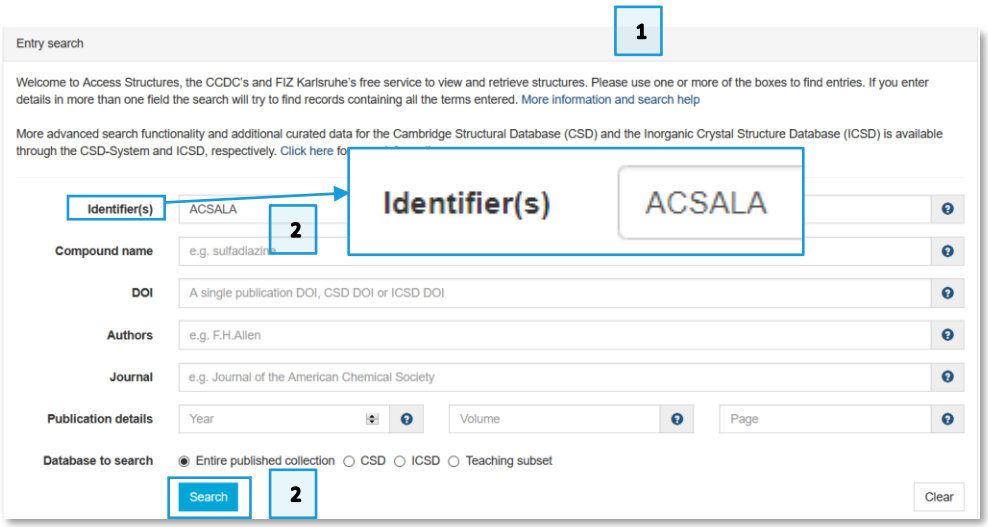
Navigating to a detailed results page

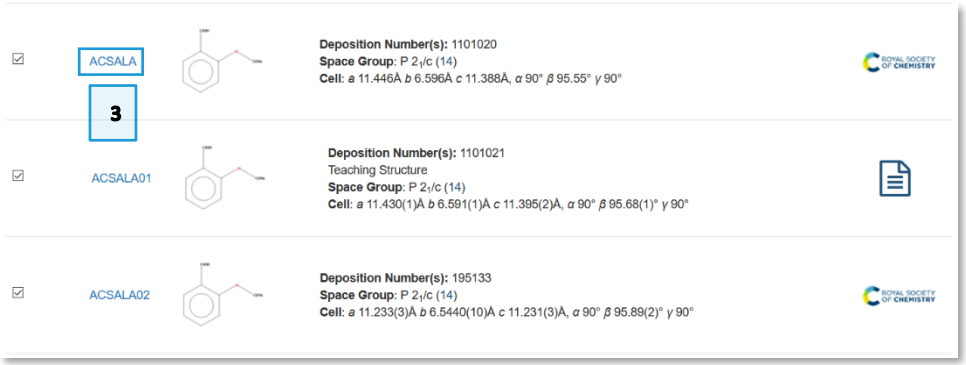
1. Open a web browser (such as Chrome) and navigate to www.ccdc.cam.ac.uk/structures. If you choose to continue from the previous example, verify that all the search parameters are back to their default values.

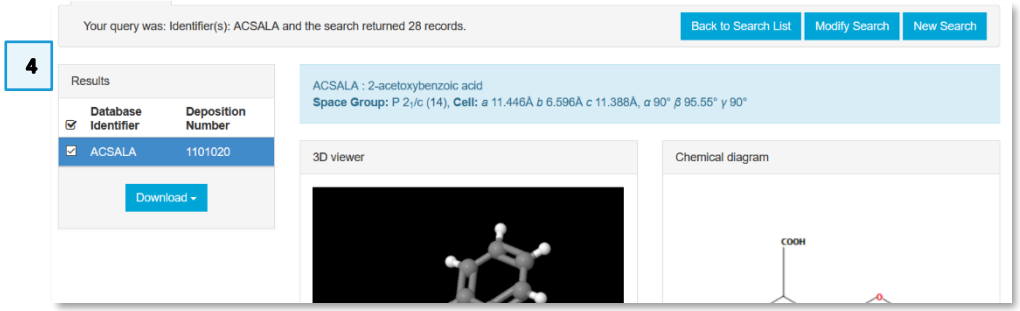
2. Enter “ACSALA” into the ***Identifier(s)*** search box and click **Search**. “ACSALA” is the refcode identifier for this particular database entry.

3. Click on the refcode of the first hit “ACSALA”.

4. This is the detailed results page. Here you can find a selection of information about the selected structure, including a 3D viewer showing a diagram of the structure in 3D, a chemical diagram, where it was published (associated publications), and links to the PubChem if the structure also exists in that database.







Exploring the 3D Viewer

The 3D viewer has lots of functionality to enable users to further their understanding of the structures they view. Here we will explore some of the main features.

5. On the detailed results page for ACSALA click on the button to expand the JSmol 3D viewer

6. How to move the view of the structure with the keyboard and the mouse:
   a. The left mouse click selects and deselects atoms.
   b. Scrolling down zooms in and scrolling up zooms out.
   c. Clicking and holding the left mouse button rotates the molecule in 3D.
   d. Holding shift, then clicking and holding the left mouse button and moving the cursor left and right rotates the structure in the plane of the screen.
   e. Holding shift, then clicking and holding the left mouse button and moving the cursor up zooms out and down zooms in.

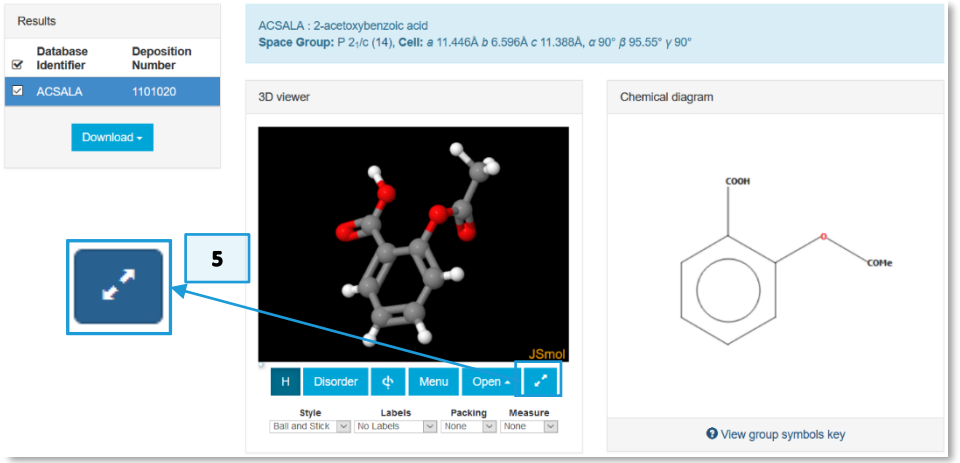
7. Clicking "**H**" either removes or adds the hydrogen atoms to the view. This function is particularly useful with larger structures.

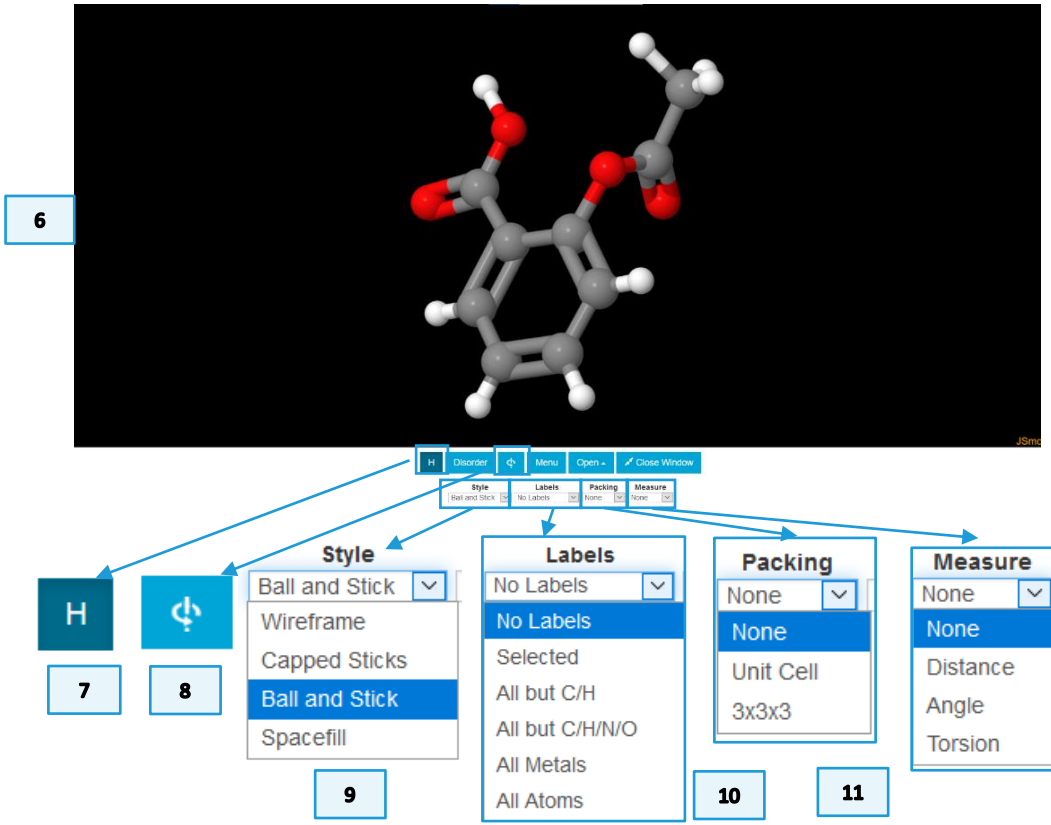
8. The button highlighted by number 8 on the right will set the molecule to rotate

9. Explore the options in the ***Style*** drop-down menu to see the different representations of the structure; wireframe, capped sticks, ball and stick or spacefill.

10. The ***Labels*** drop-down menu has different options for what labels are shown. This is particularly useful for larger structures.

11. The ***Packing*** drop-down menu gives options for visualising the unit cell as well as expanding this to 3x3x3 unit cell. Zoom out until there is some black space around the visualisation. Try moving the structure to find channels and void space. Try to line up the repeating units. You may want to turn off the view of the hydrogen atoms to make this easier.





Measuring distances

12. Revert the viewer back to showing only 1 molecule with the hydrogens by selecting ***Packing*** → none, ***Labels*** → no labels, ***Style*** → Ball and Stick, turning off the rotation and turning on **H**. Please also ensure ***Measure*** → None is selected.

13. Exit the enlarged view, either by clicking **Close Window**, the **X**, in the top right corner or by clicking anywhere on the background.

14. Click **New Search**.

15. Enter "GLUCSA" into the ***Identifier(s)*** search box and click **Search**.

16. Open the detailed results page for GLUCSA and expand the 3D viewer.

17. Orientate the molecule so that 3 carbon atoms are clearly visible, as shown on the right.

18. Select these 3 carbon atoms by clicking on them with a left mouse click.

19. Expand the ***Labels*** drop-down menu and select "Selected".

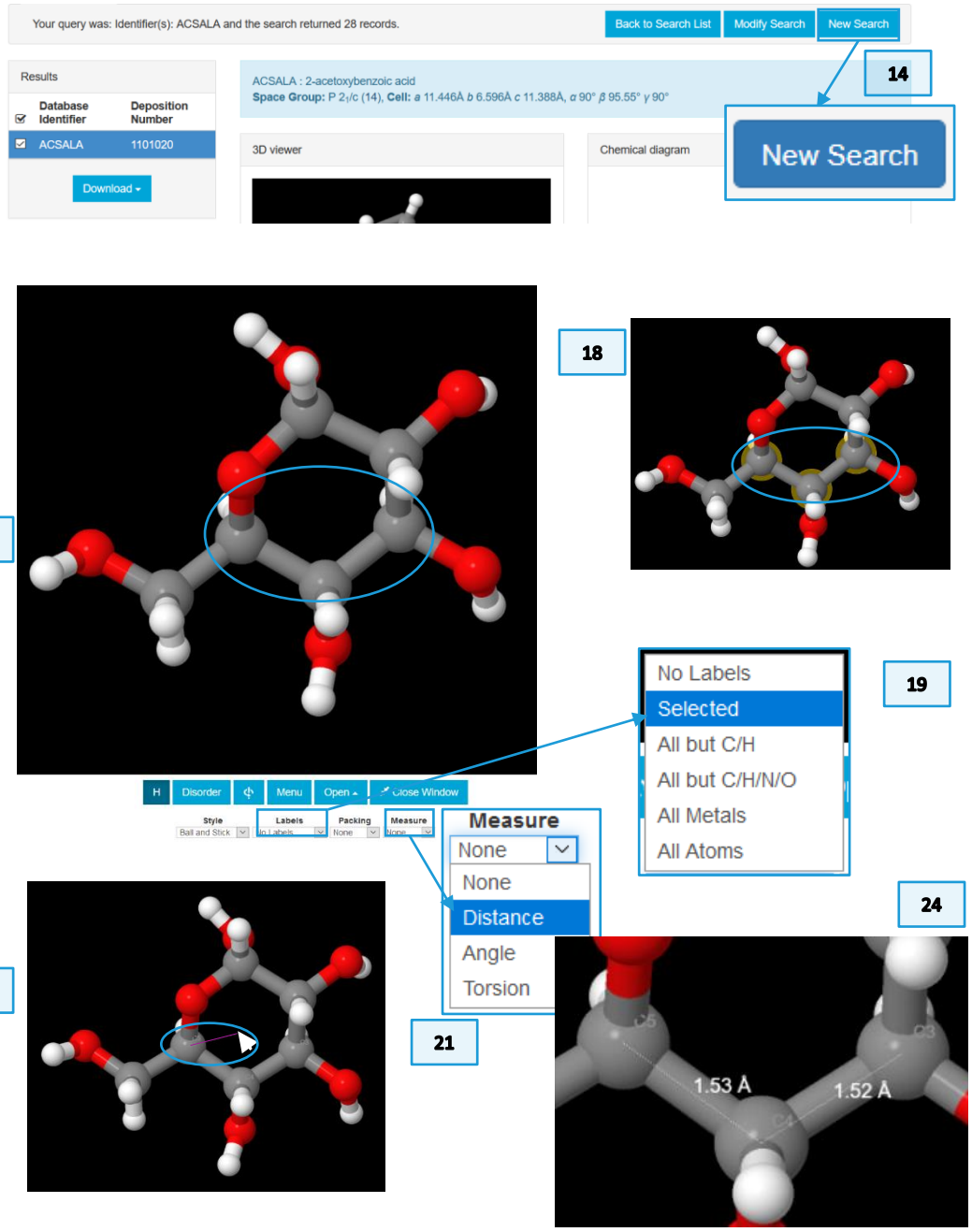
20. Deselect the 3 carbon atoms using the left mouse click.

21. Expand the ***Measure*** drop-down menu and select ***Distance***.

22. Click on atom **C5**. Slowly move the cursor and notice a faint pink line following the point of the cursor. This line is the measuring tool. It measures the distance between the selected atom (in this case **C5**) and whatever atom is hovered over with the cursor. These measurements are shown in pink.

23. Click on **C4**. The pink measurement text turns white.

24. Repeat steps 22 and 23 to measure the distance between atoms **C4** and **C3**.



Measuring angles

25. Revert the view back to its default setting by repeating step 12. Close the 3D viewer and click **New Search**.

26. Enter "BOMQOJ" into the ***Identifier(s)*** search box and click **Search**.

27. Expand the 3D viewer and orientate the molecule as shown on the right.

28. Select the atoms as shown on the right, there are 5 atoms to select.

29. Expand the ***Labels*** drop-down menu and select ***Selected***.

30. Deselect the 5 atoms.

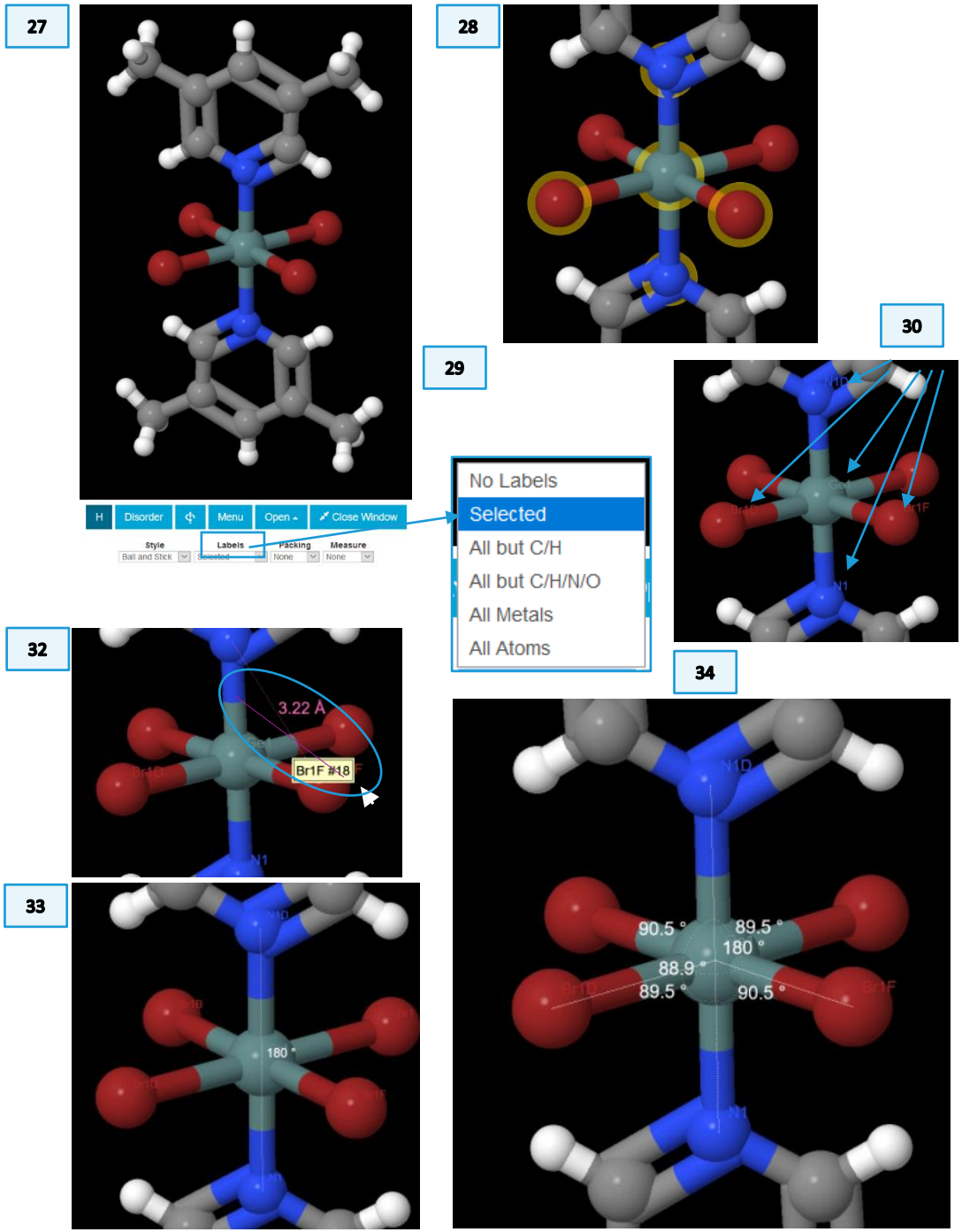
31. Expand the ***Measure*** drop-down menu and select ***Angles***.

32. Click on **N1D**, slowly move the cursor and notice a faint pink line following the point of the cursor. If the cursor is hovered over an atom, pink text appears with a distance measurement.

33. The first angle we are going to measure is between **N1D**, the central **Ge1** atom and **N1**. Click **Ge1**. You will see the faint line now drawn between **N1D** and **Ge1**. Next click **N1**. The text should turn white and show the angle measurement.

34. Measure the angle between each set of 3 atoms (as follows) by clicking on each atom in turn:
    a. **N1D**, **Ge1** and **Br1D**
    b. **N1D**, **Ge1** and **Br1F**
    c. **Br1D**, **Ge1** and **Br1F**
    d. **Br1D**, **Ge1** and **N1**
    e. **Br1F**, **Ge1** and **N1**. You should have 6 angle measurements.

    You can see that even though this is a fairly regular octahedral complex, the deviations from the 90° angles are still noticeable.

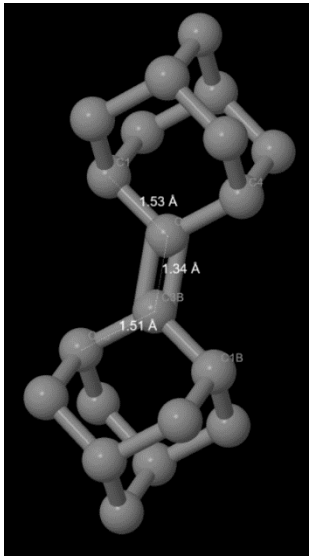


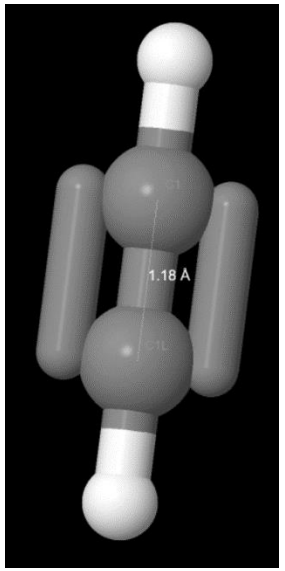
Conclusion

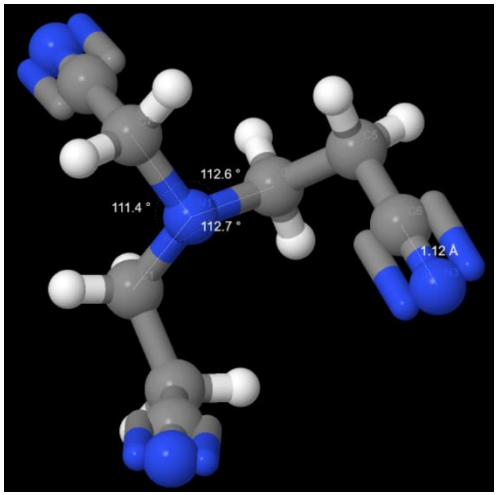
In this example we have explored some of the main functionality of the 3D viewer on the detailed results page of Access Structures. This has included different styles, packing the unit cell and expanding to 3x3x3, as well as measuring distances and angles between atoms.

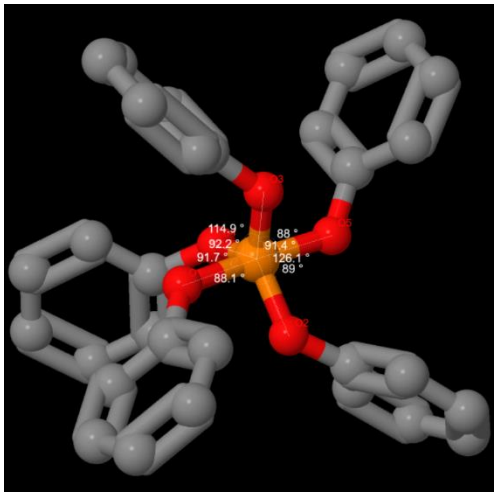
Further exercises

- For extra practise measuring distances search identifier ADYLAD and note the difference in distances between C-C single bonds and the C=C double bond. This can be compared to the C-C triple bond in ACETYL03. Have a look at the C-N triple bond in AWUDIF.
- Measure the 3 angles around the central nitrogen atom in AWUDIF.
- For extra practise measuring angles investigate the angles around the central atom in the trigonal bipyramidal structure PPHOXP.
- Navigate to the CCDC's educational resources webpage (www.ccdc.cam.ac.uk/Community/educationalresources/teaching-database/), download the annotated spreadsheet of all the structures in the Teaching Database. Scroll through to see if there are any compounds you would like to visualise and investigate.









Example 3. Linking

Over the years, CCDC has developed collaborations with other data repositories, such as the PDB, ChemSpider, DrugBank and PubChem. This example will highlight the advantages of these collaborations by showing you how to access structures in some of these other resources from their CSD entries.

Protein Data Bank (PDB)

The PDB collects, organises and disseminates data on biological macromolecular structures. The PDB assigns a 3- or 4-character code to each chemical component found in PDB entries, which is shown in the synonym field in Access Structures. Linking to the PDB gives users easy access to this additional wealth of information. (ref: www.ebi.ac.uk/pdbe/about) Over 1,500 entries in the CSD link to PDB entries.

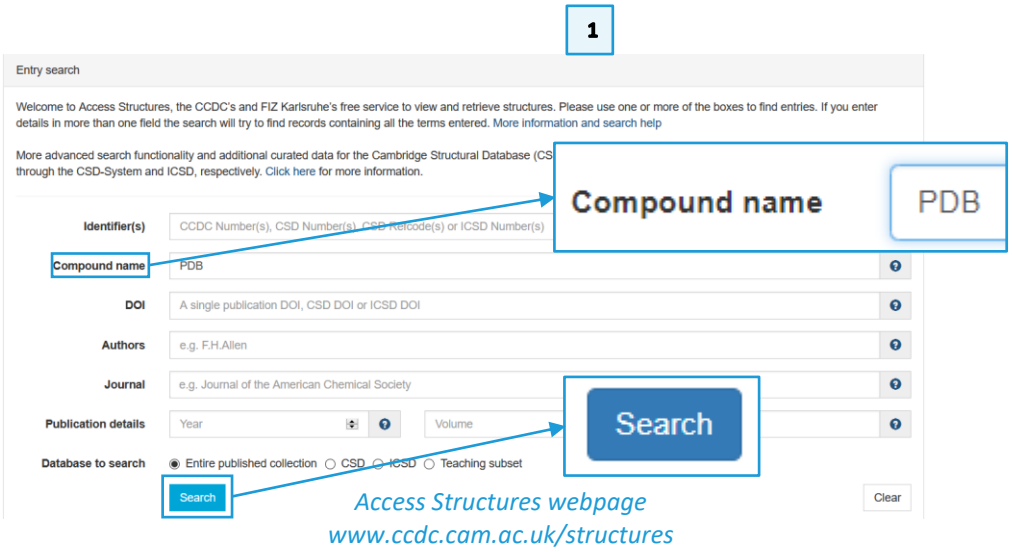
1. Navigate to www.ccdc.cam.ac.uk/structures and enter "PDB" into the ***Compound name*** search box and click **Search**.

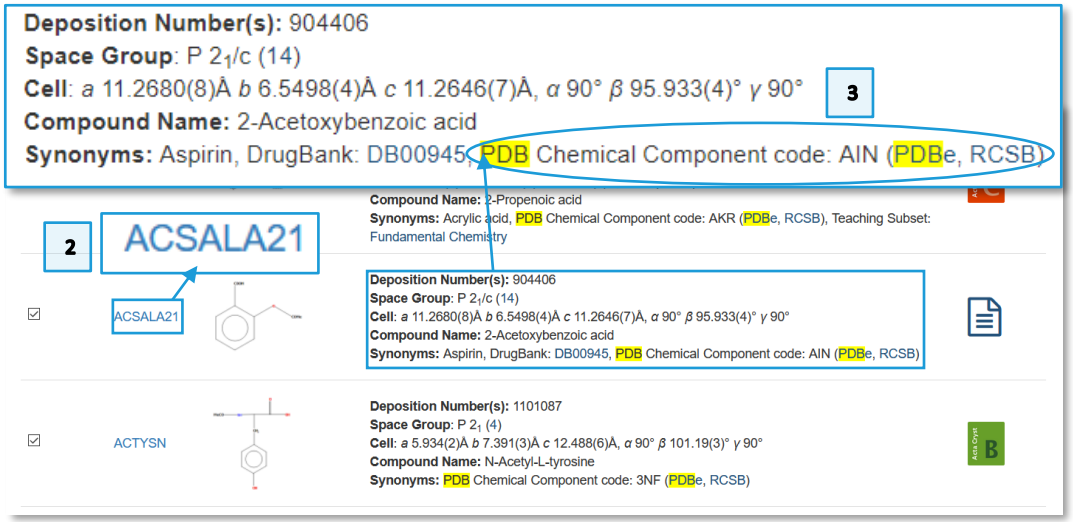
2. The search will return more than 30 records. The results are all the structures in the CSD which contain a component which also appears in the PDB. Scroll through the results to find the refcode "ACSALA21". This is the structure of aspirin.

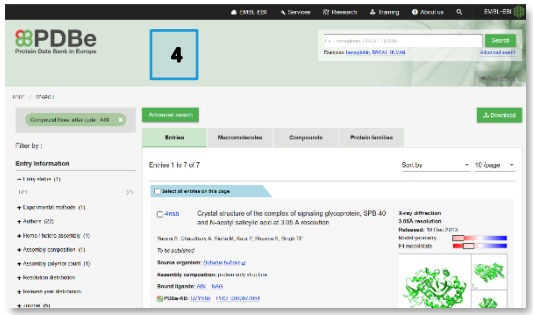
3. In the ***Synonyms*** field you can see the 3-character code, the "PDB Chemical Component code", is "AIN" and the links to the relevant PDBe and RCSB webpages.

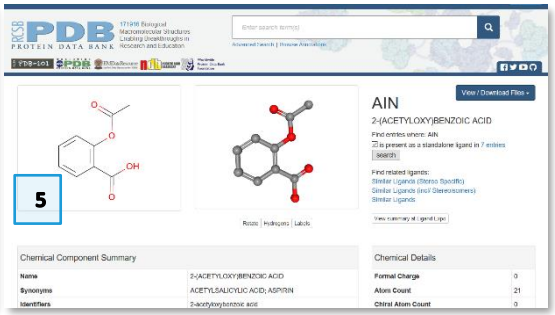
4. Click on "PDBe", this opens the related PDBe webpage in a new tab, showing entries that contain aspirin.

5. Return to the Access Structures results page tab in your browser and click on the second hyperlink, "RCSB". This opens the related RCSB webpage, showing more details about aspirin as a free ligand.







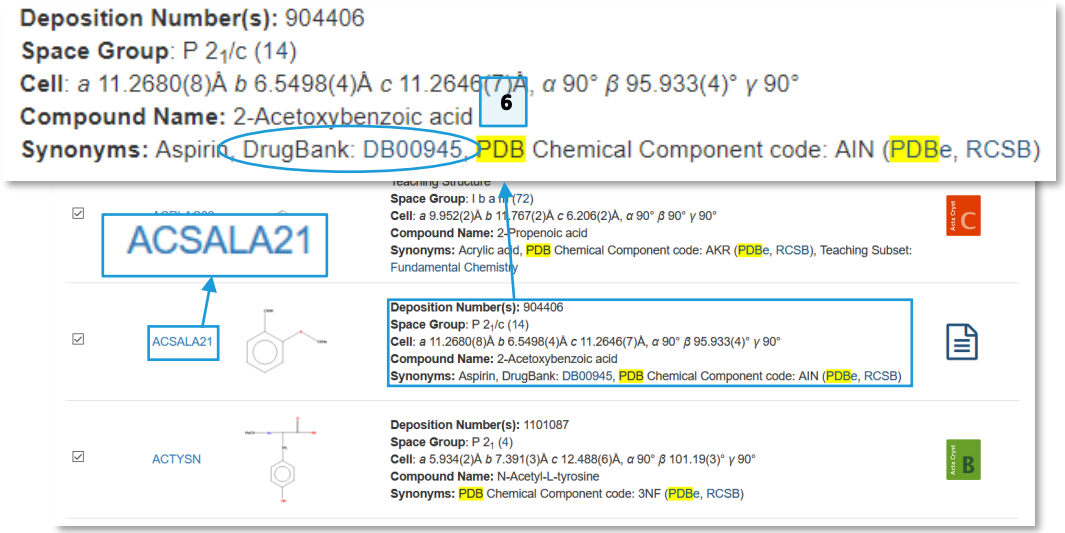


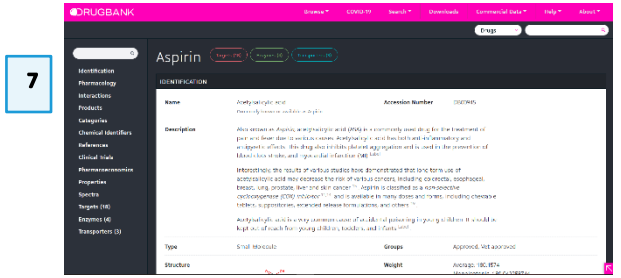
DrugBank

The DrugBank database is a comprehensive, freely accessible, online database containing information on drugs and drug targets. DrugBank is widely used by the drug industry, medicinal chemists, pharmacists, physicians, students and the public. Its extensive drug and drug-target data has enabled the discovery and repurposing of several existing drugs to treat rare and newly identified illnesses. (ref: https://www.drugbank.ca/about) Over 1,800 entries in the CSD link to DrugBank entries.

6. Return to the Access Structures results page tab in your browser and look again at the ***Synonyms*** field. DrugBank assigns a DrugBank ID to each entry in its database. This ID is a unique number consisting of a 2-letter prefix (DB) and a 5 number suffix. It is used to access the drug entry via the URL. The DrugBank ID for aspirin is "DB00945".

7. Click on "DB00945". This opens the relevant DrugBank webpage in a new tab in your browser. Scroll down the page to see what information you can find.



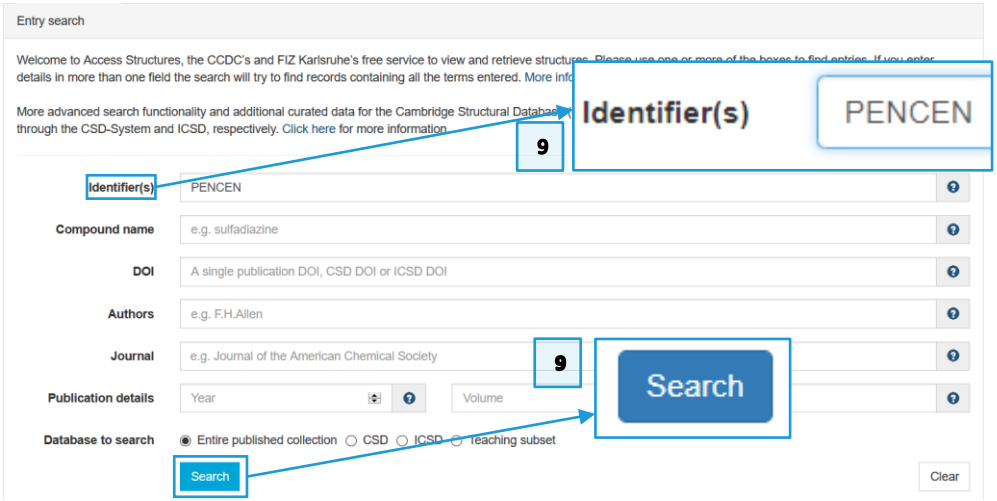


PubChem

PubChem is a database of chemical structures, identifiers, chemical and physical properties, biological activities, patents, health, safety and toxicity data, mostly of small molecules. (ref: https://pubchemdocs.ncbi.nlm.nih.gov/about) Over 61,000 entries in the CSD link to PubChem entries.

8. Click the **New Search** button on the Access Structures results page. (Not shown)

9. Enter "PENCEN" into the ***Identifier(s)*** search box and click **Search**.

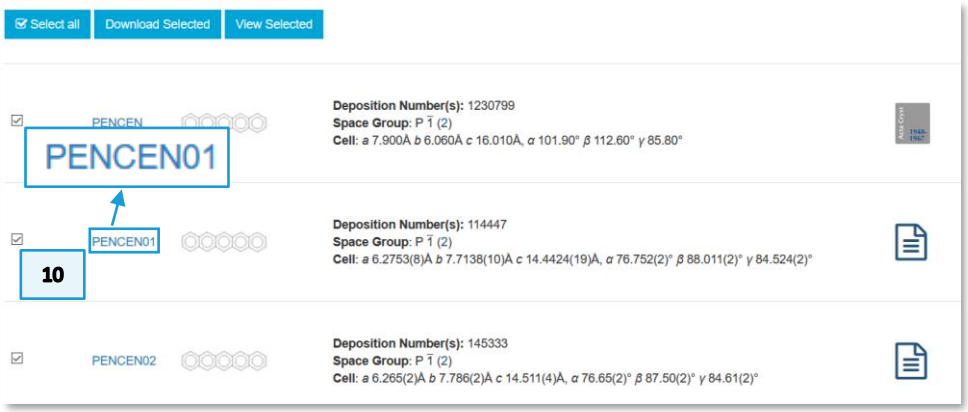


10. Scroll through the results to find refcode “PENCEN01” and click to on “PENCEN01” to open the detailed results page.

11. Scroll down the information on pentacene and find the section called “Links”. Here you find PubChem’s compound identifier (CID), a numeric code for a unique chemical structure.

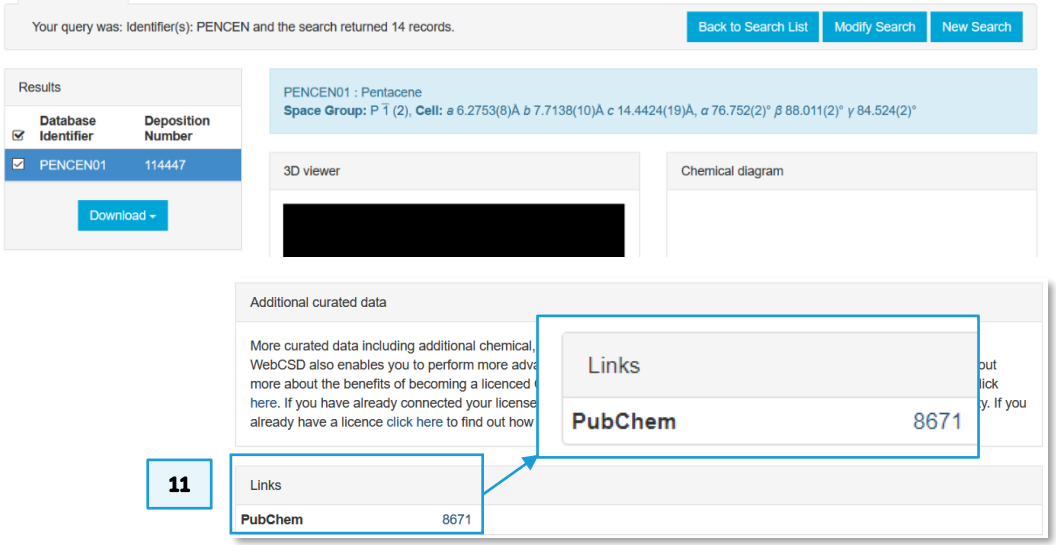
12. Click on this link to open the relevant PubChem information page on pentacene in a new tab in your browser.

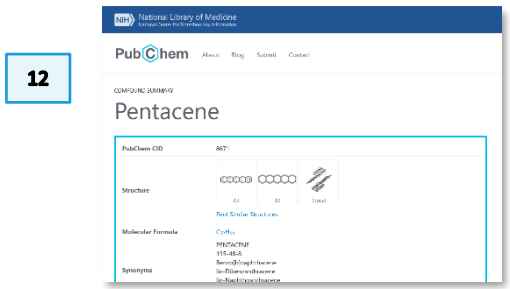
13. Scroll down to section 1.3 to view the crystal structure information. Notice that links to each refcode in the PENCEN family are given.

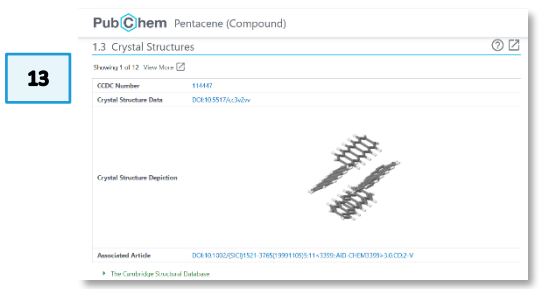


Conclusion

This example has shown where to find the links in Access Structures to other databases that the CCDC collaborates with. This allows users easy access to a wealth of additional information from other resources.







Example 4. Downloading

Access Structures allows the download of CIF files, checkCIF reports and gcd files.

The Crystallographic Information File (CIF) is a standard format for information interchange in crystallography. It is a well-established way of reporting crystal structure determinations. (ref: Brown, I. D. & McMahon, B. (2002). Acta Cryst. B58, 317-324. http://dx.doi.org/10.1107/S0108768102003464)

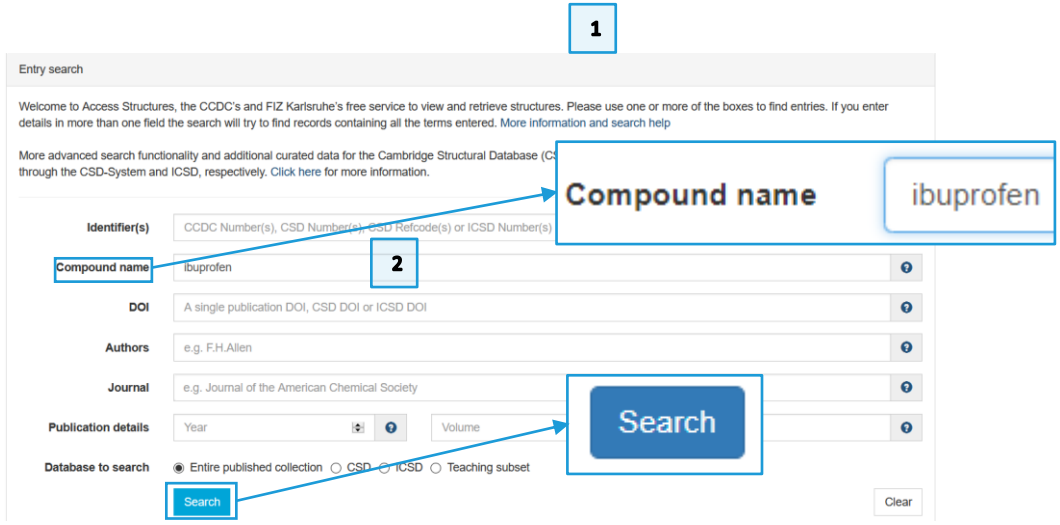
checkCIF is an online service offered by the International Union of Crystallography (IUCr) for checking structural data. (ref: Spek, A. L. (2009). Acta Cryst. D65, 148-155 http://dx.doi.org/10.1107/S090744490804362X)

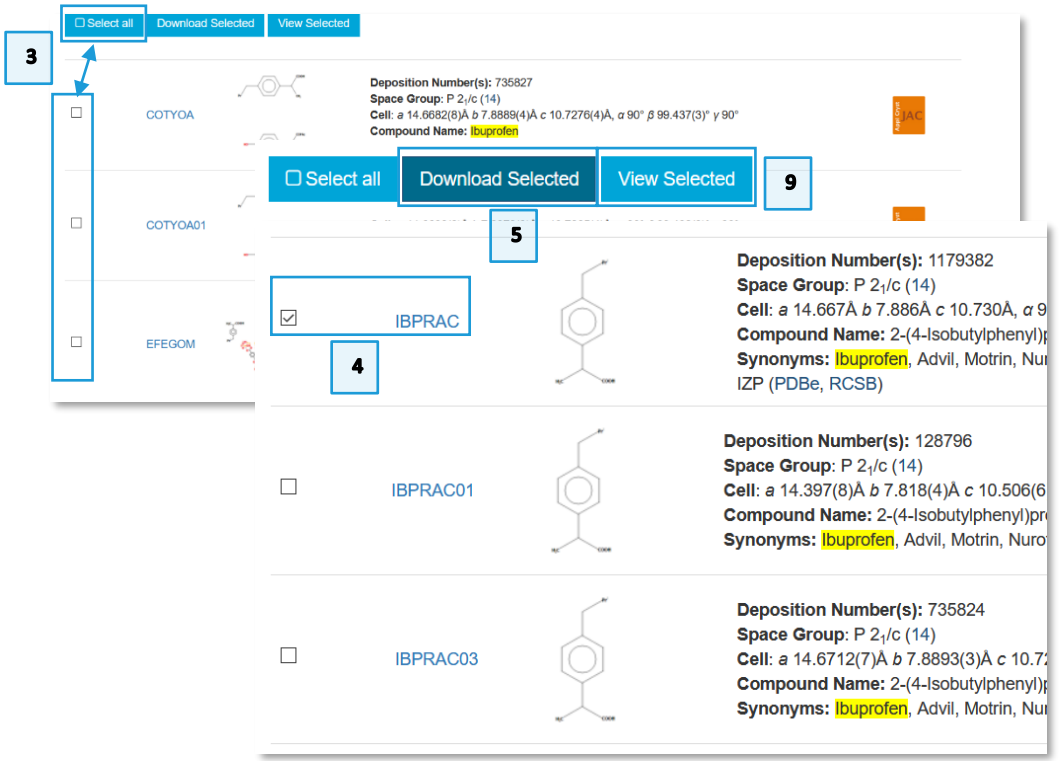
Downloading and viewing these files give users access to in-depth structural information.

Download options from initial results page

These steps will show you how to download one or more different CIF files from the initial results page in Access Structures.

1. Open your web browser and navigate to www.ccdc.cam.ac.uk/structures
2. Into the ***Compound name*** search field type “Ibuprofen” and click the **Search** button.
3. Your search will return more than 30 records. Notice that the **Select all** box is automatically ticked. Untick the **Select all** button to deselect all the structures in the hit list.
4. Scroll down the page to find “IBPRAC” and “IBPRAC15”. Tick the boxes by each of these refcodes to select them.
5. Click **Download Selected**.

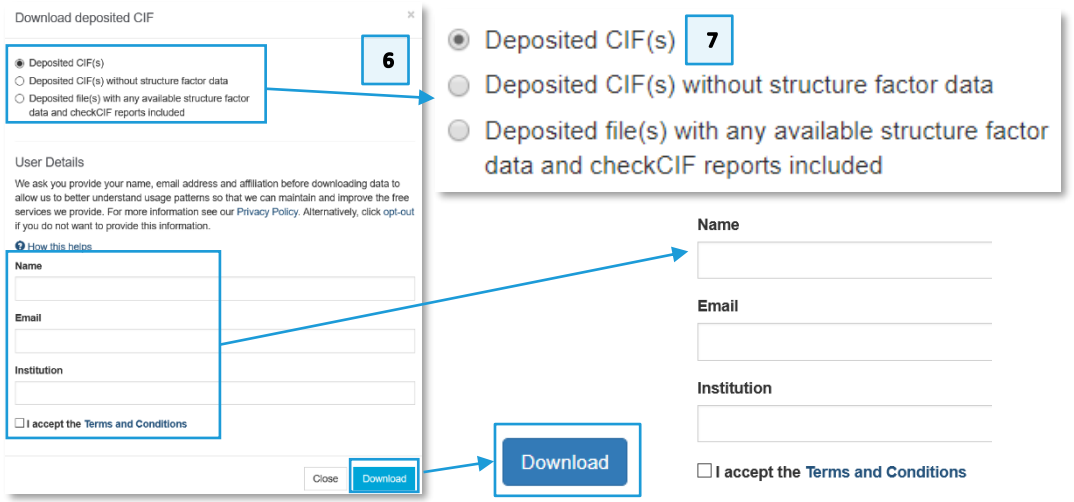




6. You then have 3 options for downloading the deposited CIFs for the 2 structures selected (IBPRAC and IBPRAC15); (i) the deposited CIFs, (ii) the deposited CIF(s) without structure factor data or (iii) the deposited files with any available structure factor data as well as any included checkCIF reports.

7. Select “Deposited CIF(s)”, enter your **Name**, **Email** and **Institution** details, tick the box to accept the terms and conditions and click the **Download** button.

8. You now have a CIF file containing both structures saved to your specified downloads location. You can open, view and edit these files in software programmes such as Encifer, Mercury or Notepad. Compare the differences between the two data sets. These determinations are particularly interesting as IBPRAC was deposited in 1974 and IBPRAC15 was deposited in 2015. The dictionary of CIF terminology has expanded over time.



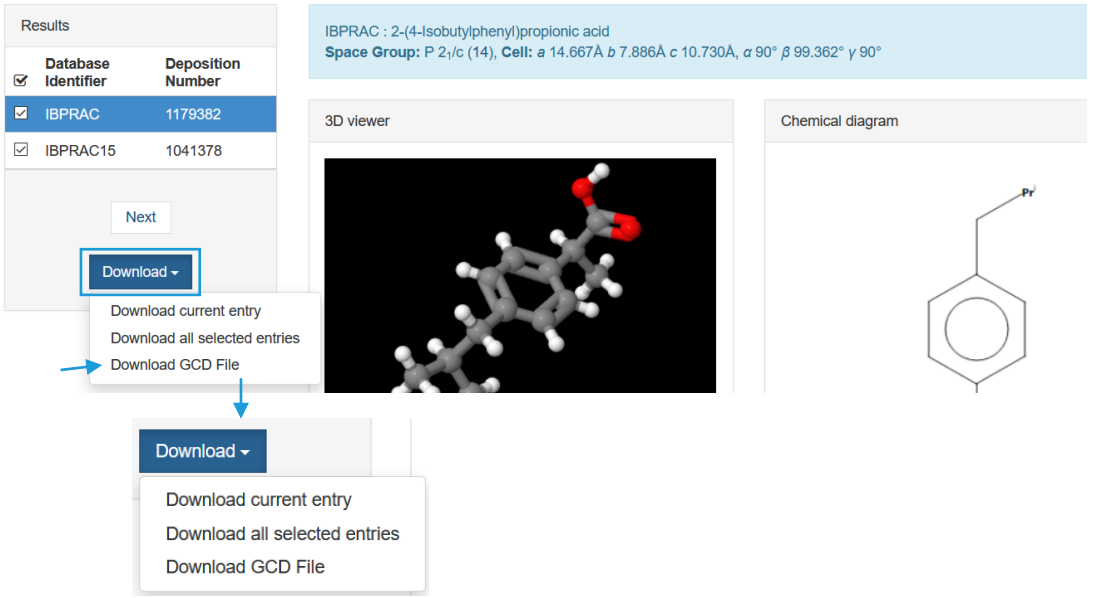
Download options from detailed results page

9. Return to the Access Structures results page tab in your browser and click the **View Selected** button (see above).

10. On the left you can see that both structures are selected. Click the drop-down arrow on the **Download** button.

11. Now you can see an extra option to download the GCD file. Click on this option.

12. This downloads a text file containing a list of the refcodes of the selected entries. This can be opened in software programs such as Mercury and Conquest.



Citing the CSD

If you use data from the CSD in your research or publication that is not your data, you should provide a reference to the data. For example, when referring to IBPRAC, you should refer to the structure by its refcode in the style: CSD-REFCODE e.g CSD-IBPRAC. You should also cite the original publication where the structure

was published in your references section. This can be found in the associated publications section on the results page for the structure. The preferred general citation for the CSD is:
The Cambridge Structural Database C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, Acta Cryst. (2016). B72, 171-179 10.1107/S2052520616003954

The CSD web interface can be cited as:
CCDC (2017). CSD web interface – intuitive, cross-platform, web-based access to CSD data. Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, UK.

Conclusion

This example shows the different download options available and how to cite data obtained via Access Structures.

Conclusions

This workshop introduced our Access Structures web based. You should now be familiar with:

- Conducting various text-based searches in access structures
- Visualizing structures and measuring bond lengths and angles
- Exploring links to other databases
- Downloading cifs of structures

Next steps

You can check out other CSD-Community based workshop materials here

Feedback

We hope this workshop improved your understanding of Access Structures 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 WCSD-001. Thank you!