CSD Tutorial

Suzanna Ward and Eric Rogers

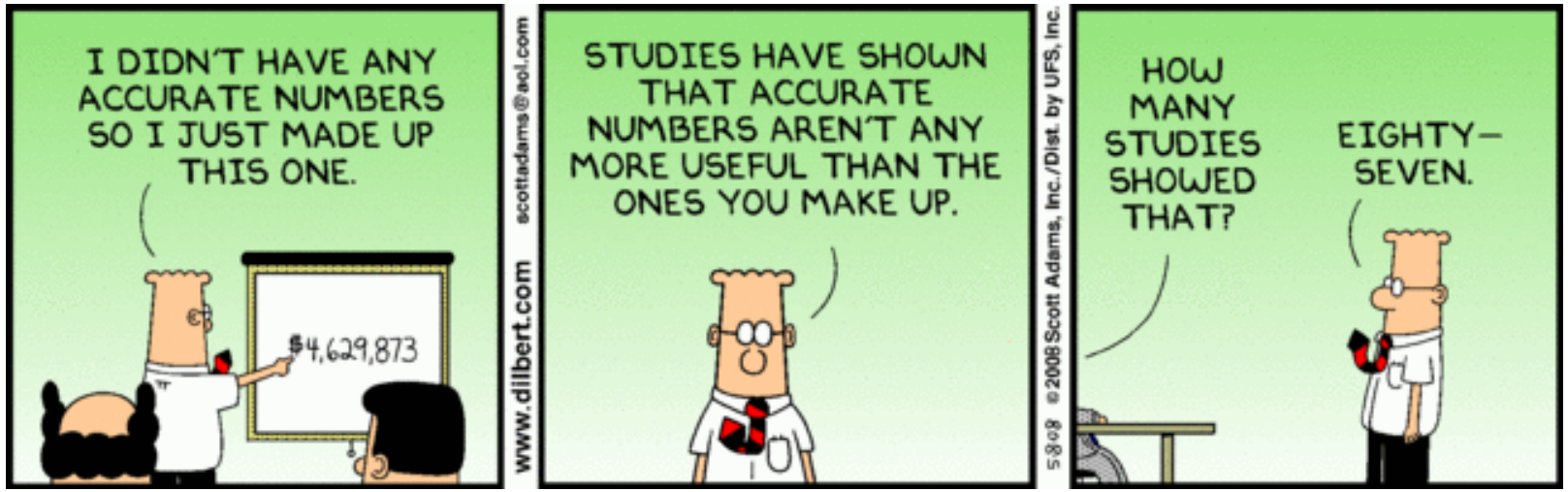
The Cambridge Crystallographic Data Centre

What is a crystallographic database?

- A structured collection of data about crystals and crystal structures
- Typically containing information on:
  - Chemical composition, position of atoms, unit cell dimensions, space group, experimental parameters and bibliographic details
  - Plus more information about structures if relevant to the scientific area
- What makes crystallography special?
  - Standard agreed file format
  - *Every* published structure is added to the appropriate database
  - Established *curated* databases rather than just collections of data
- A database implies some level of quality control and some level of processing & validation to the structures
  - Allowing you to learn from the data without having to re-process it

Importance of data quality

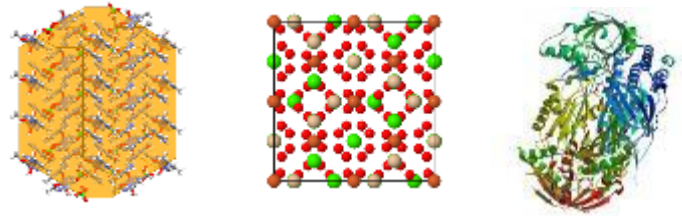
- The ability of a database search to answer a question is limited by the quality of the data



- Not just the quality of the diffraction experiment & the refinement, but also the CIF & data included!

Before we start….

- What kind of crystallography are you interested in?
  - Organic or metal-organic
  - Inorganic
  - Protein



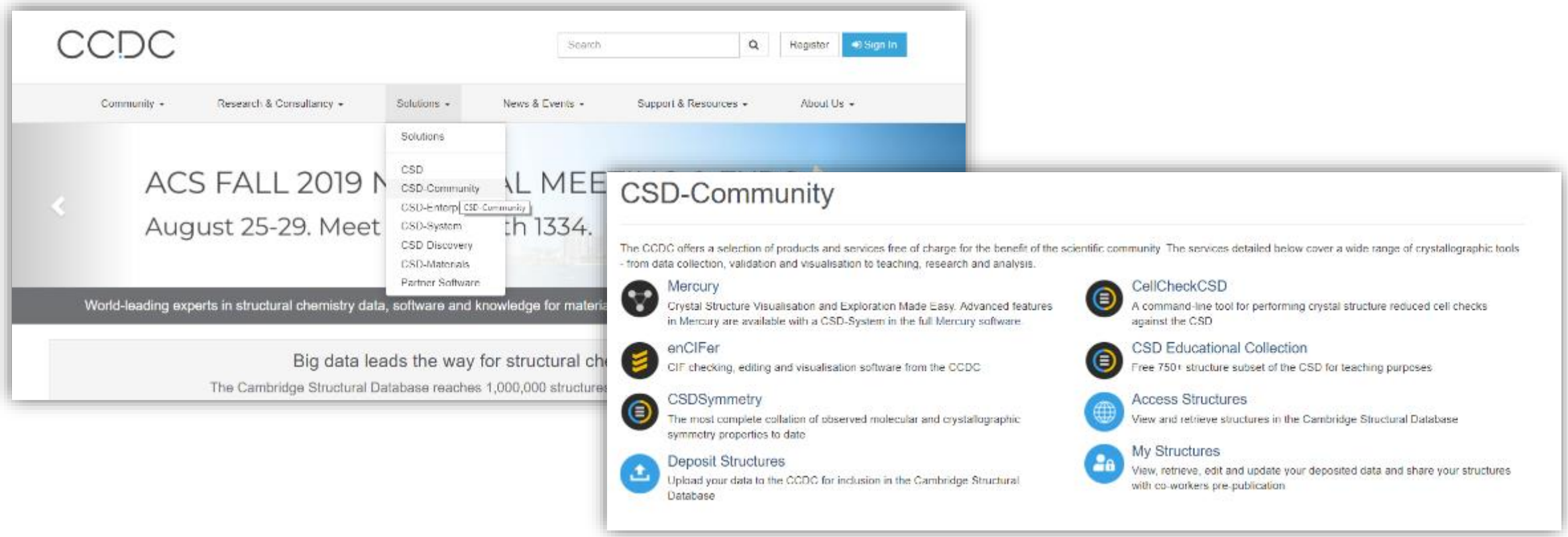
- Who had heard of the CSD?
- Who has searched the CSD online?
- Who has used desktop software such as Mercury, Conquest or Mogul to look at the CSD?
- Who has determined their own crystal structure?
- Who has used our deposition service?
- Who has got their own structure in the CSD?

What do you need for today?

- Who still needs the CSD installed on their laptop?
- Who still needs the workshop materials and example files?

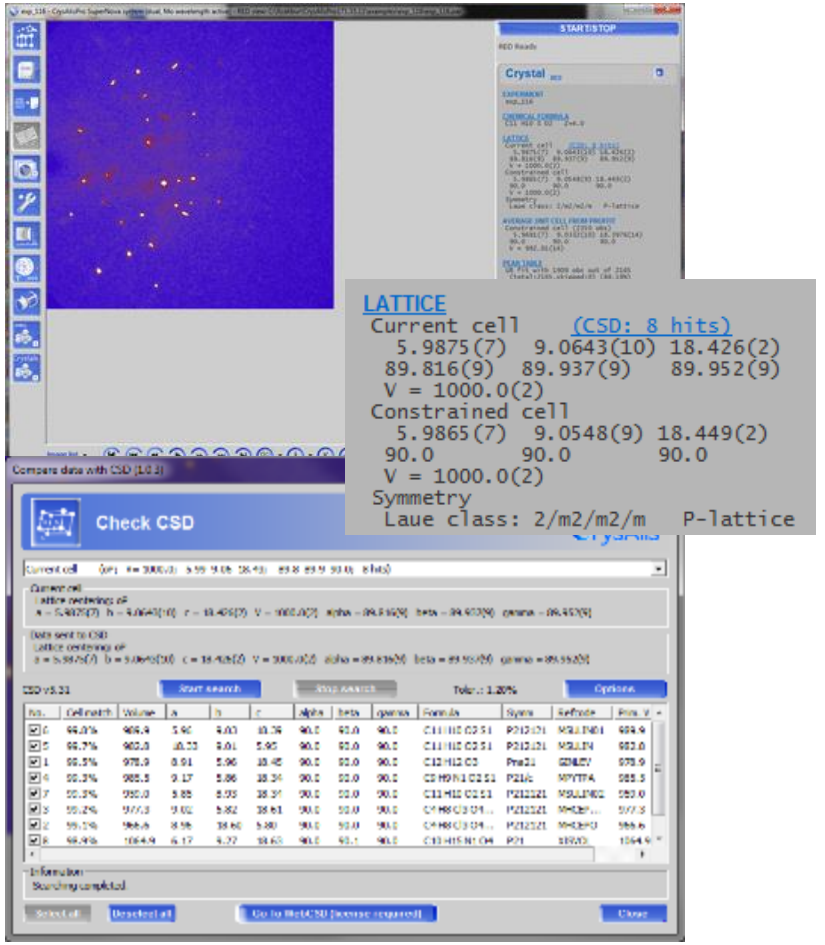
How you can access our tools?

- www.ccdc.cam.ac.uk



CSD-Community - CellCheckCSD

- Freely available through CSD-Community
- An automated tool for checking unit cells against the CSD during data collection
- Used to:
  - Match existing cell dimensions
  - Check sample is novel
  - Check crystal is not starting material or a by-product
  - Ensure diffractometer time used effectively



Data Deposition

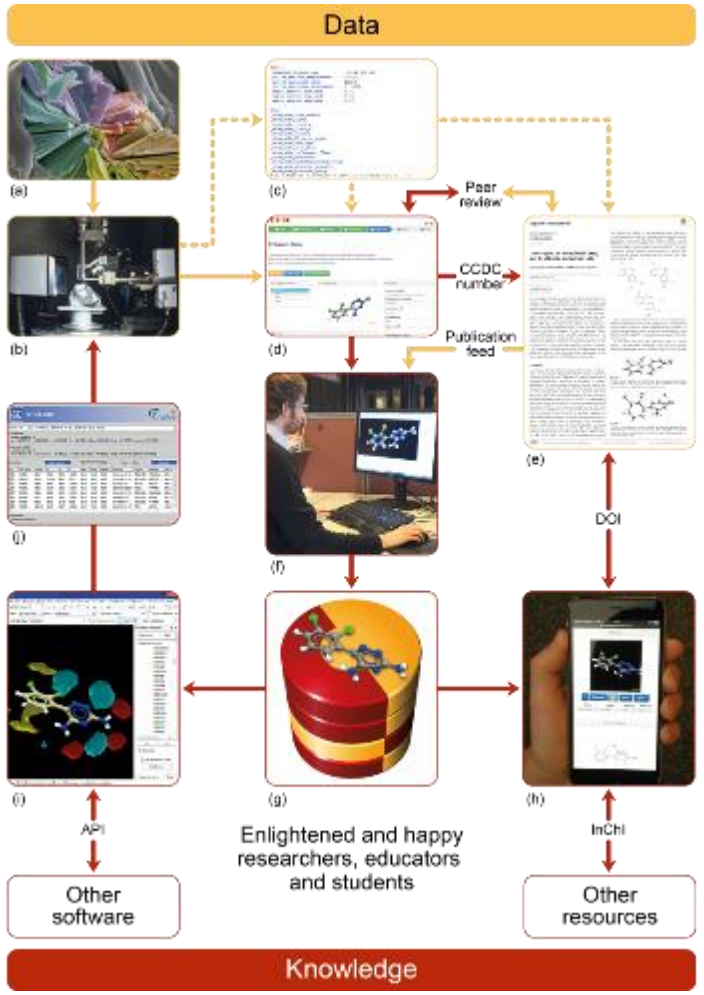
Deposition

- Data deposited pre-publication
  - When you know most about your structure
  - Enables links to data sets at the point of publication
  - Tailored deposition service

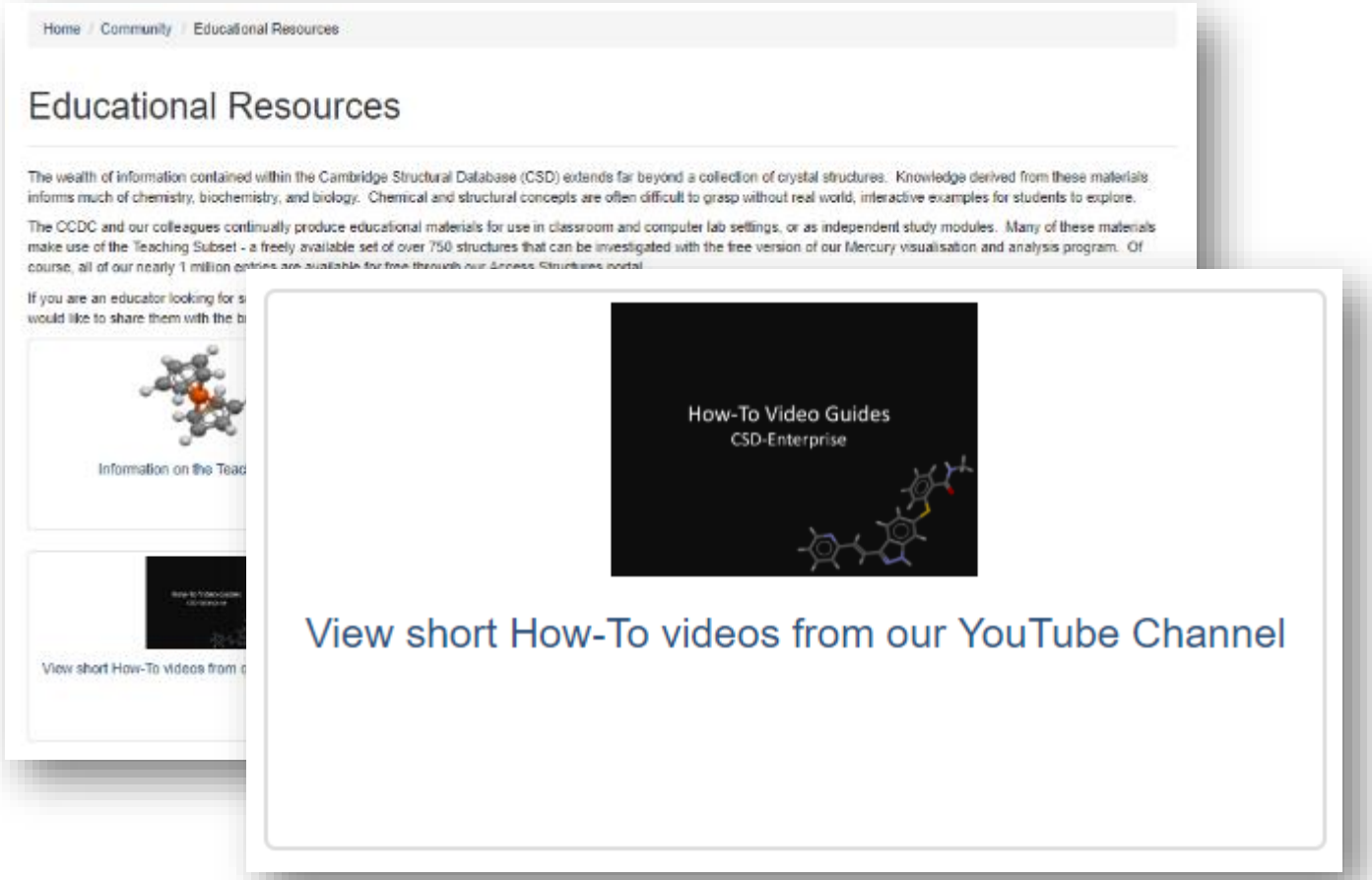
Small molecule single crystal data

Authors should present their crystal data in a CIF (Crystallographic Information File) format and deposit any organic or organometallic structural information with the Cambridge Crystallographic Data Centre (CCDC) before they submit their manuscript to us. Data will be held in the CCDC's confidential archive until publication of the article, when data for organic and organometallic compounds will be entered into the Cambridge Structural Database. Authors are encouraged to deposit inorganic crystal structures with the ICSD, hosted by FIZ Karlsruhe.

During submission of a manuscript to the Royal Society of Chemistry using our online submission system, authors will be asked to provide CCDC reference numbers; CIFs should not be submitted with the manuscript (these should have already been deposited with the CCDC/ICSD, see above). Any revised CIFs obtained subsequently should be deposited directly with the CCDC before the revised manuscript is submitted to us. CCDC or ICSD numbers should be included in the manuscript prior to submission.



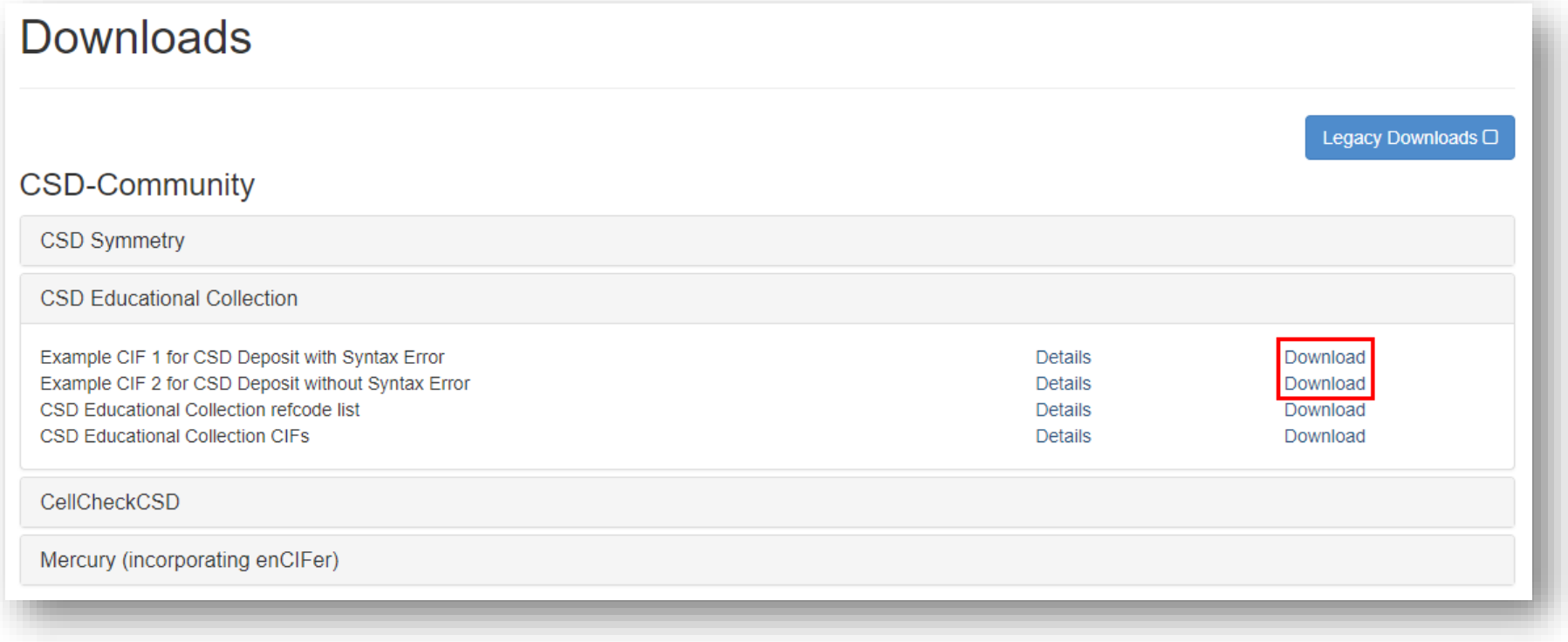
Instructional videos



https://www.ccdc.cam.ac.uk/Community/educationalresources/

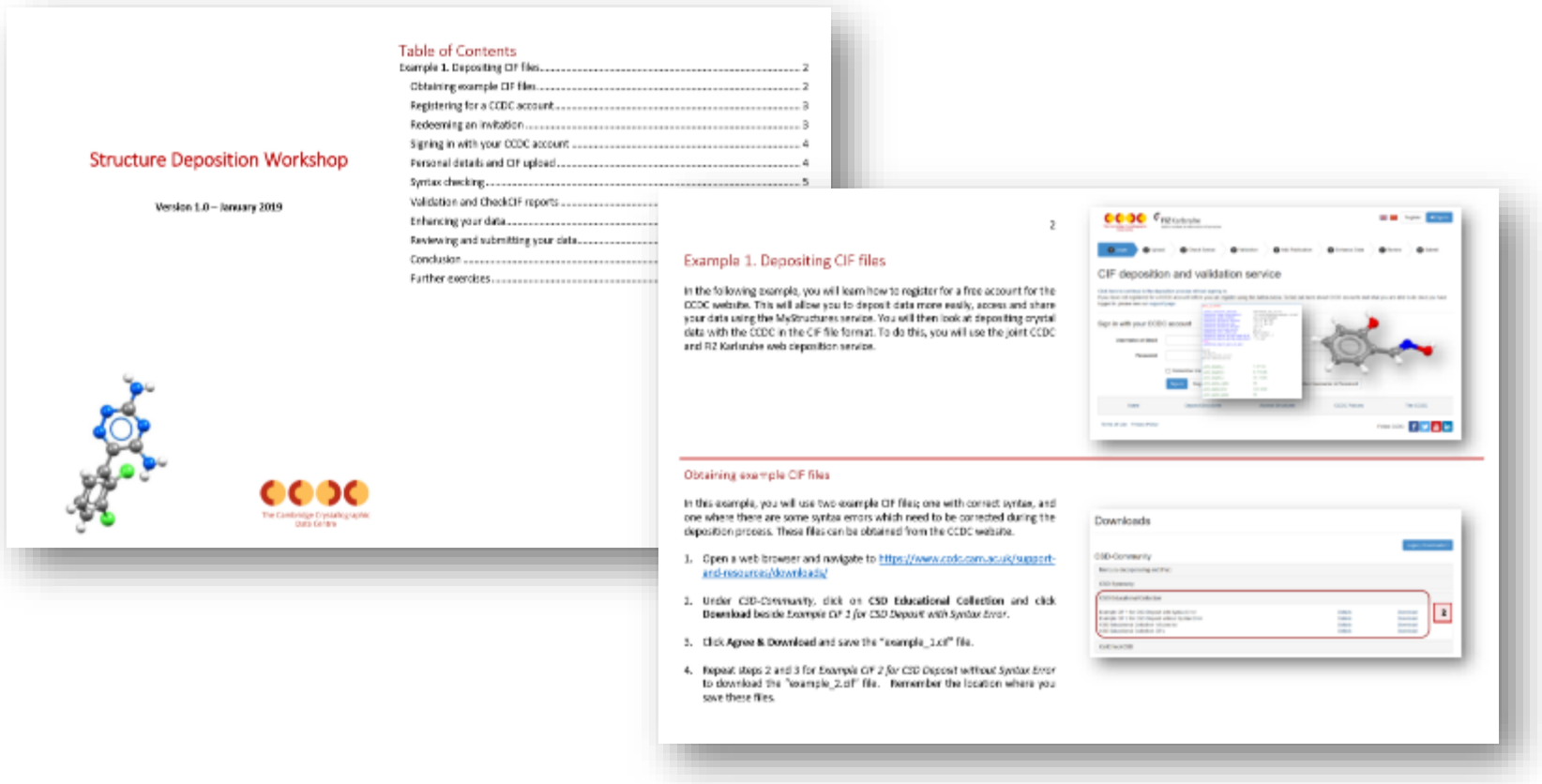
Demo Files

- https://www.ccdc.cam.ac.uk/support-and-resources/downloads

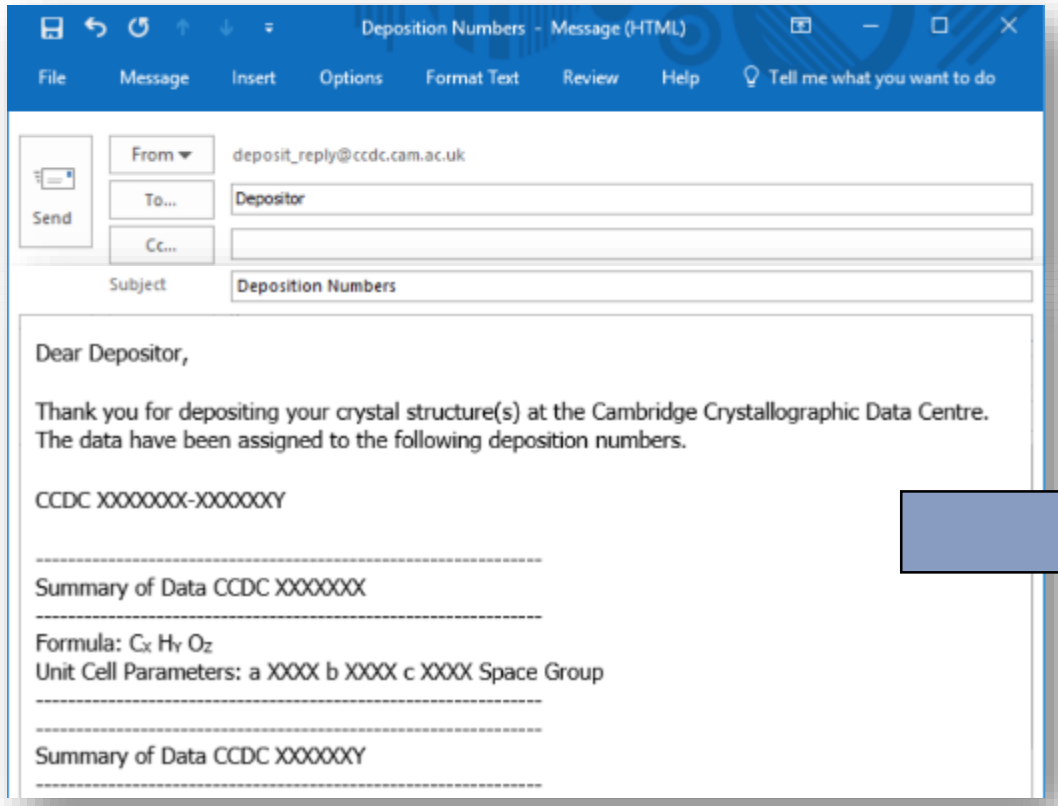


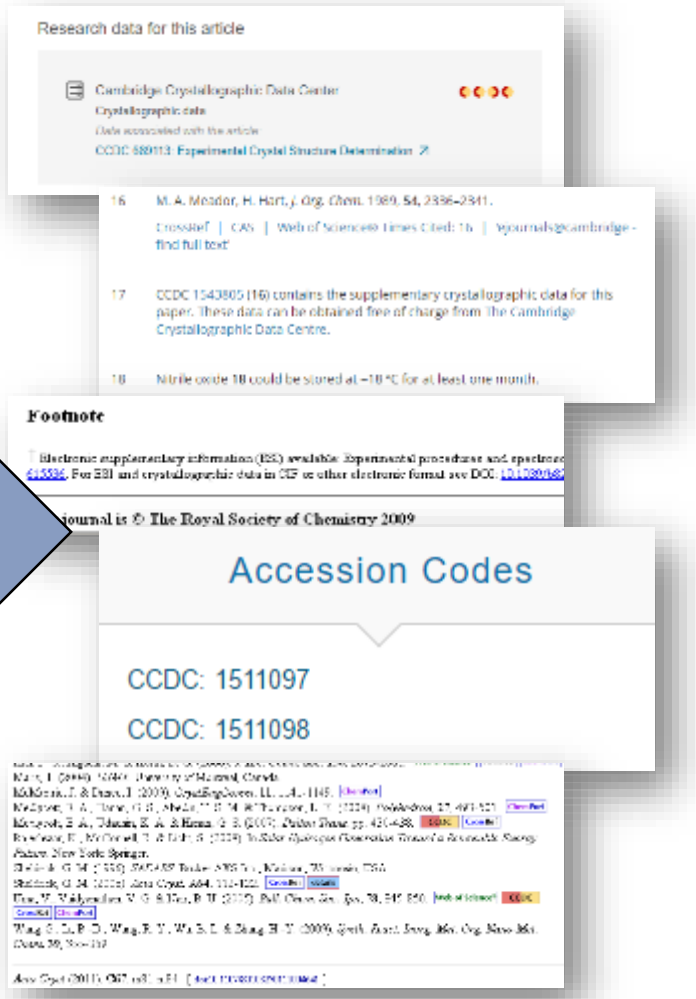
Deposition Tutorial

- https://www.ccdc.cam.ac.uk/deposit/



Dataset identifiers

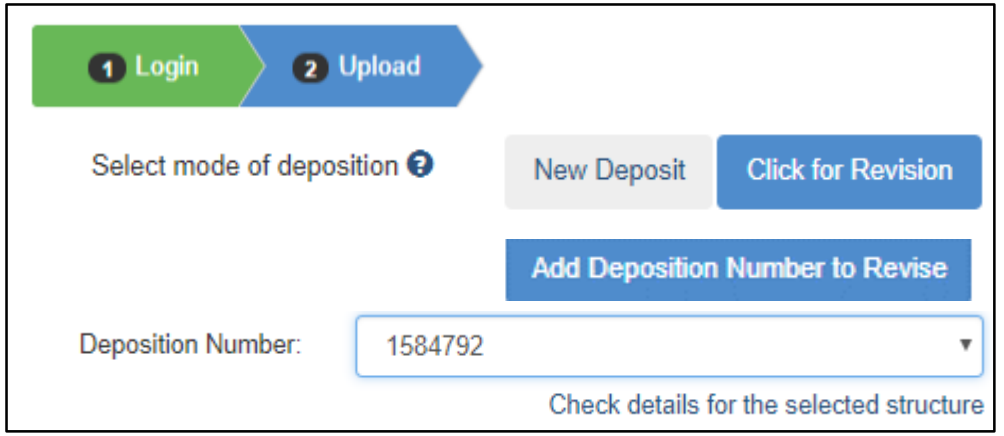




Identifier sent after deposition so it can be included in a manuscript

What if I want to revise my data?

- You can revise your deposited data prior to publication
- During deposition:
    1. **If logged in:**
        a) ‘Click for Revision’
        b) ‘Add Deposition Number to Revise’
        c) Select Deposition Number you wish to revise
    2. **If not logged in:**
        a) Add Deposition Number to ‘Deposition Number(s) for revision box

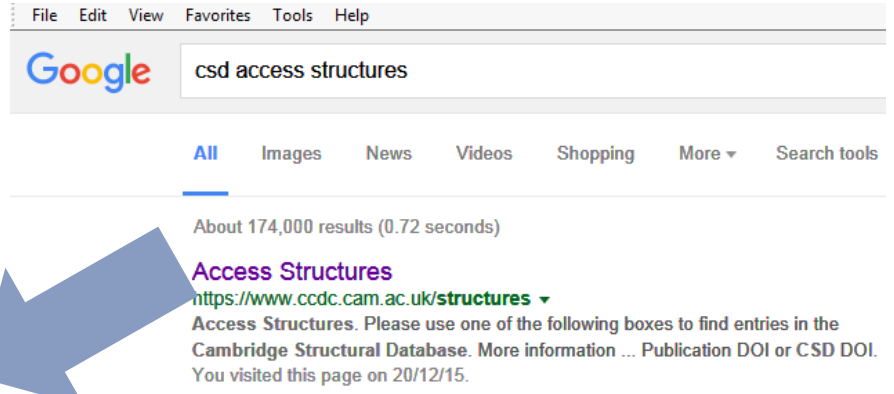


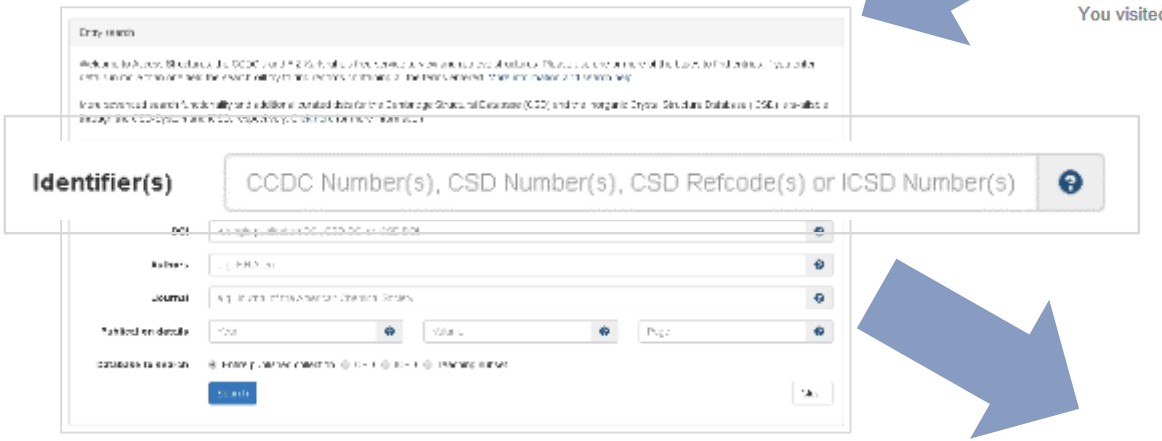
- Revised dataset will be assigned the same Deposition Number
- Old revisions are stored but not visible

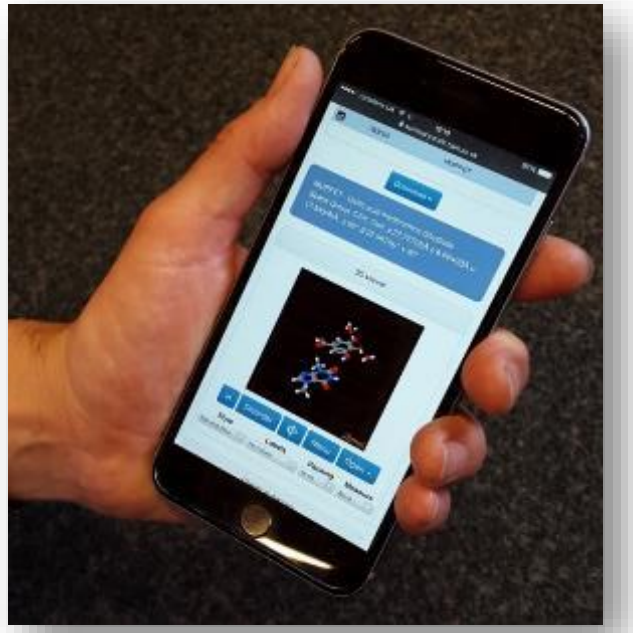
Data Access

Every individual structure is accessible……

- Google CSD Access Structures
- Got to CCDC website
- Search and view

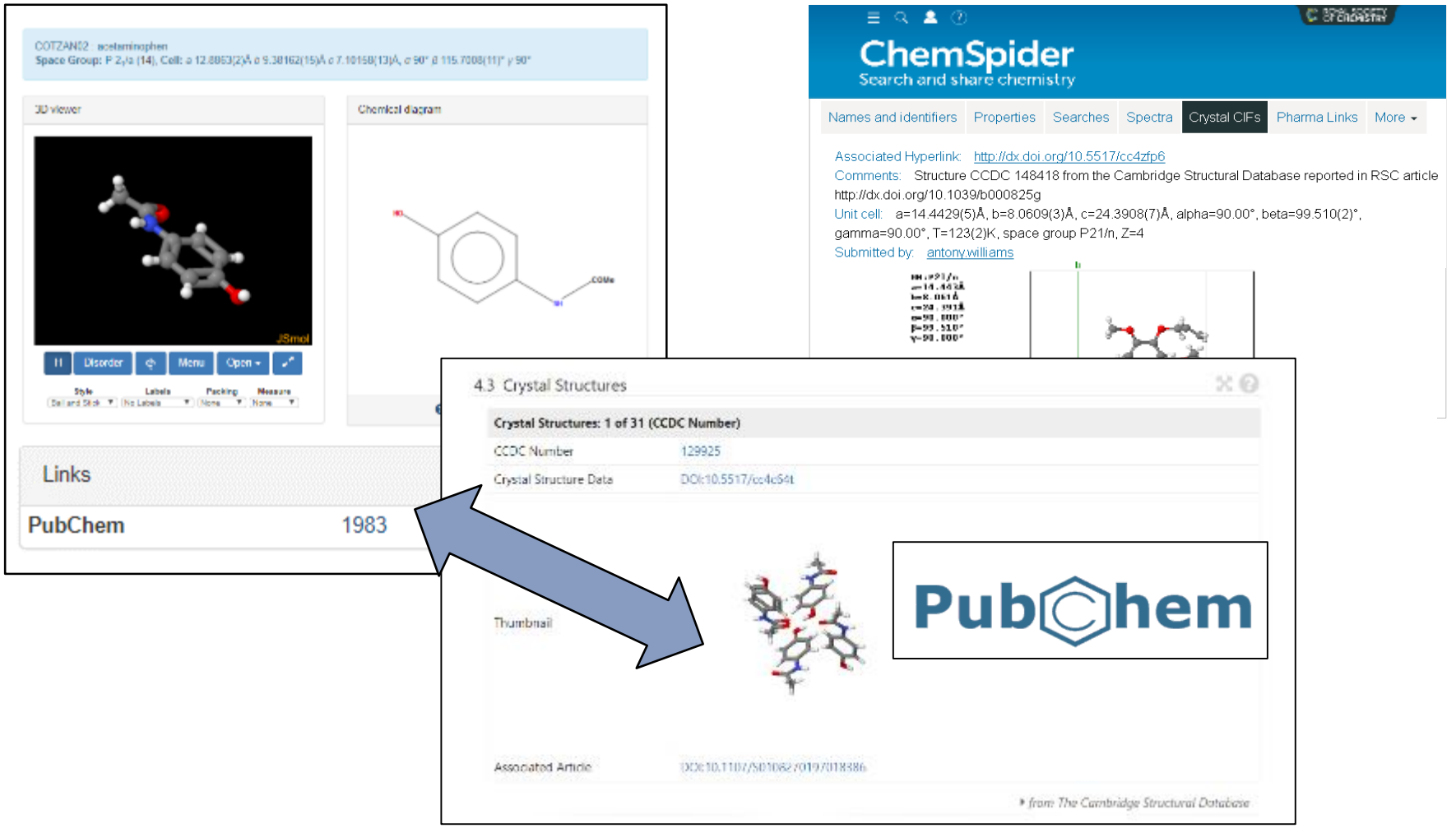






- Some of my favourite CSD Refcodes:
  - KITTEN, DISNEY, TURKEY, BADBOY, ITALIA
- Tell us yours!

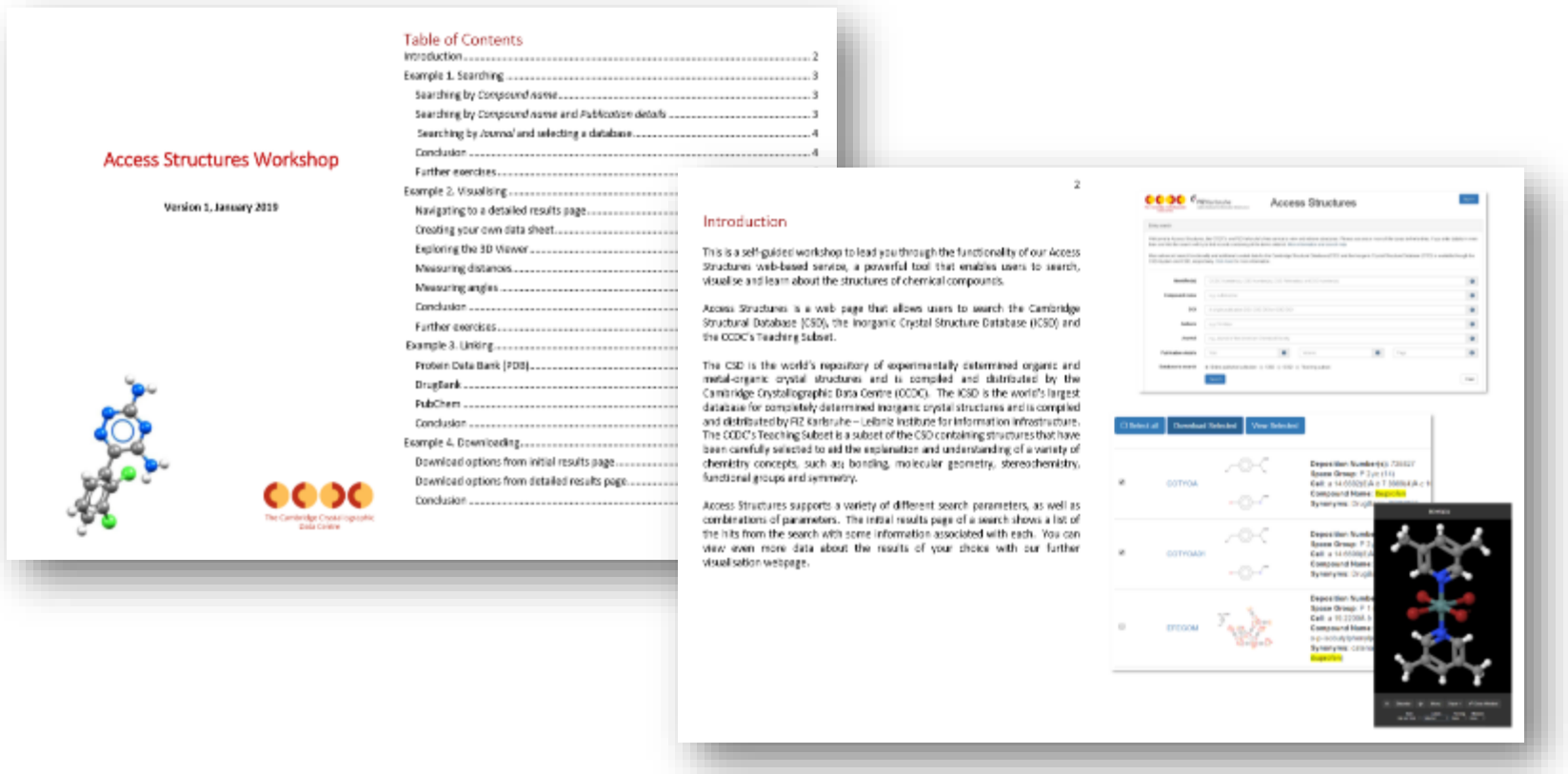
Accessing data from general resources



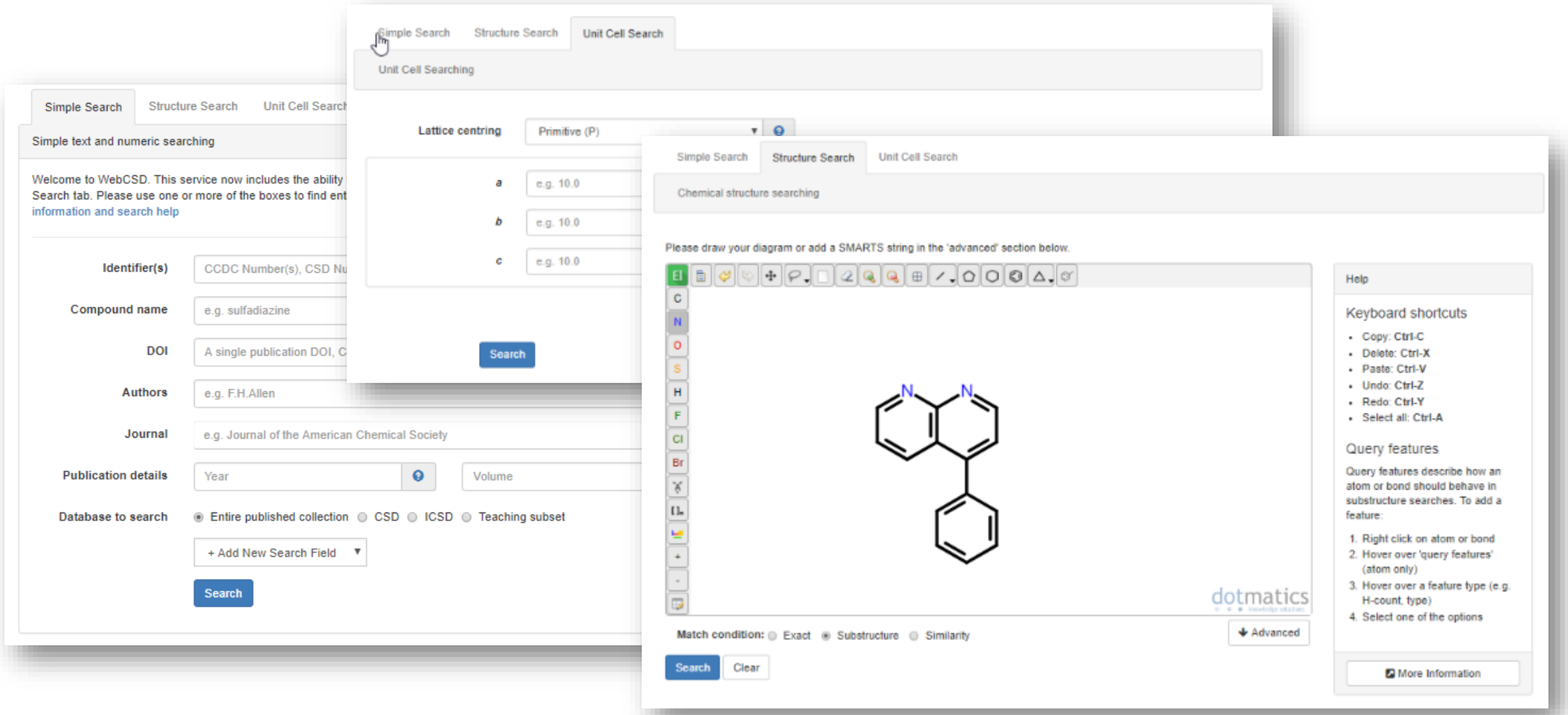
https://pubchem.ncbi.nlm.nih.gov/compound/1983#section=Crystal-Structures

Access Structures Tutorial

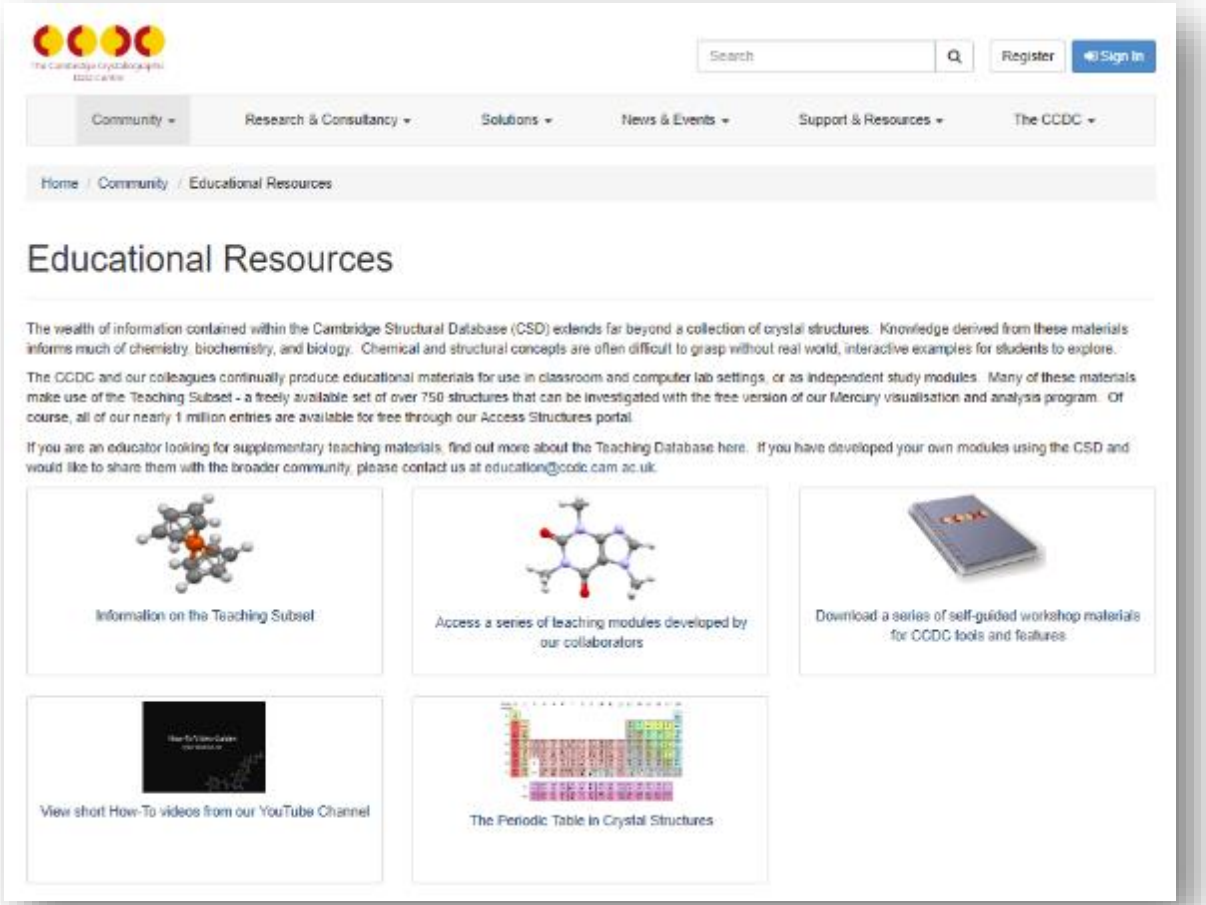
- http://www.ccdc.cam.ac.uk/structures

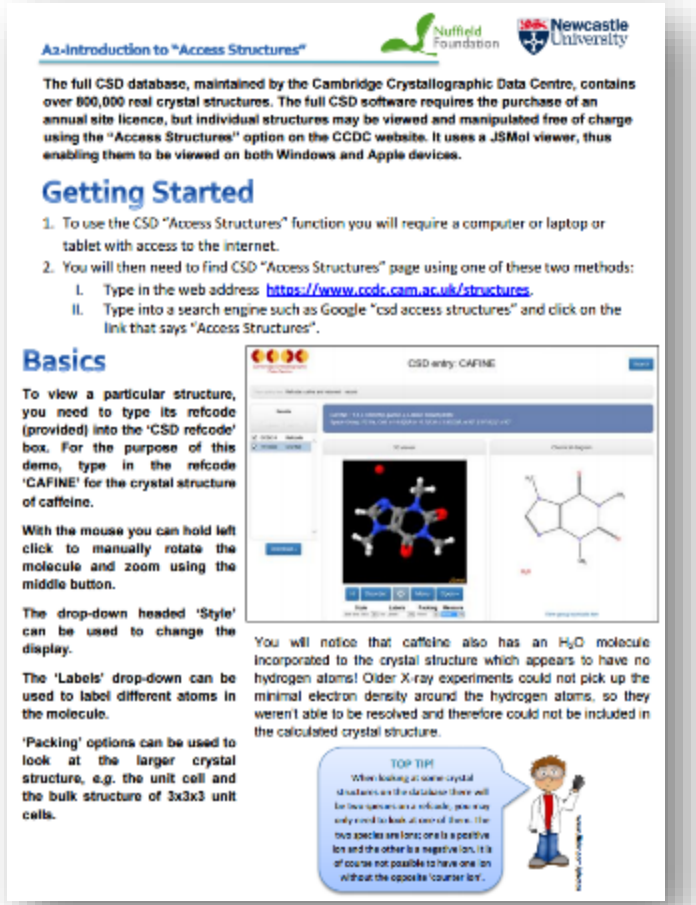


WebCSD


Simple Search
Structure Search
Unit Cell Search
Simple text and numeric searching
Welcome to WebCSD. This service now includes the ability
Search tab. Please use one or more of the boxes to find ent
information and search help
Identifier(s)
CCDC Number(s), CSD Nu
Compound name
e.g. sulfadiazine
DOI
A single publication DOI, C
Authors
e.g. F.H.Allen
Journal
e.g. Journal of the American Chemical Society
Publication details
Year
Volume
Database to search
Entire published collection
CSD
ICSD
Teaching subset
+ Add New Search Field
Search
Simple Search
Structure Search
Unit Cell Search
Unit Cell Searching
Lattice centring
Primitive (P)
a
e.g. 10.0
b
e.g. 10.0
c
e.g. 10.0
Search
Simple Search
Structure Search
Unit Cell Search
Chemical structure searching
Please draw your diagram or add a SMARTS string in the 'advanced' section below.
El
C
N
O
S
H
F
Cl
Br
dotmatics
Match condition:
Exact
Substructure
Similarity
Advanced
Search
Clear
Help
Keyboard shortcuts
Copy: Ctrl-C
Delete: Ctrl-X
Paste: Ctrl-V
Undo: Ctrl-Z
Redo: Ctrl-Y
Select all: Ctrl-A
Query features
Query features describe how an atom or bond should behave in substructure searches. To add a feature:
1. Right click on atom or bond
2. Hover over 'query features' (atom only)
3. Hover over a feature type (e.g. H-count, type)
4. Select one of the options
More Information

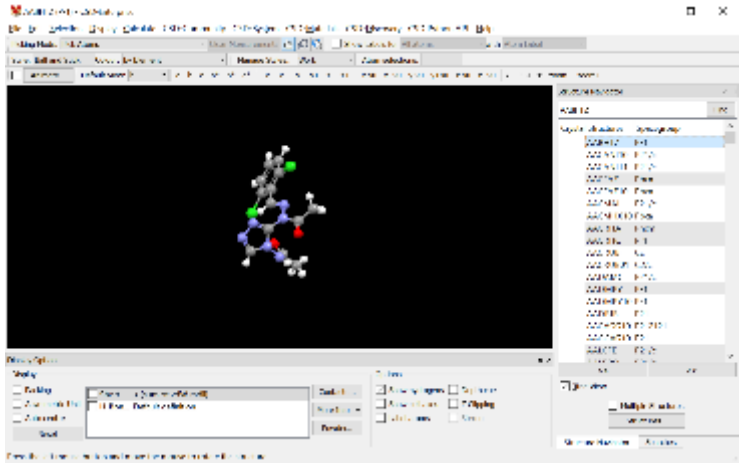
Educational resources – A teaching DB

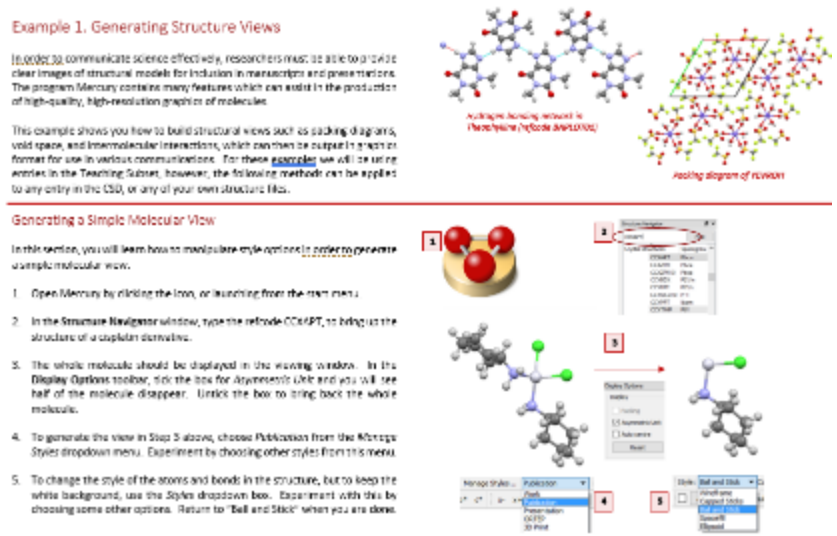


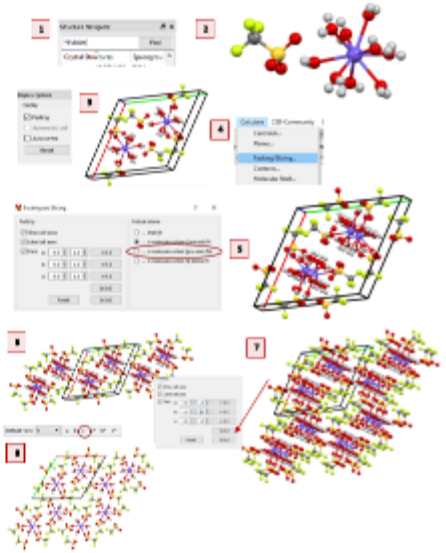


https://www.ccdc.cam.ac.uk/Community/educationalresources/

Educational resources and tutorials







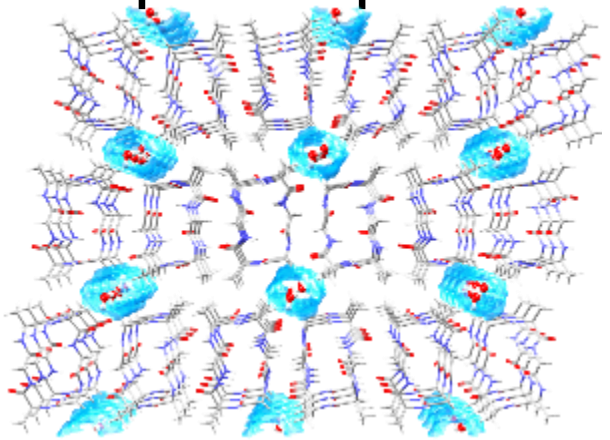
https://www.ccdc.cam.ac.uk/Community/educationalresources/workshop-materials/

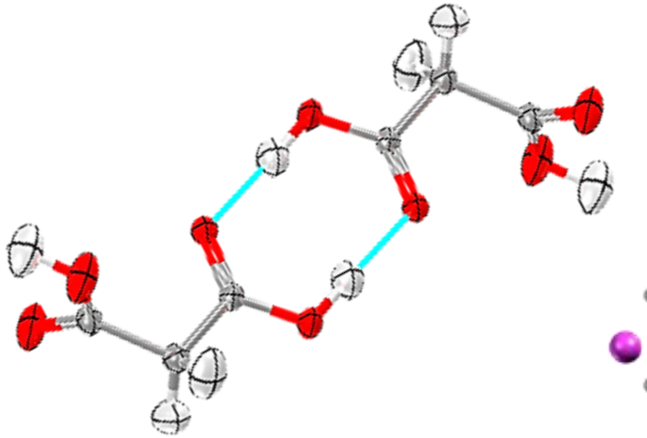
Exploring Mercury

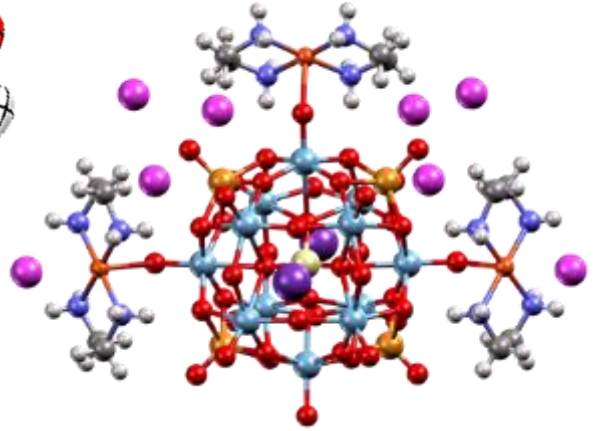
Visulalisation

Scientific communication

- The right representation can convey complex topics easily

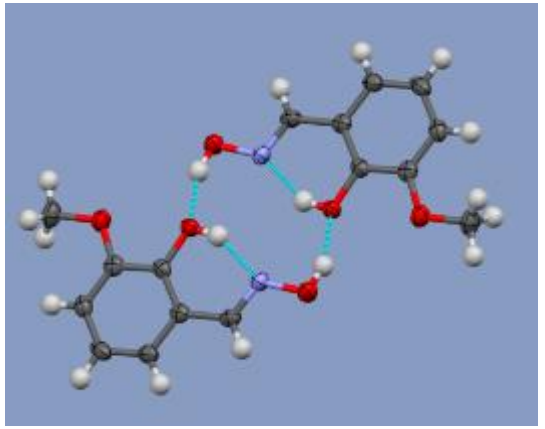


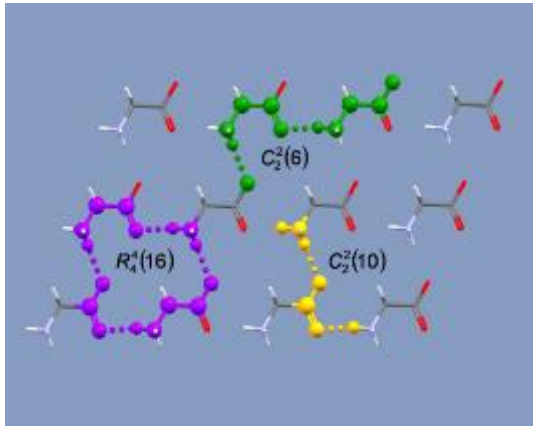


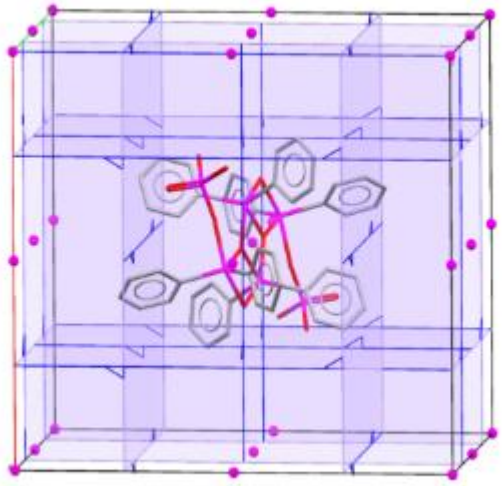


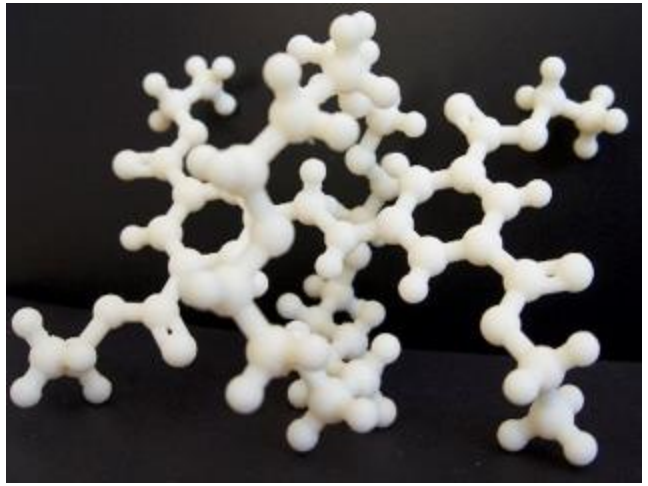
Structure visualisation

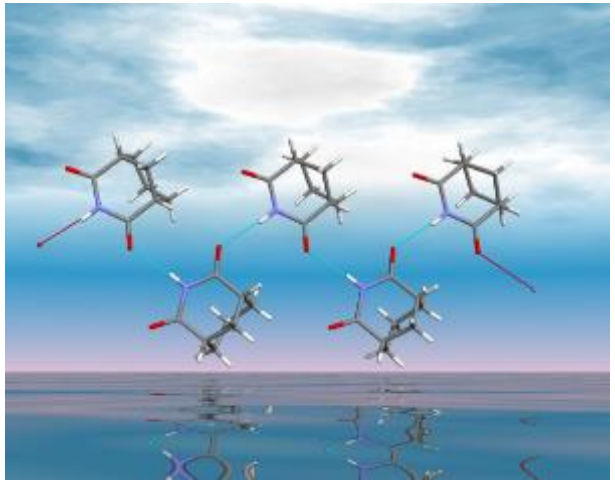
- Symmetry elements
- Displacement ellipsoids etc.
- POV-Ray
- 3D printing
- Graph sets
- Movies



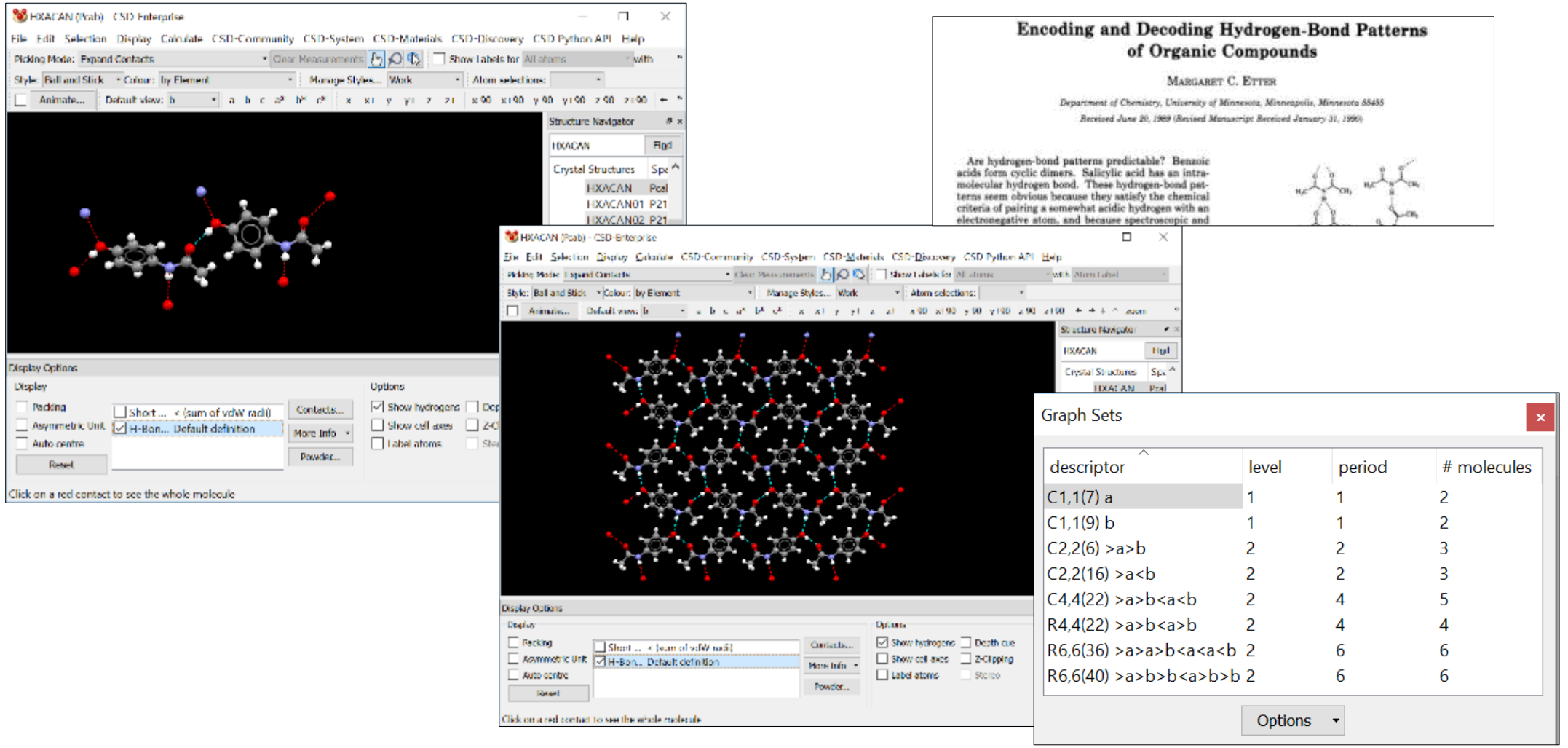




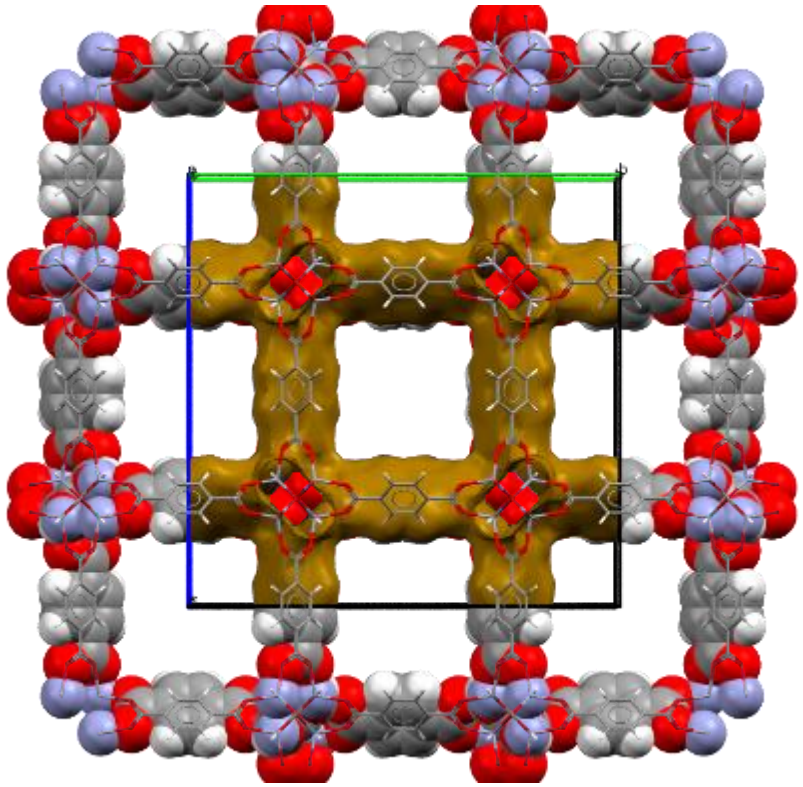




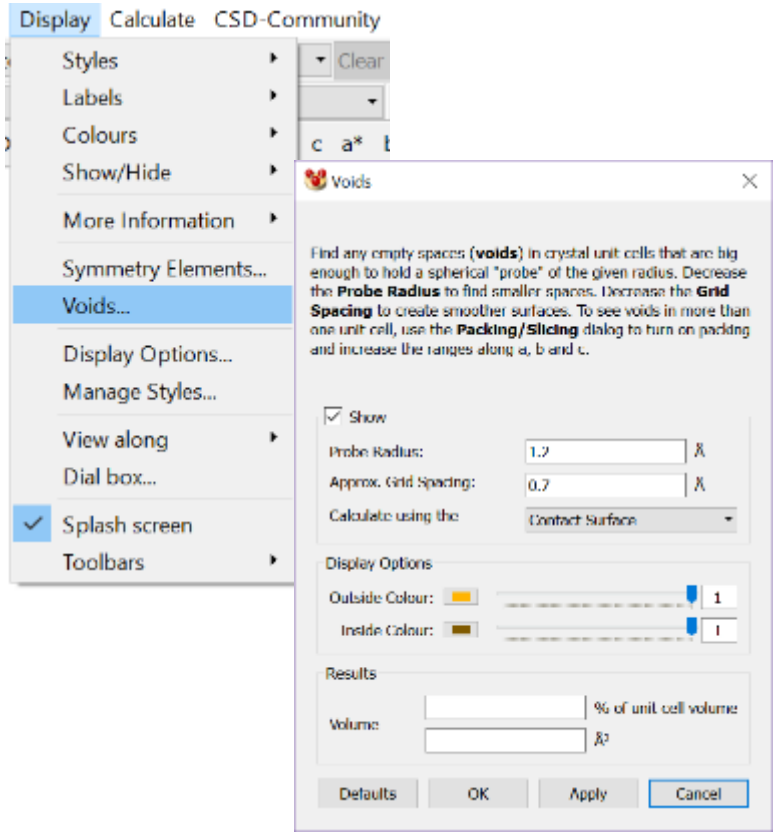
Hydrogen bonds



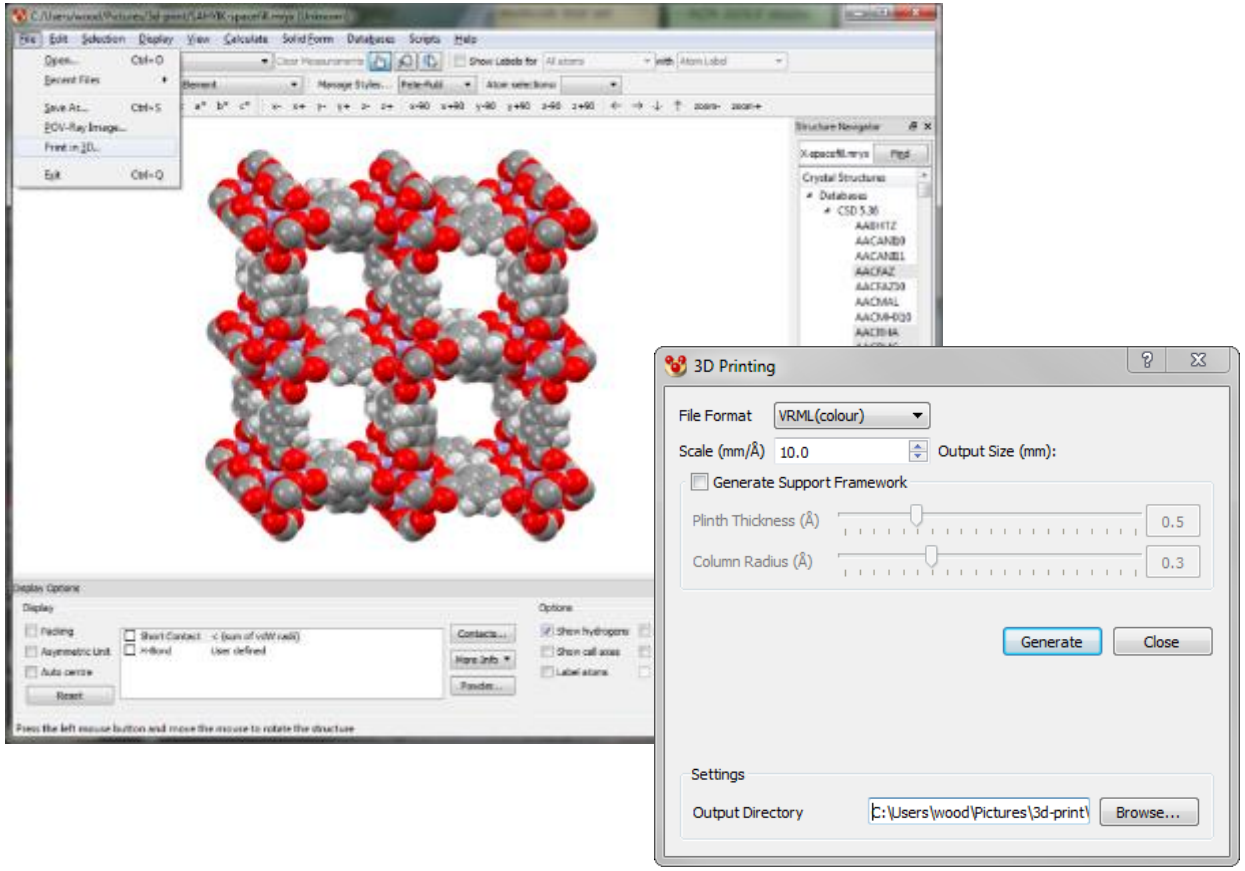
Empty space visualisation



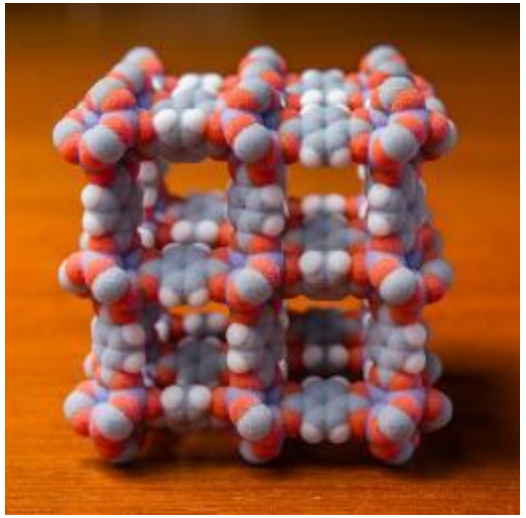
SAHYIK – MOF5



3D printing



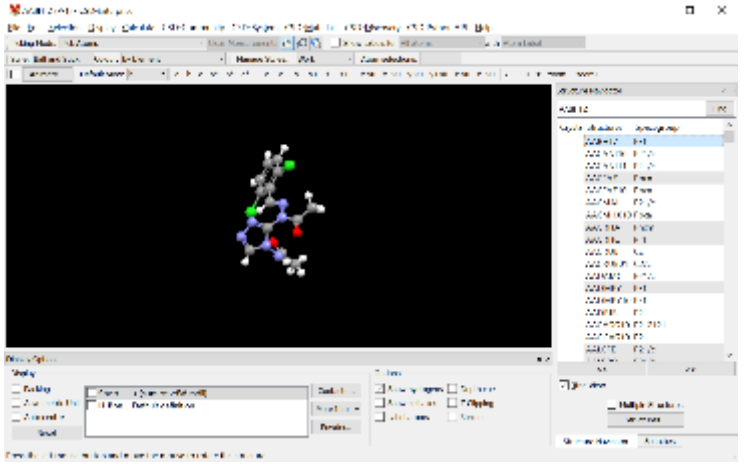
Simple interface to set up file production + 3D print style

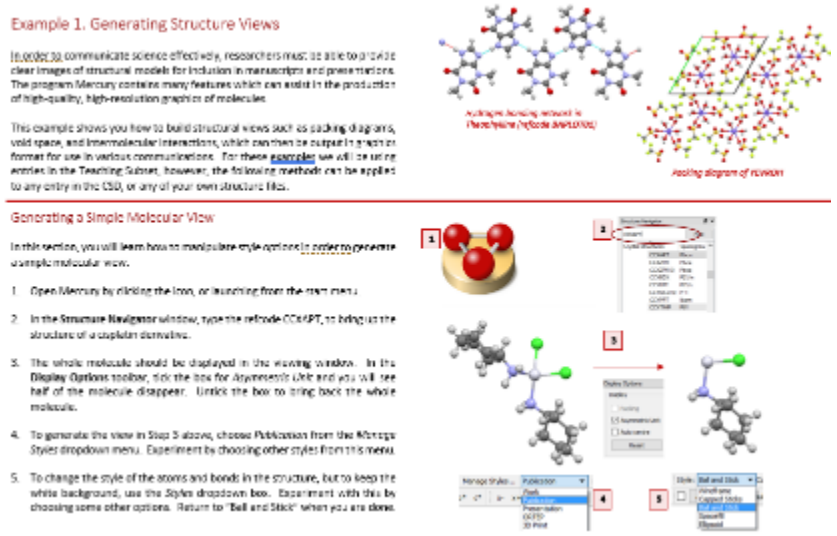


P. A. Wood, A. A. Sarjeant, et al., *CrystEngComm*, 2017, **19**, 690-698

Example 1

- Simple molecular views
- Generating packing diagrams
- Displaying intermolecular contacts
- Generating graph sets





Example 1. Generating Structure Views

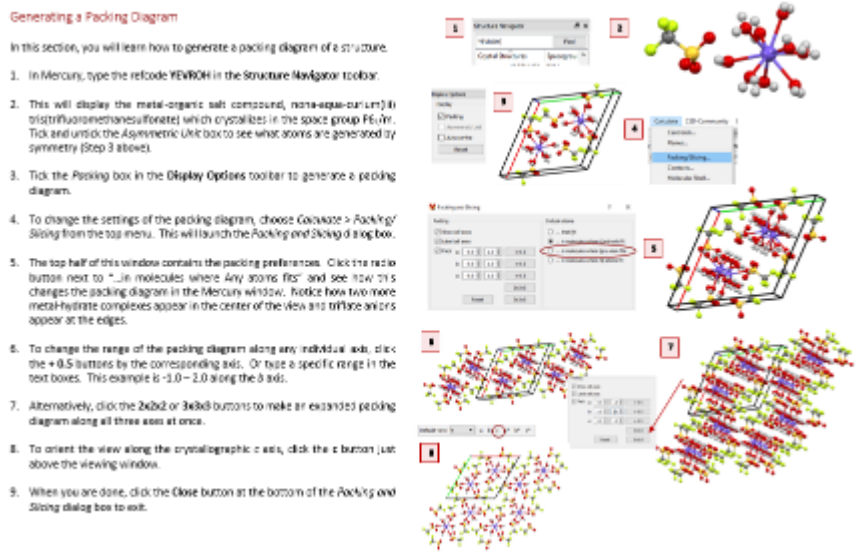
In order to communicate science effectively, researchers must be able to provide clear images of structural models for inclusion in manuscripts and presentations. The program Mercury contains many features which can assist in the production of high-quality, high-resolution graphics of molecules.

This example shows you how to build structural views such as packing diagrams, void space, and intermolecular interactions, which can then be output in graphics format for use in various communications. For these exercises we will be using entries in the Teaching Subset, however, the following methods can be applied to any entry in the CSD, or any of your own structure files.

Generating a Simple Molecular View

In this section, you will learn how to manipulate style options in order to generate a simple molecular view.

1. Open Mercury by clicking the icon, or launching from the start menu.
2. In the **Structure Navigator** window, type the refcode CCRAPT, to bring up the structure of a cisplatin derivative.
3. The whole molecule should be displayed in the viewing window. In the **Display Options** toolbar, tick the box for *Asymmetric Unit* and you will see half of the molecule disappear. Untick the box to bring back the whole molecule.
4. To generate the view in Step 3 above, choose *Publication* from the *Manage Styles* dropdown menu. Experiment by choosing other styles from this menu.
5. To change the style of the atoms and bonds in the structure, but to keep the white background, use the *Styles* dropdown box. Experiment with this by choosing some other options. Return to "Ball and Stick" when you are done.



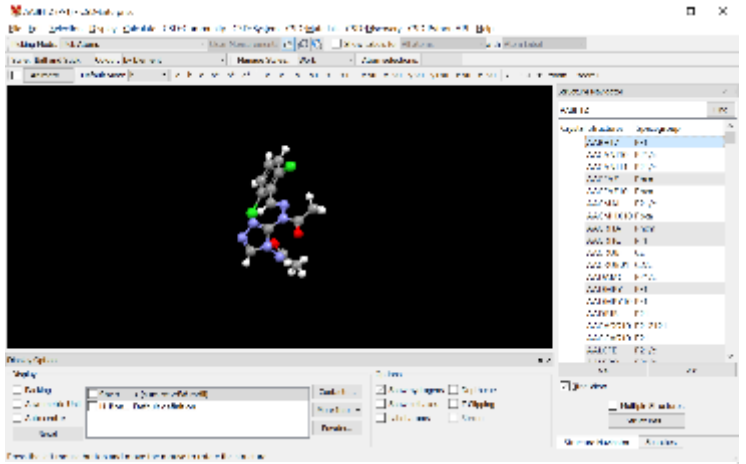
Generating a Packing Diagram

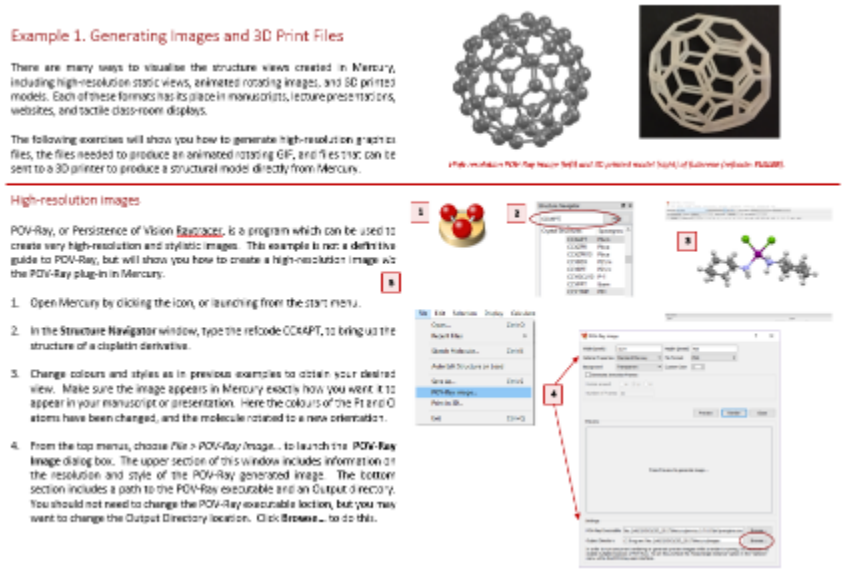
In this section, you will learn how to generate a packing diagram of a structure.

1. In Mercury, type the refcode YEVROH in the **Structure Navigator** toolbar.
2. This will display the metal-organic salt compound, nona-aqua-cerium(III) tristrifluoromethanesulfonate) which crystallizes in the space group P6$_3$/m. Tick and untick the *Asymmetric Unit* box to see what atoms are generated by symmetry (Step 3 above).
3. Tick the *Packing* box in the **Display Options** toolbar to generate a packing diagram.
4. To change the settings of the packing diagram, choose *Calculate > Packing/Slicing* from the top menu. This will launch the *Packing and Slicing* dialog box.
5. The top half of this window contains the packing preferences. Click the radio button next to "...in molecules where Any atoms fits" and see how this changes the packing diagram in the Mercury window. Notice how two more metal-hydrate complexes appear in the center of the view and triflate anions appear at the edges.
6. To change the range of the packing diagram along any individual axis, click the +0.5 buttons by the corresponding axis. Or type a specific range in the text boxes. This example is -1.0 – 2.0 along the *b* axis.
7. Alternatively, click the 2x2x2 or 3x3x3 buttons to make an expanded packing diagram along all three axes at once.
8. To orient the view along the crystallographic *c* axis, click the *c* button just above the viewing window.
9. When you are done, click the **Close** button at the bottom of the *Packing and Slicing* dialog box to exit.

Example 2

- High resolution images
- Generating rotating images
- Generating a 3D print model file





Example 1. Generating Images and 3D Print Files

There are many ways to visualise the structure views created in Mercury, including high-resolution static views, animated rotating images, and 3D printed models. Each of these formats has its place in manuscripts, lecture presentations, websites, and tactile class-room displays.

The following exercises will show you how to generate high-resolution graphics files, the files needed to produce an animated rotating GIF, and files that can be sent to a 3D printer to produce a structural model directly from Mercury.

High-resolution images

POV-Ray, or Persistence of Vision Raytracer, is a program which can be used to create very high-resolution and stylistic images. This example is not a definitive guide to POV-Ray, but will show you how to create a high-resolution image via the POV-Ray plug-in in Mercury.

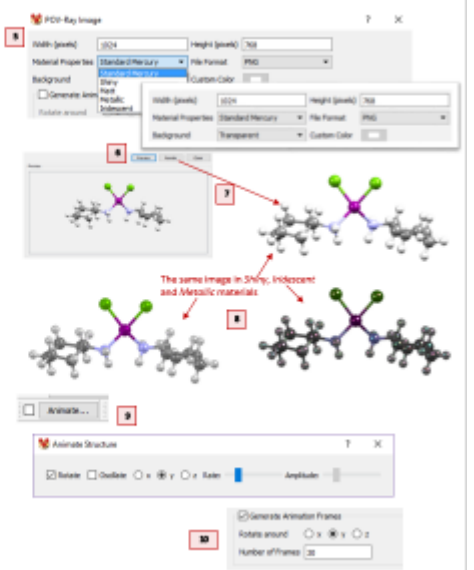
1. Open Mercury by clicking the icon, or launching from the start menu.
2. In the **Structure Navigator** window, type the refcode CCX4PT, to bring up the structure of a cisplatin derivative.
3. Change colours and styles as in previous examples to obtain your desired view. Make sure the image appears in Mercury exactly how you want it to appear in your manuscript or presentation. Here the colours of the Pt and O atoms have been changed, and the molecule rotated to a new orientation.
4. From the top menus, choose *File > POV-Ray Image...* to launch the **POV-Ray Image** dialog box. The upper section of this window includes information on the resolution and style of the POV-Ray generated image. The bottom section includes a path to the POV-Ray executable and an Output Directory. You should not need to change the POV-Ray executable location, but you may want to change the Output Directory location. Click **Browse...** to do this.

5. Change the *Material Properties* to the desired setting by clicking in the dropdown box and clicking the appropriate type. Select *Transparent* from the *Background* dropdown menu. Leave all other default values the same.
6. For complicated or large images, or on slow computers, POV-Ray may take a while to create the image file. Click **Preview** to check what the image will look like before rendering the full-size file. The POV-Ray window will briefly appear as the rendering proceeds.
7. When you are satisfied with the output, click **Render** to create the file. You can find the .png file in the output directory shown in the bottom of the window.
8. Using different *Material Properties* creates different image styles. Try different materials to see the effects.

Generating Rotating Images

For presentations or websites, you may want to display a rotating image of your molecule, to highlight stereochemistry, or other spatial effects that are not obvious in static images. This section will show you how to generate images that can be used to make animated (rotating) molecule graphics.

9. To see how the rotation will look, click the **Animate...** button in the Mercury window. Tick the box next to **Rotate** and select the radio button for the "y" axis. This will start the molecule rotating in the Mercury viewer. Select different axes (one at a time) and note how the molecule rotates.
10. Once you have decided on the orientation of your molecule and the axis of rotation, return to the **POV-Ray Image** dialog box. Tick box next to "Generate Animation Frames." Now specify the axis of rotation (x, y, or z) and the "Number of Frames" to generate. The structure will always rotate through 360 degrees, so the higher number of frames, the higher the "resolution" of the rotating image. The default value of 30 is typically a good compromise between file size and resolution. Click **Render** to generate the 3D images.

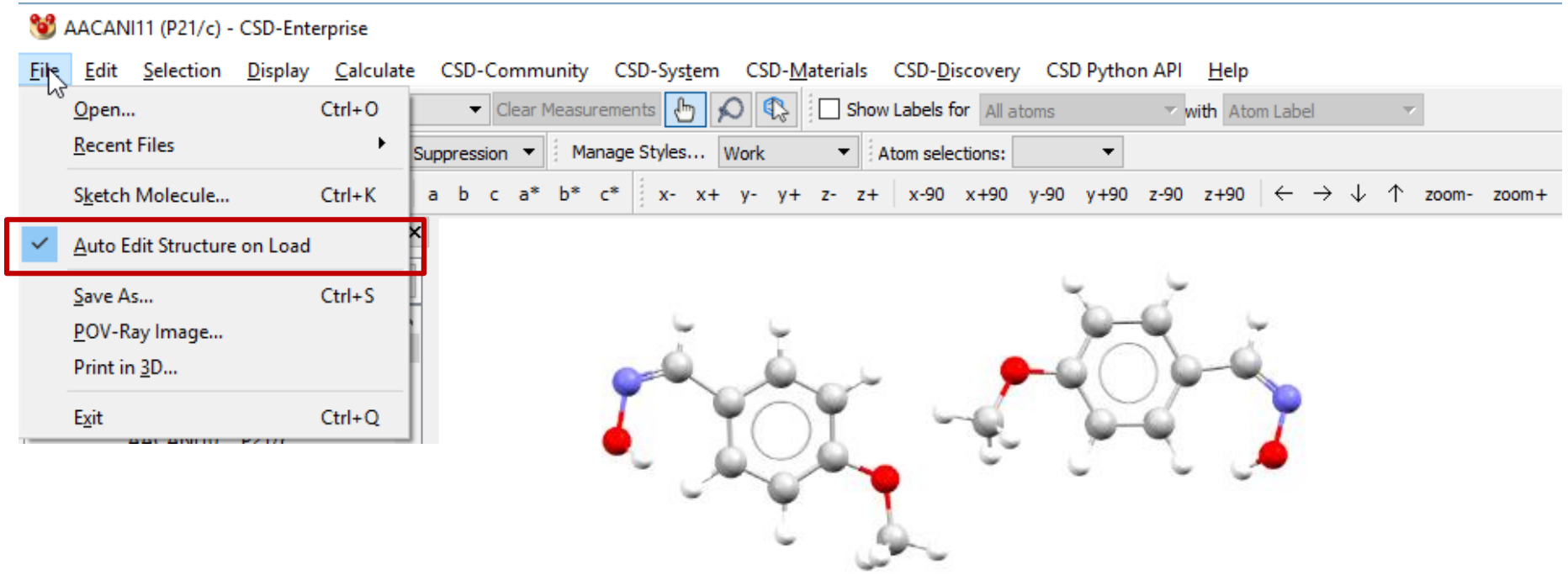


Exploring Mercury

Tips and Tricks

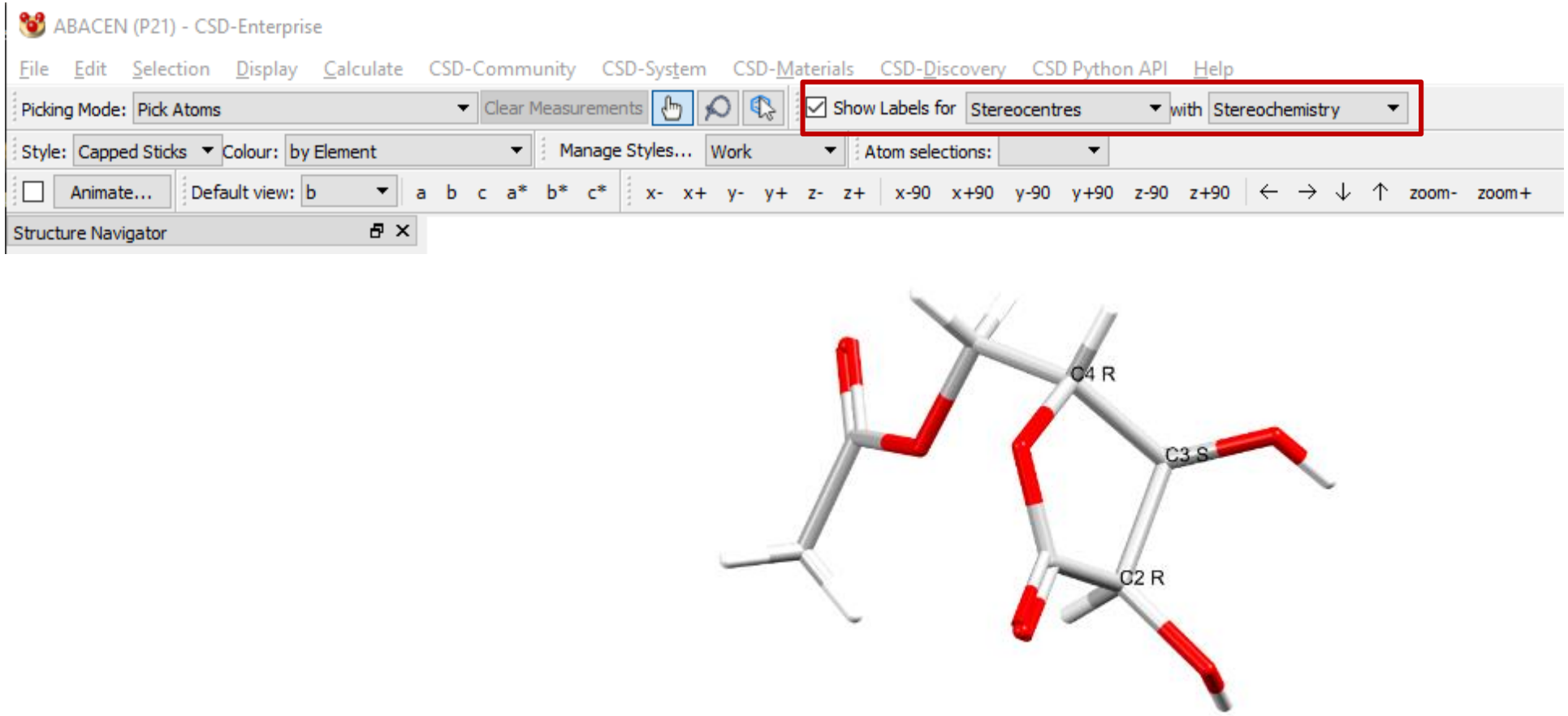
It’s the simple things

- Do you load .cif files into Mercury a lot?



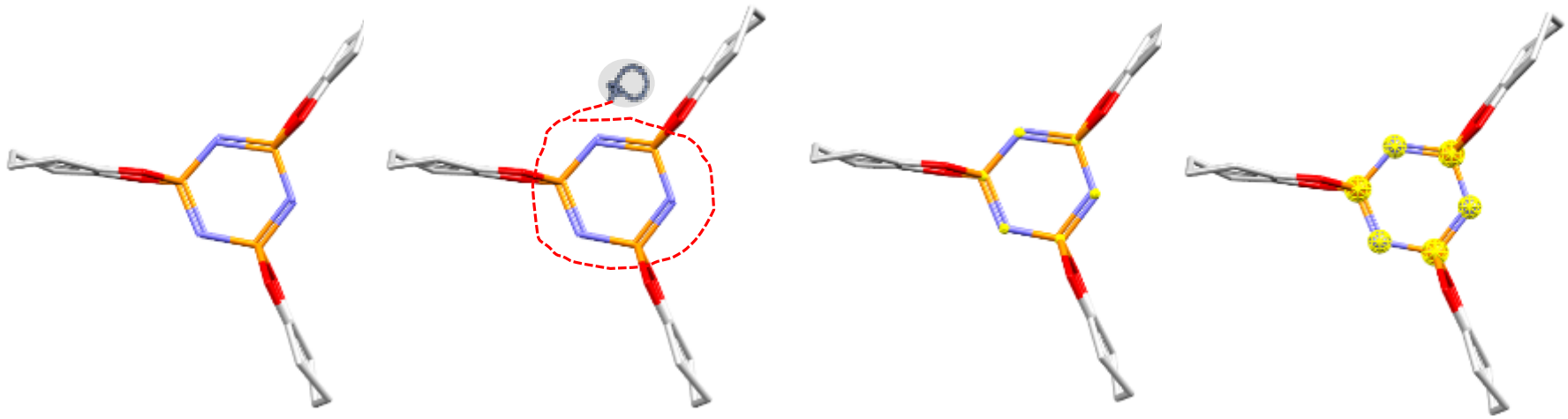
It’s the simple things

- Do you look at chiral molecules a lot?



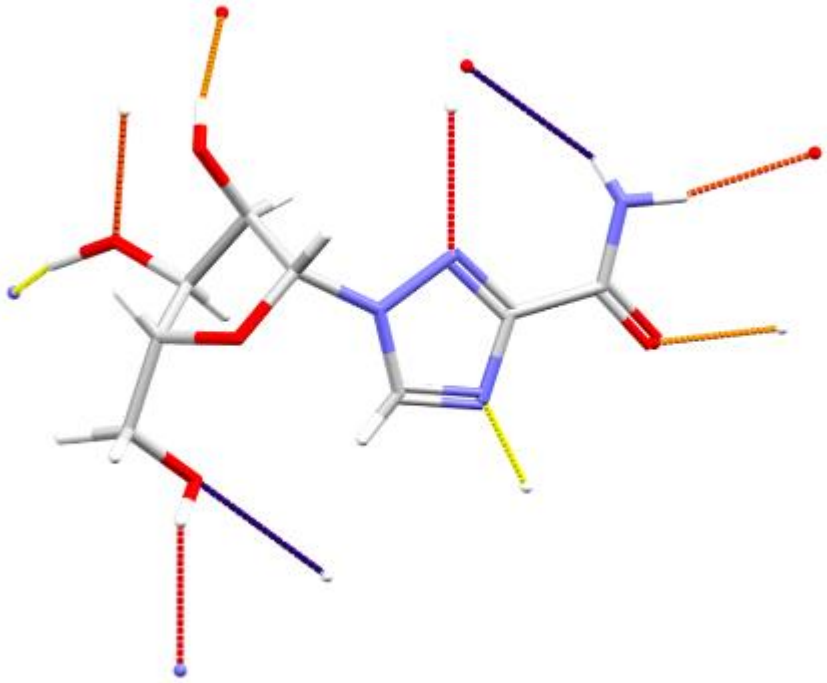
It’s the simple things

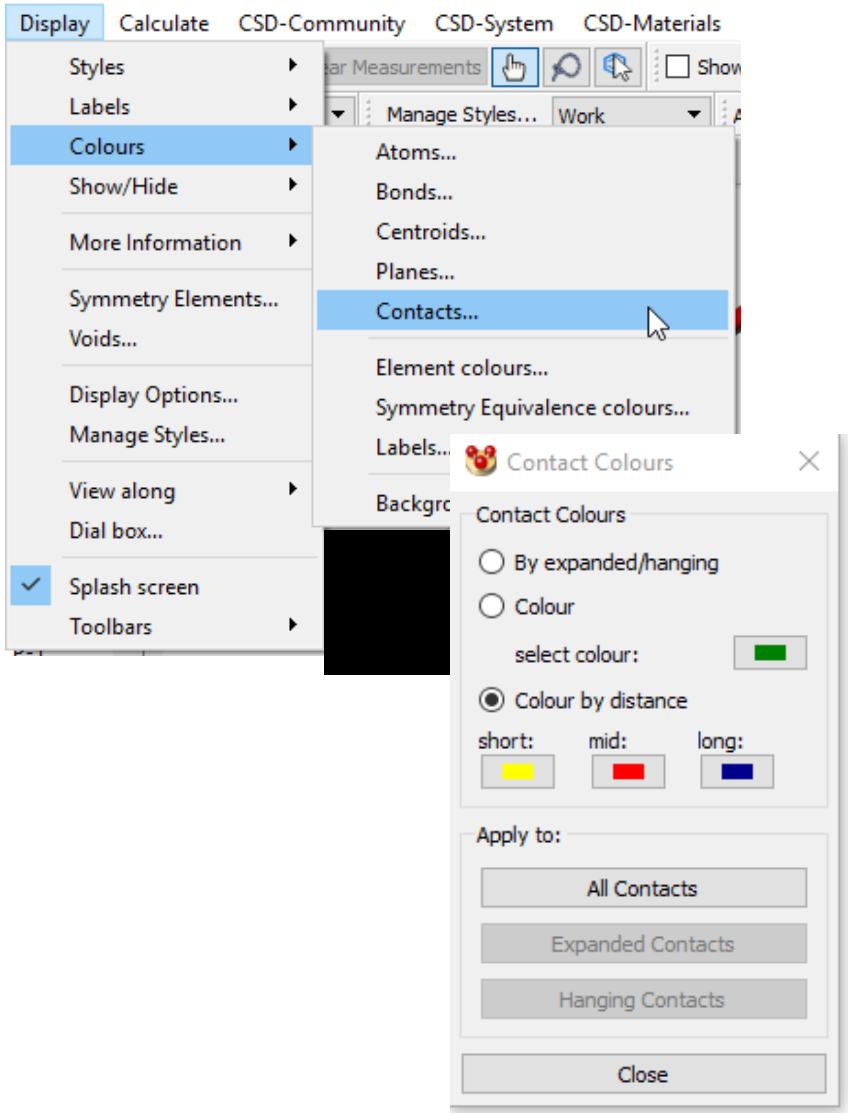
- Do you select parts of molecules often (to change display styles for example?)
- Alt-Left Mouse button invokes lasso selection



It’s the simple things

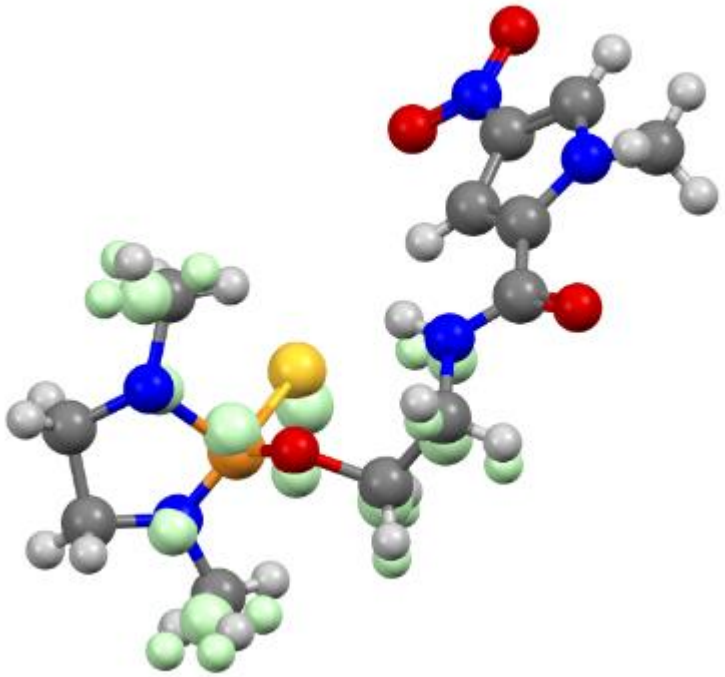
- Do you look at hydrogen bonds a lot?
- Colour by distance

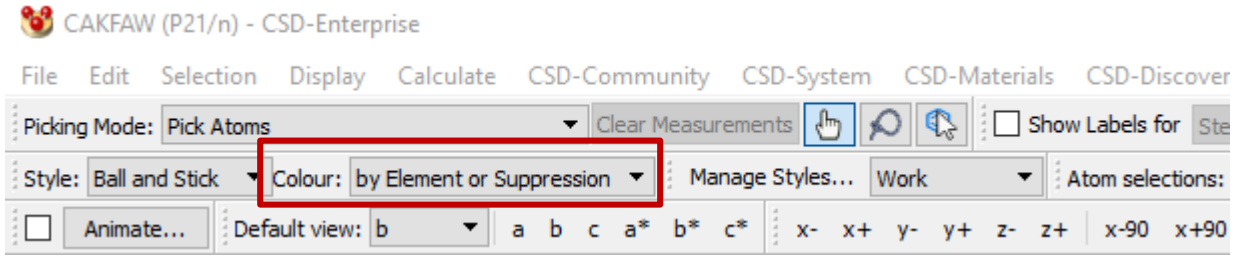




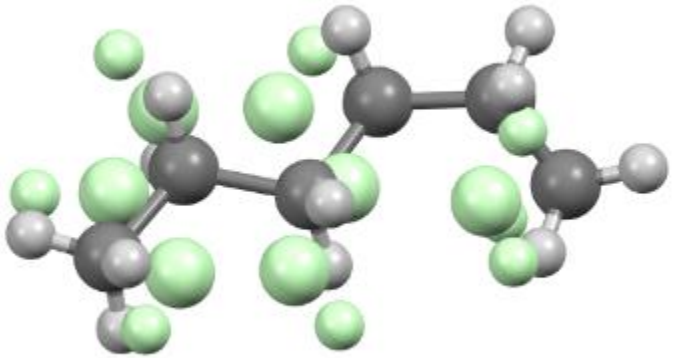
It’s the simple things

- Disordered structures?



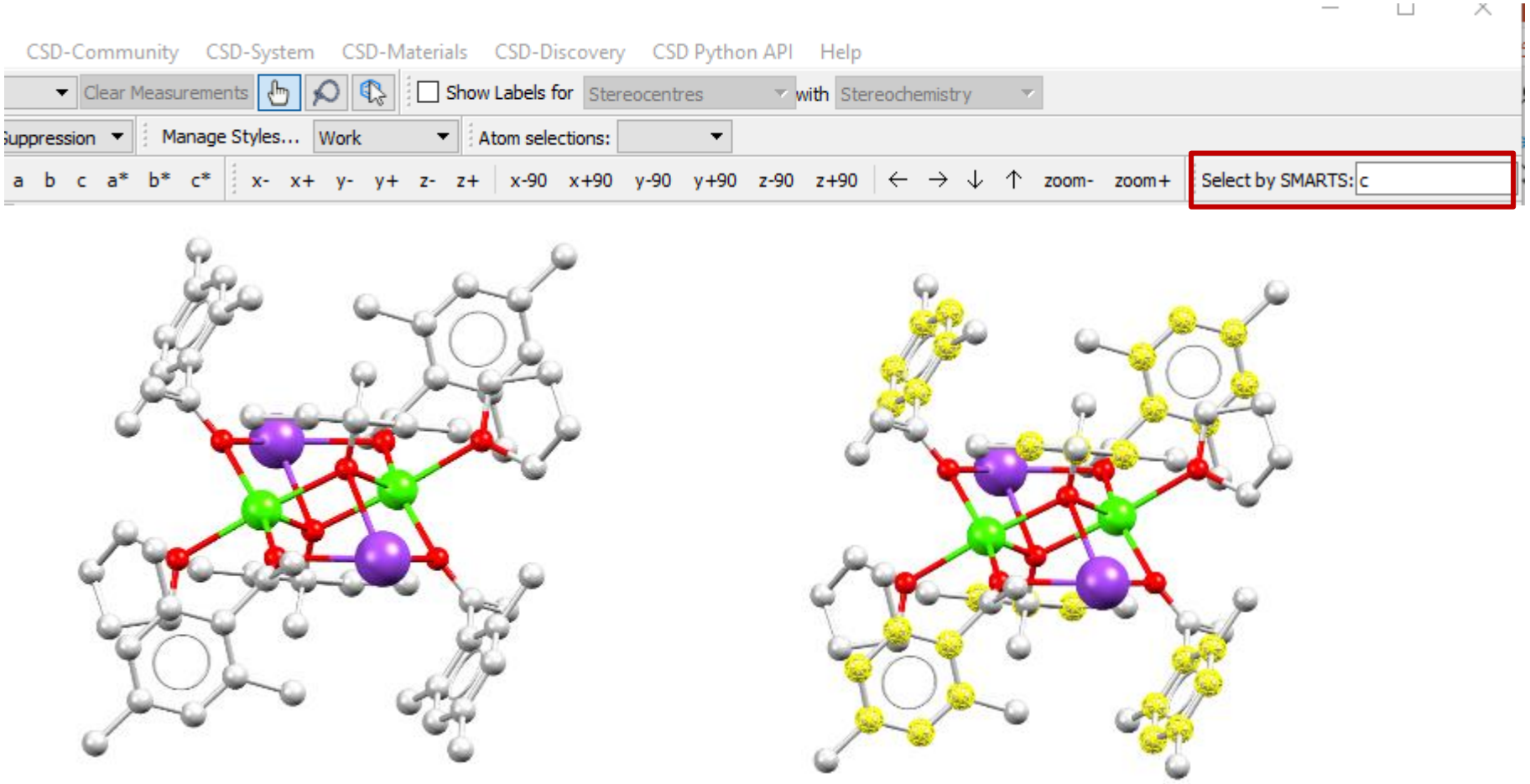


“Suppression” is our term for disordered atoms



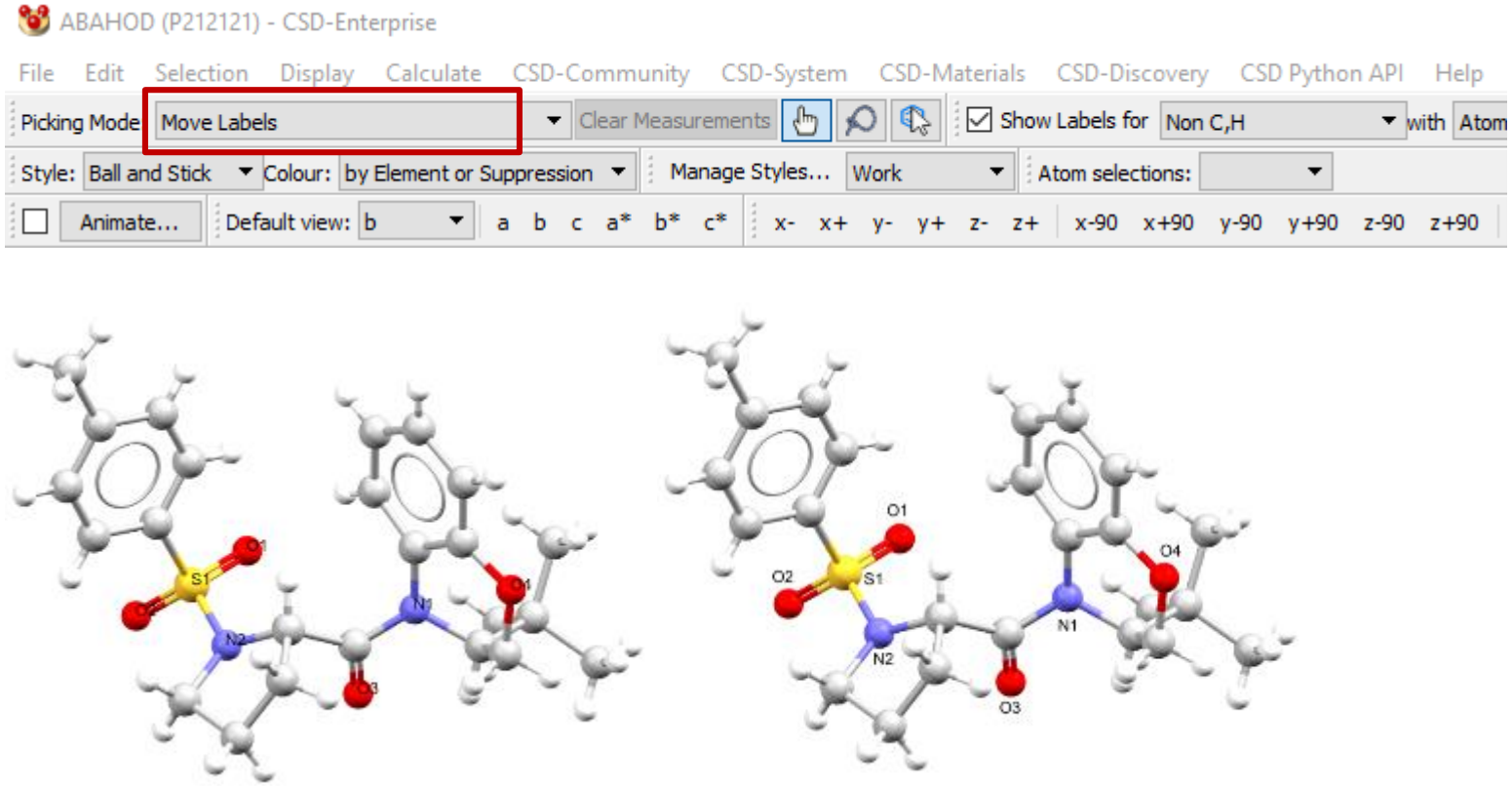
It’s the simple things

- Select by SMARTS



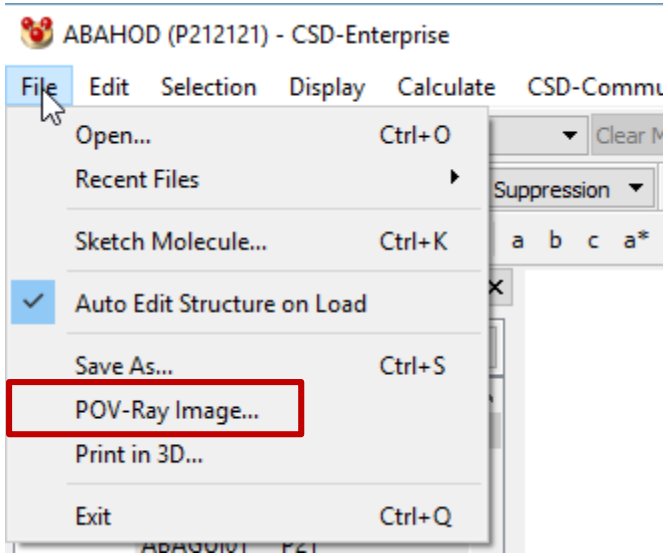
It’s the simple things

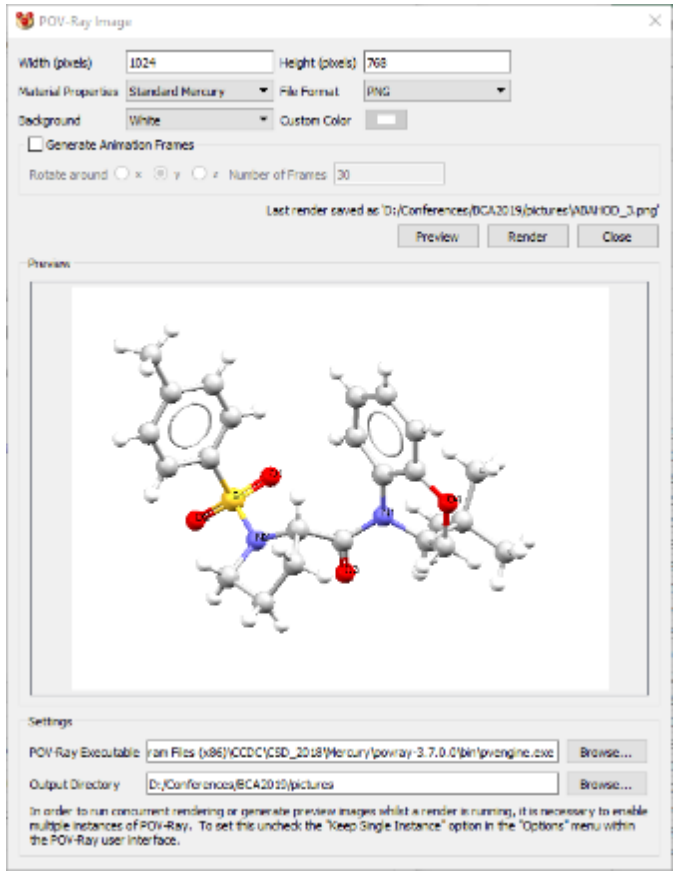
- Make pictures for publications or slides?



It’s the simple things

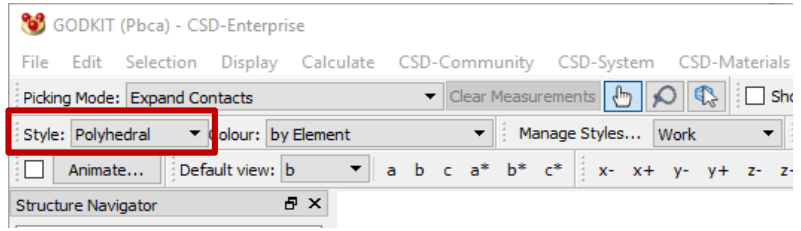
- Make pictures for publications or slides?

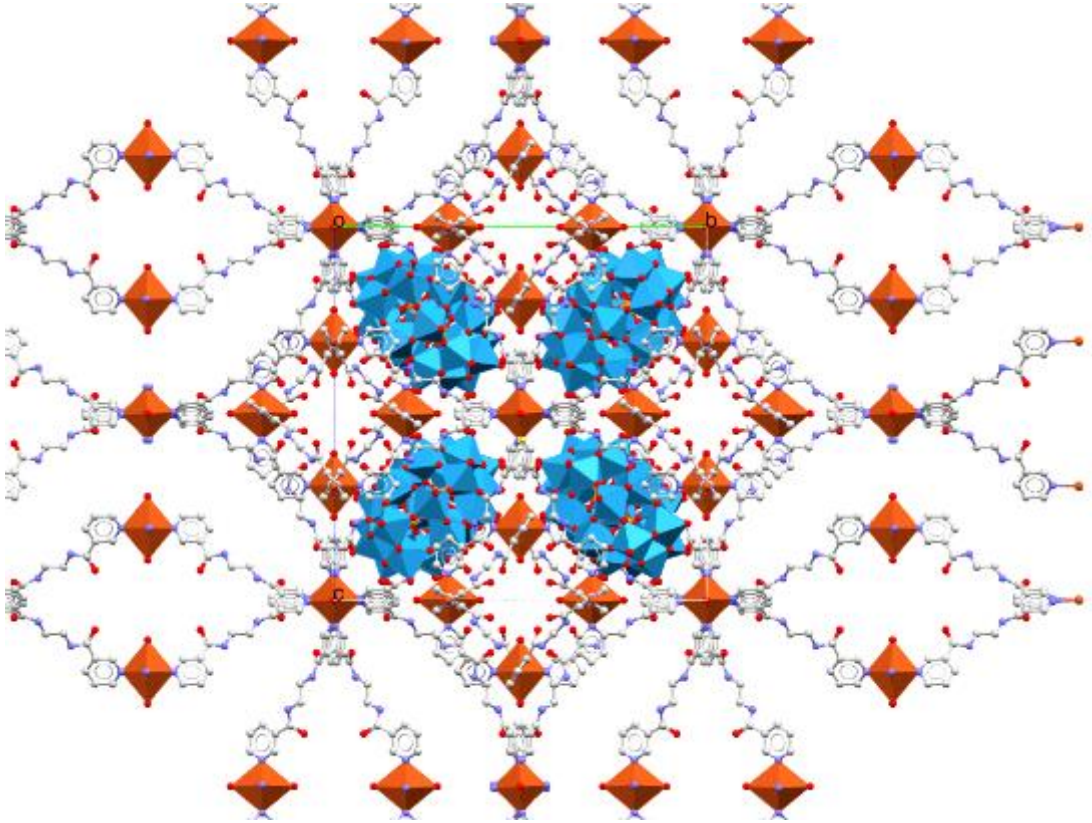




It’s the simple things

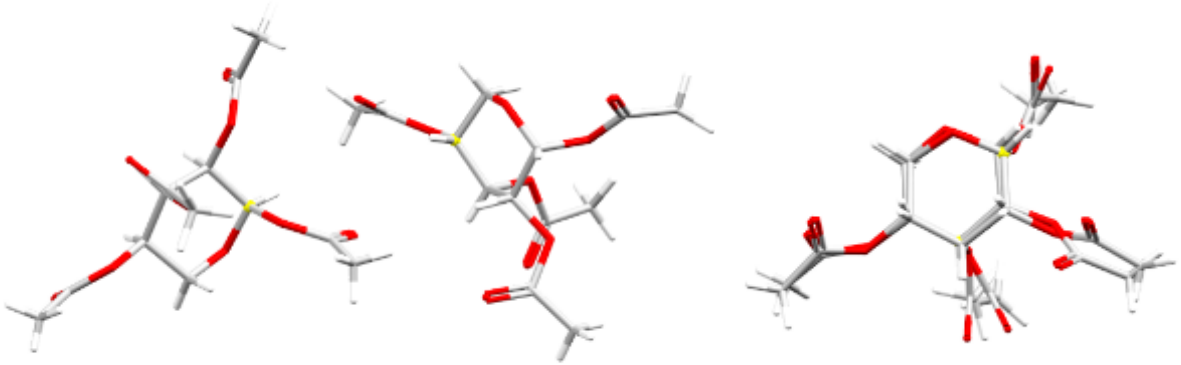
- Work with coordination metals?

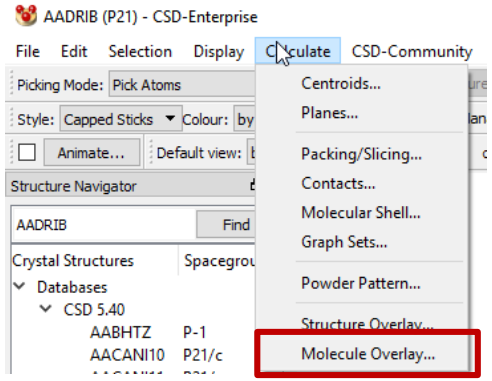


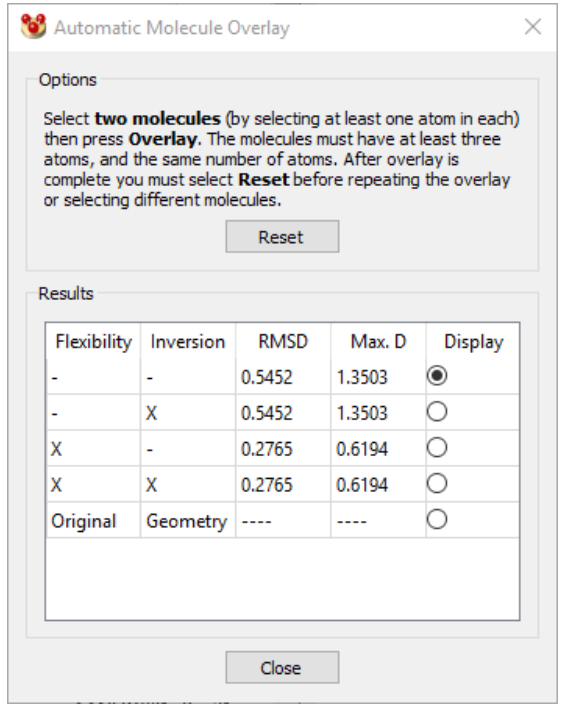


And a little more involved

- Molecule Overlay
- 2 molecules in the asymmetric unit and want to compare their conformations?

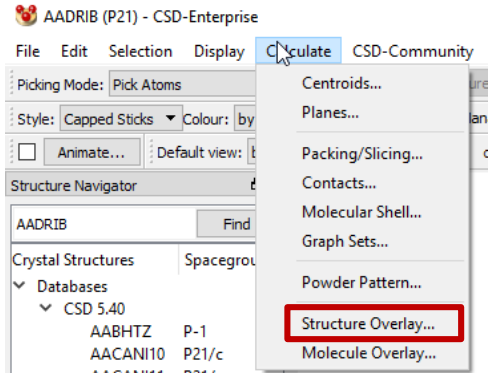


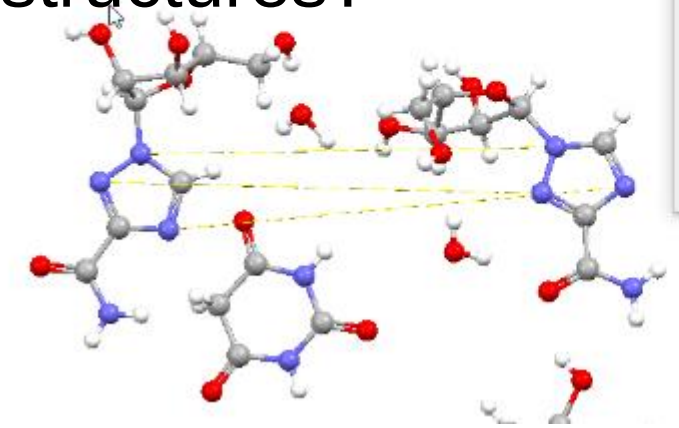


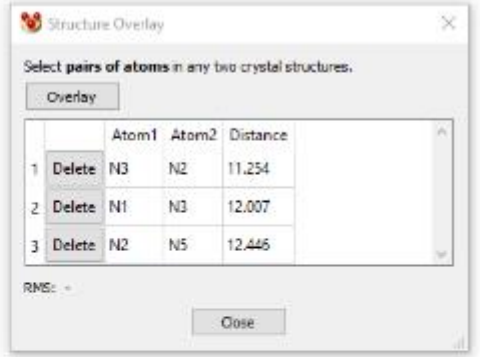


And a little more involved

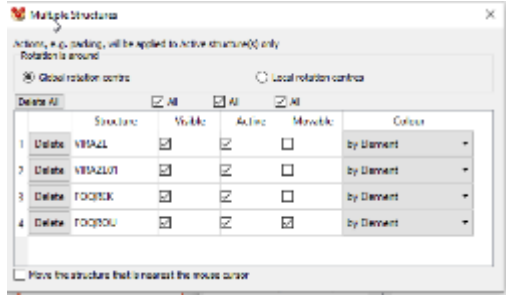
- Overlaying more than 2 structures?







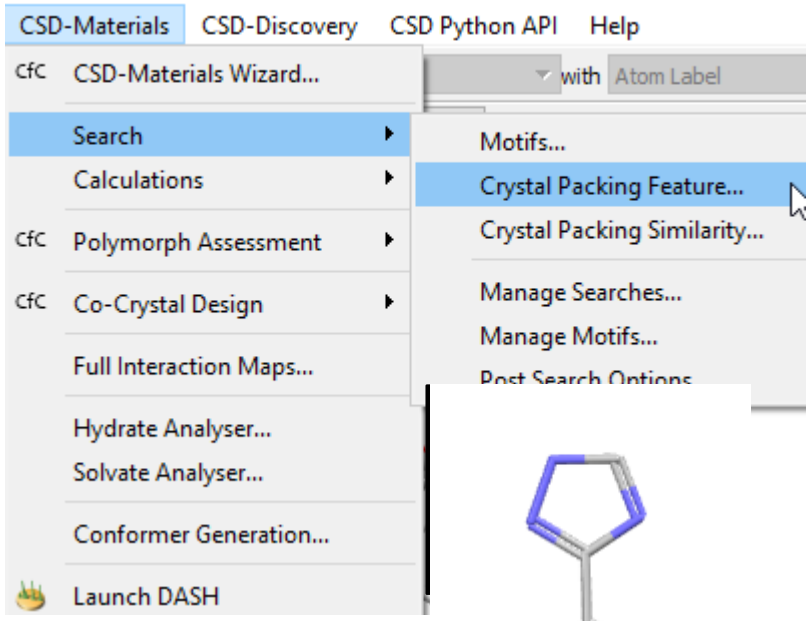
Last molecule selected is the one that moves

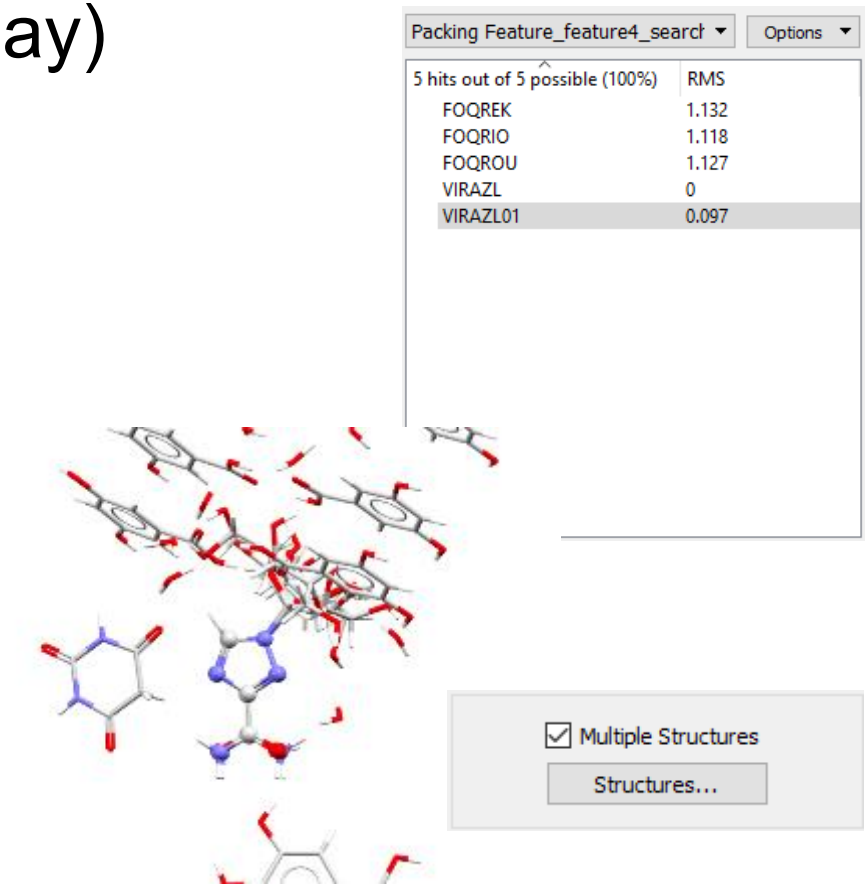


Multiple structures dialog allows you to “freeze” and hide the overlaid molecules

And a little more involved

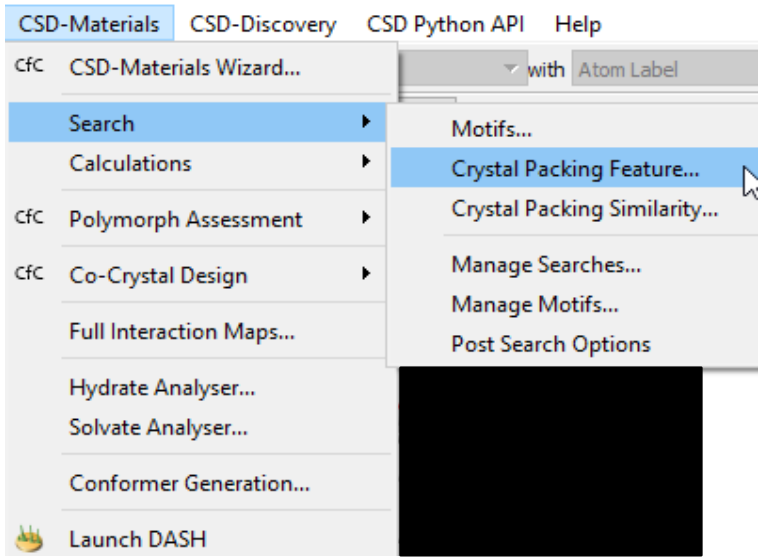
- Want to overlay molecules from different structures? (and you don’t want to use structure overlay)

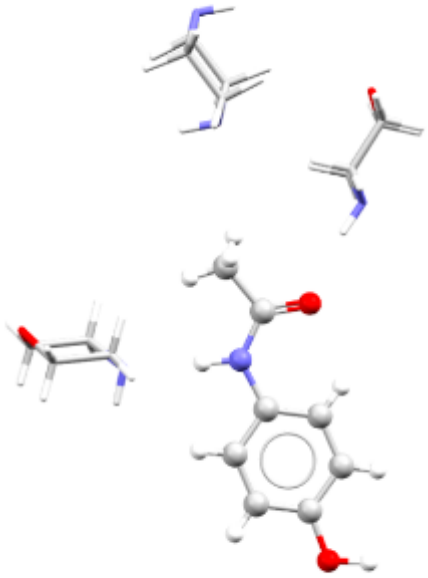


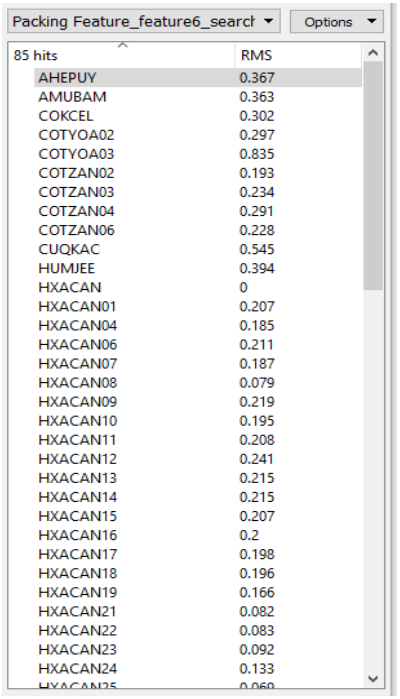


And a little more involved

- Do you want to find all crystal structures containing a particular molecule?
  - Solid form landscape of paracetamol, ethanol solvates, anything with a common substructure, not just packing features…







And a little more involved

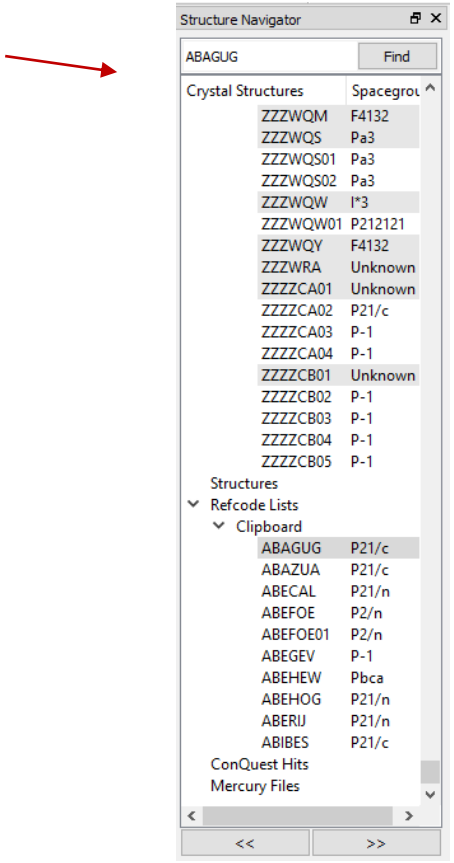
- Do you build subsets of structures based on a particular property? Cut and paste from Excel….

	A	B	C	G	H	I	J	K	L	M	N	O	P	R	S	T
7	Refcode	spaceg	zprime	mol_co	molecu	active.r	laxis	maxis	saxis	lm_rati	ls_ratio	ms_rati	shape_	Atoms	Aromat	Ring_A
54	ABAGUG	P21/c	0.5	1269.882	1269.882	1269.882	20.021	16.593	15.394	1.207	1.301	1.078	0.054	180	6	36
134	ABAZUA	P21/c	1	556.5447	556.5447	556.5447	19.086	12.307	11.725	1.551	1.628	1.05	0.055	80	6	36
148	ABECAL	P21/n	1	354.4418	354.4418	354.4418	20.035	9.386	3.592	2.135	5.577	2.613	0.13	46	6	35
166	ABEFOE	P2/n	0.5	618.7596	618.7596	618.7596	13.965	12.457	9.76	1.121	1.431	1.276	0.091	82	6	36
167	ABEFOE01	P2/n	0.5	618.7596	618.7596	618.7596	13.965	12.457	9.76	1.121	1.431	1.276	0.091	82	6	36
175	ABEGEV	P-1	1	620.7754	620.7754	620.7754	14.918	13.838	13.373	1.078	1.116	1.035	0.069	84	6	36
186	ABEHEW	Pbca	1	744.746	744.746	744.746	16.992	15.583	10.971	1.09	1.549	1.42	0.084	95	6	36
191	ABEHOG	P21/n	1	700.6936	700.6936	700.6936	15.065	14.07	10.708	1.071	1.407	1.314	0.087	88	6	36
283	ABERIJ	P21/n	1	694.7266	694.7266	694.7266	15.2	15.144	10.921	1.004	1.392	1.387	0.091	86	6	36
308	ABIBES	P21/c	1	492.756	492.756	492.756	14.536	12.656	9.635	1.148	1.509	1.314	0.09	64	6	36

Property in this column is the number of aromatic rings in molecule. Chosen structures with 6 aromatic rings.

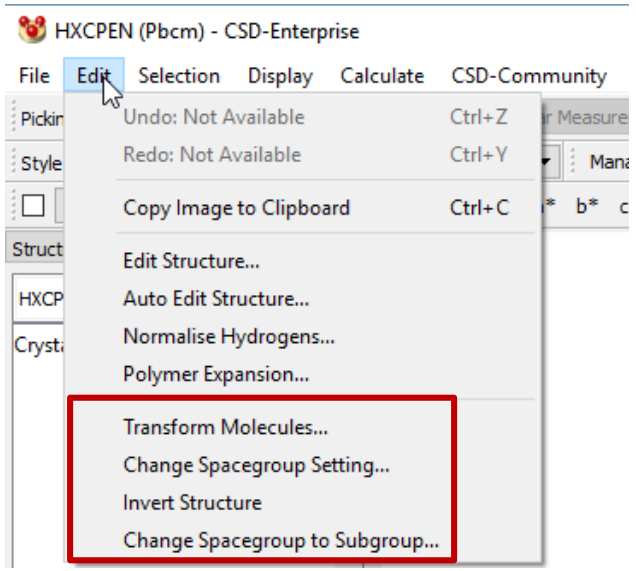
And a little more involved

- Do you build subsets of structures based on a particular property? Cut and paste from Excel….into Mercury



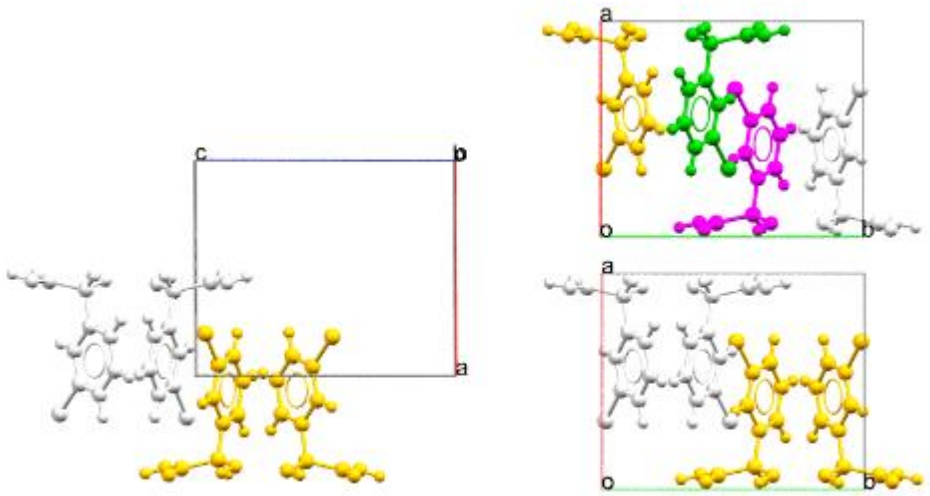
And a little more involved

- Editing a structure
  - Invert a structure/ Invert a molecule
  - Removing symmetry by choosing a subgroup
  - Changing the space group setting: P21/n to P21/c



And a little more involved

- Do you explore structural changes due to phase transitions?
- FIZPEL/ FIZPEL01



$P2_1/c$
Room Temp

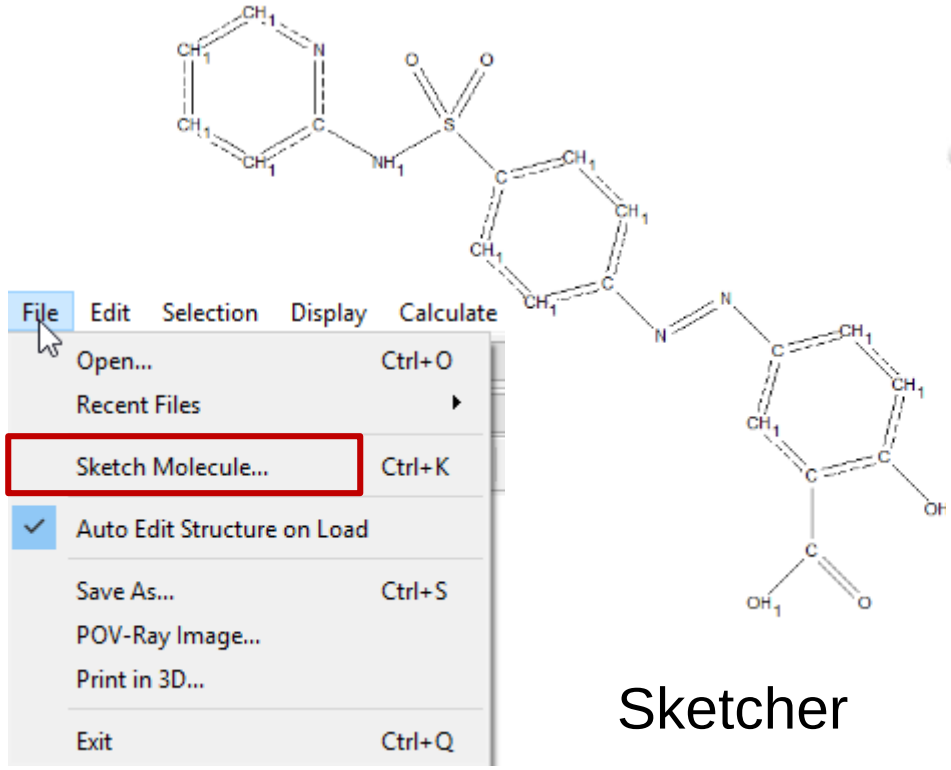
Mirror symmetry relationship formalised in P21/c

Transformed to subgroup P-1

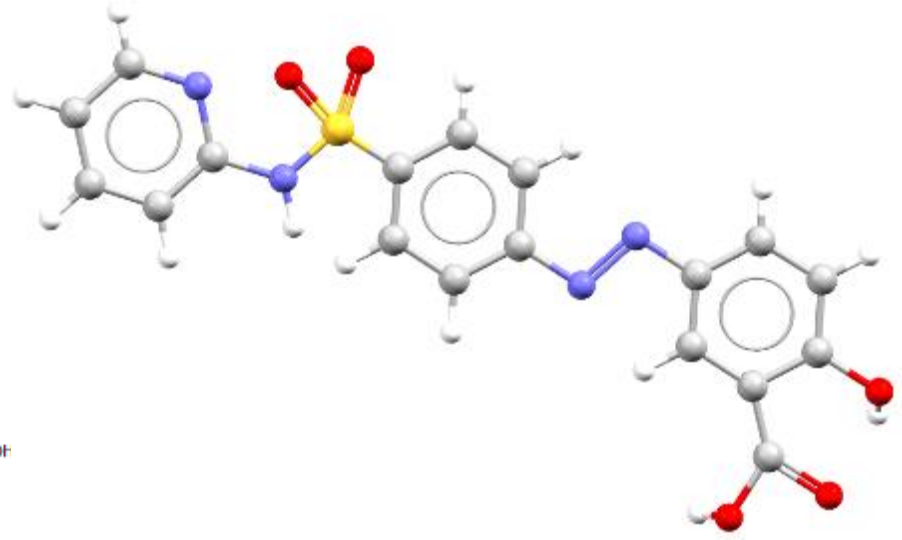
P-1
Low temp

And a little more involved

- Sketch a molecule and generate a conformation



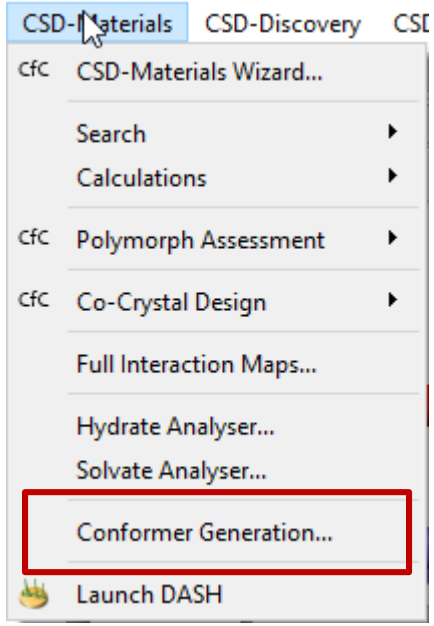
Sketcher

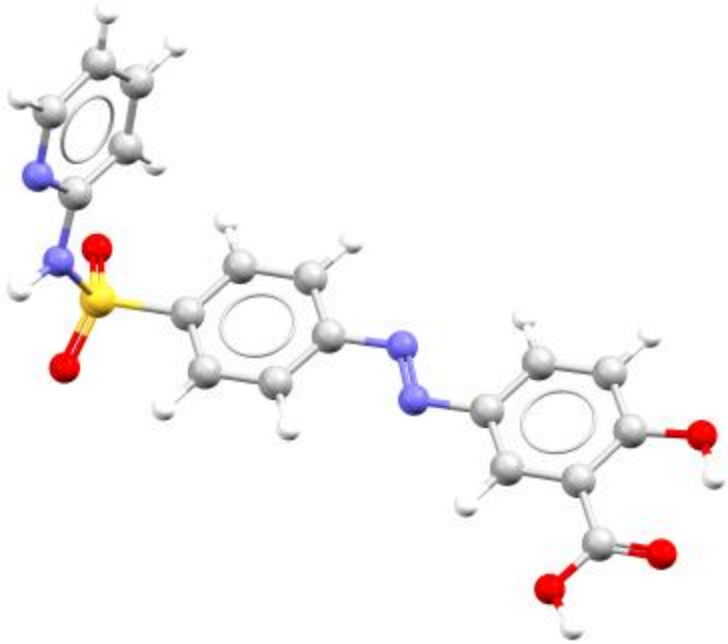


Mercury

And a little more involved

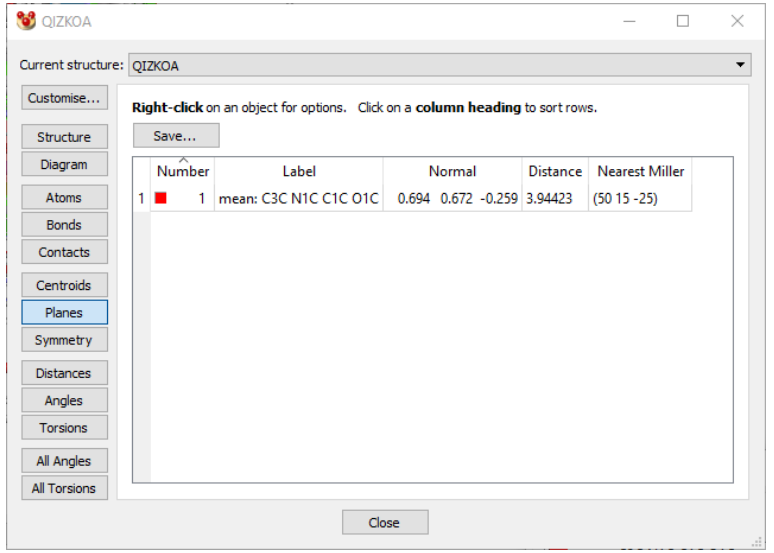
- Sketch a molecule and generate a conformation….





Too many to mention:

- Custom view
- Powder pattern simulation
- Links to Isostar, GOLD, Mogul etc
- Motif searching and hydrogen bond propensity calculations
- More info dialog
  - Nearest Miller Plane
  - Selecting atoms from Atoms list
- MOPAC – point group symmetry



Mercury

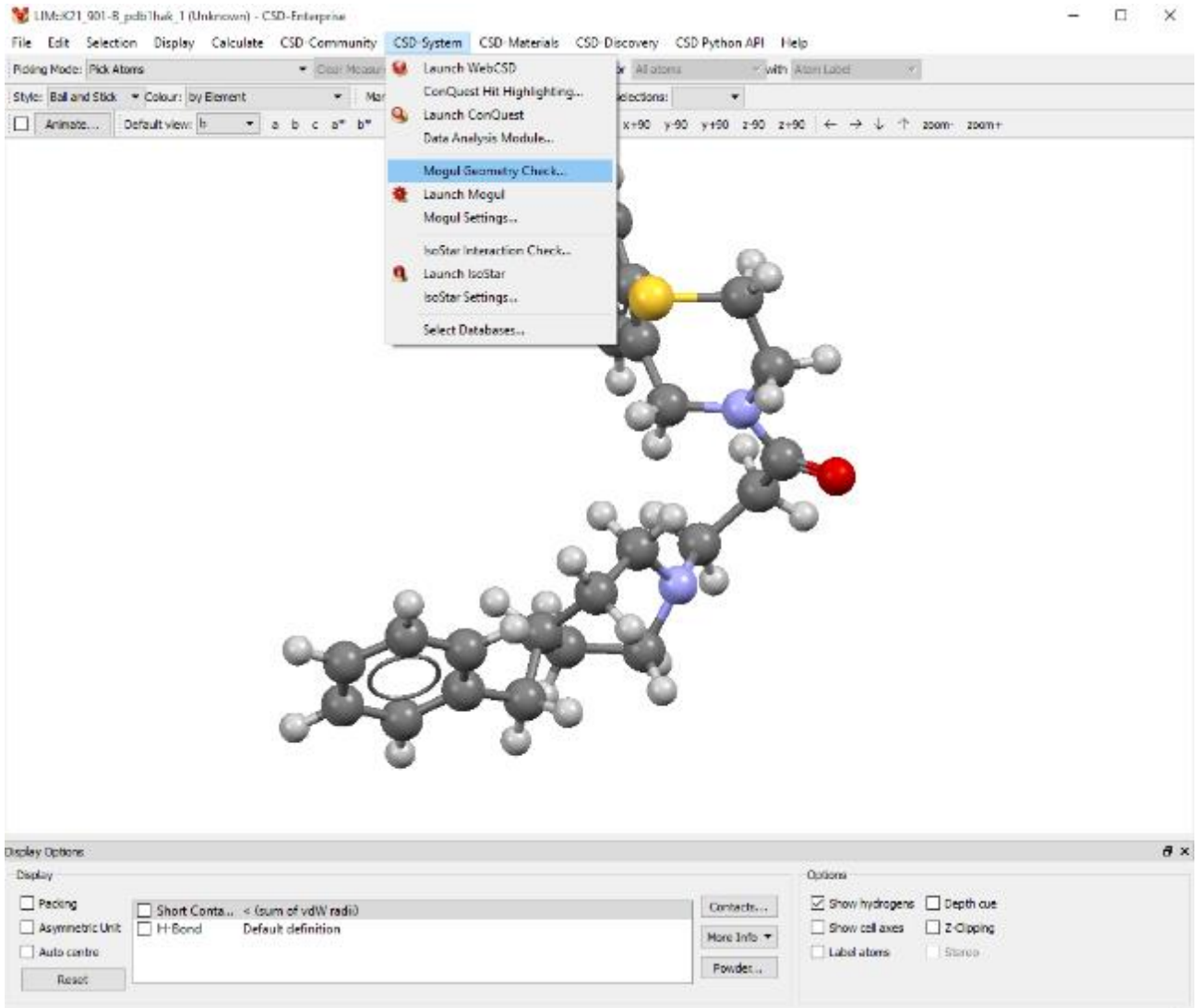
Mogul, Full Interaction Maps, Crystal Packing

Mogul

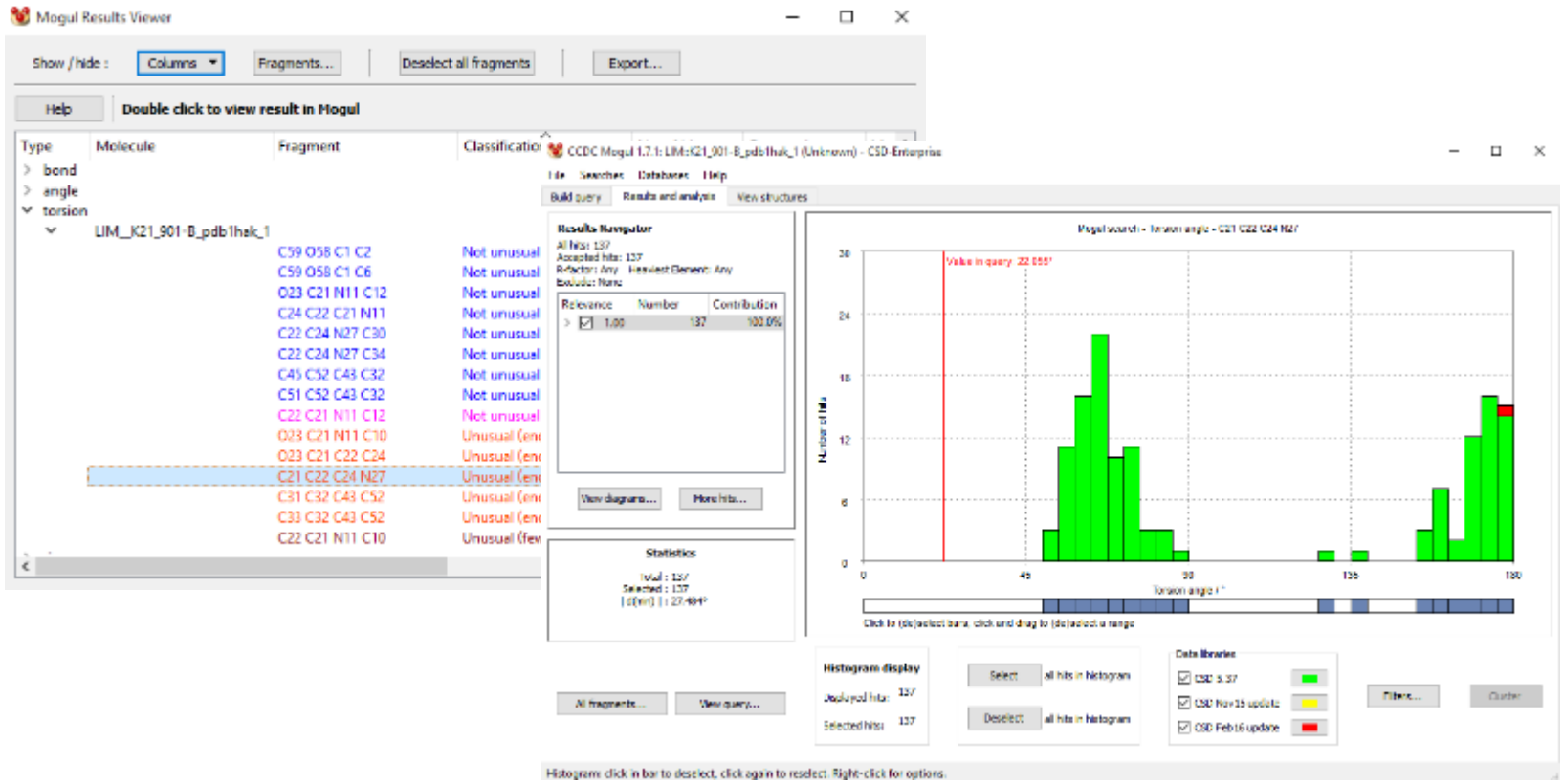
- Incorporates pre-computed libraries of bond lengths, valence angles, torsion angles and ring conformations derived entirely from the CSD
- Validate complete geometry: retrieve distributions, and figures of merit for all bonds, angles, torsions and rings in the molecule
- Fragment Generalisation: If fragment specified is rare, Mogul will include CSD results from the most similar fragments
- Hyperlinking to the CSD: View CSD entries in specific areas of histogram
- Integration with other software: Instruction file permits other programs to interact with Mogul

Bruno et al., J. Chem. Inf. Comput. Sci., 44, 2133-2144, 2004

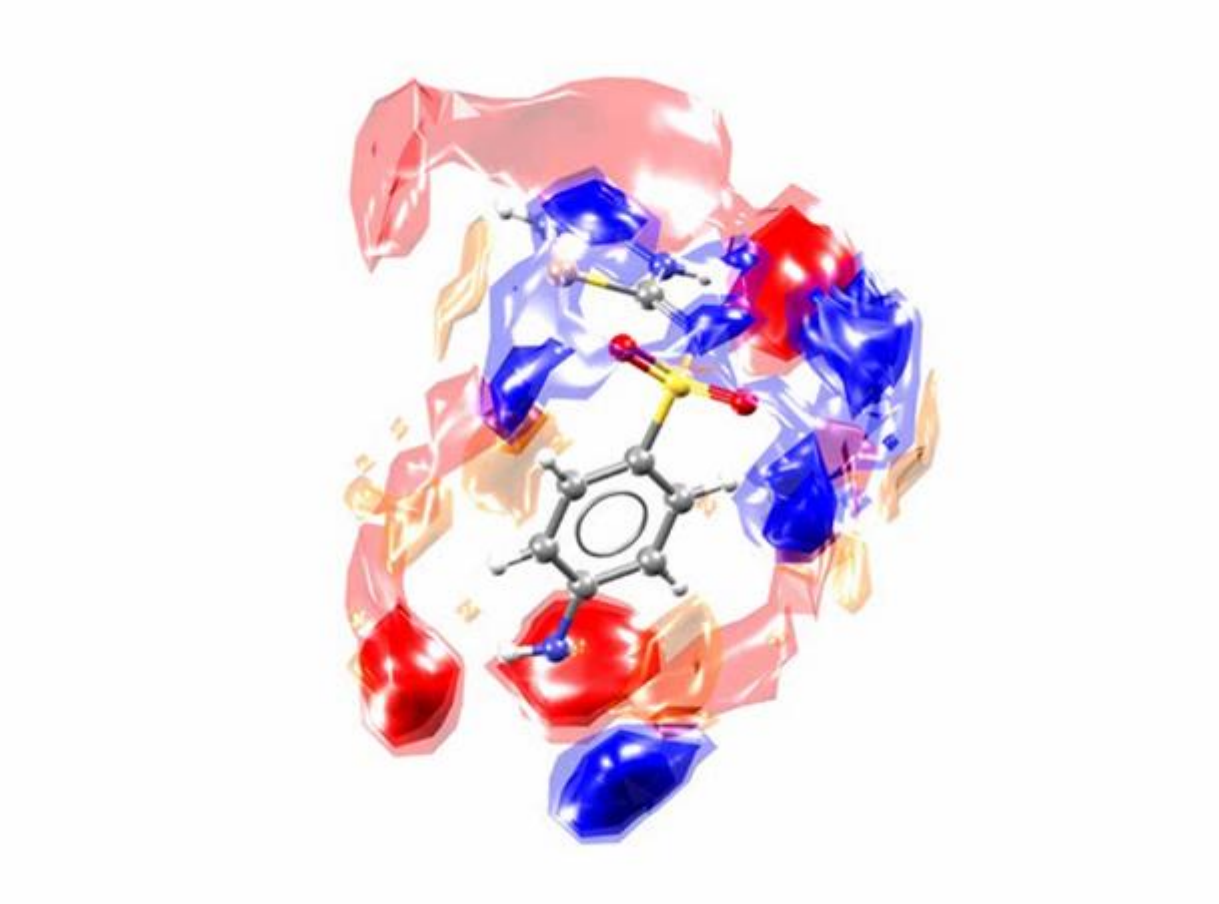
Mogul


File Edit Selection Display Calculate CSD Community CSD System CSD Materials CSD Discovery CSD Python API Help
Launch WebCSD
ConQuest Hit Highlighting...
Launch ConQuest
Data Analysis Module...
Mogul Geometry Check...
Launch Mogul
Mogul Settings...
IsoStar Interaction Check...
Launch IsoStar
IsoStar Settings...
Select Databases...
Display Options
Display
Packing
Asymmetric Unit
Auto centre
Reset
Short Conta... < (sum of vdW radii)
H-Bond Default definition
Contacts...
More Info
Powder...
Options
Show hydrogens
Depth cue
Show cell axes
Z-Clipping
Label atoms

Mogul


Mogul Results Viewer
Show / hide :
Columns
Fragments...
Deselect all fragments
Export...
Help
Double click to view result in Mogul
Type
Molecule
Fragment
bond
angle
torsion
LIM_K21_901-B_pdb1hak_1
C59 O58 C1 C2
C59 O58 C1 C6
O23 C21 N11 C12
C24 C22 C21 N11
C22 C24 N27 C30
C22 C24 N27 C34
C45 C52 C43 C32
C51 C52 C43 C32
C22 C21 N11 C12
O23 C21 N11 C10
O23 C21 C22 C24
C21 C22 C24 N27
C31 C32 C43 C52
C33 C32 C43 C52
C22 C21 N11 C10
Not unusual
Not unusual
Not unusual
Not unusual
Not unusual
Not unusual
Not unusual
Not unusual
Not unusual
Unusual
Unusual
Unusual
Unusual
Unusual
Unusual
CCDC Mogul 1.7.1: LIM:K21_901-B_pdb1hak_1 (Unknown) - CSD Enterprise
File
Searches
Databases
Help
Build query
Results and analysis
View structures
Results Navigator
All hits: 137
Accepted hits: 137
R-factor: Any
Heaviest Element: Any
Exclude: None
Relevance
Number
Contribution
1.00
137
100.0%
View diagrams...
More hits...
Statistics
Total : 137
Selected : 137
Mogul search - torsion angle - C21 C22 C24 N27
Value in query: 22.655°
Number of hits
torsion angle /°
Click to (de)select bars, click and drag to (de)select a range
Histogram display
Accepted hits: 137
Selected hits: 137
Select all hits in histogram
Deselect all hits in histogram
Data libraries
CSD 5.37
CSD Nov16 update
CSD Feb16 update
Filters...
Cluster
All fragments...
View query...
Histogram: click in bar to deselect, click again to reselect. Right-click for options.

Full Interaction Mapping

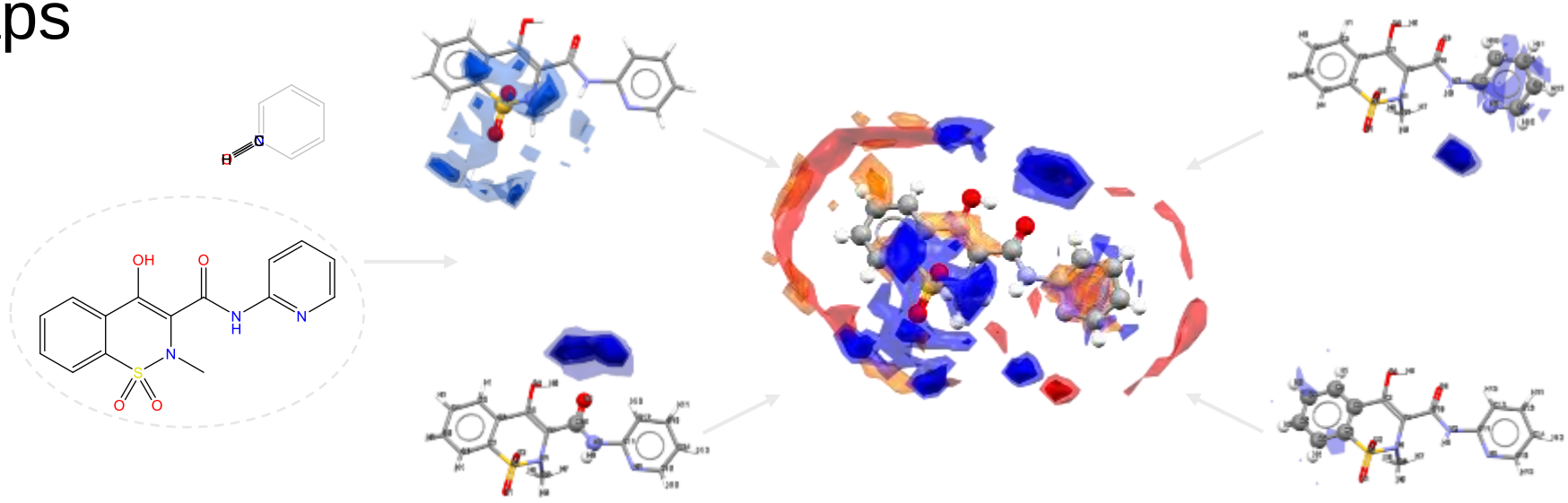


- Provides your molecule’s interaction preferences, in the context of the crystal structure, at the click of a button
- Based on validated experimental structural knowledge
- Interactions visualised simply in 3D within Mercury

Wood *et al.*; *CrystEngComm,* **2013**, *15*, 65-72. DOI: 10.1039/C2CE25849H

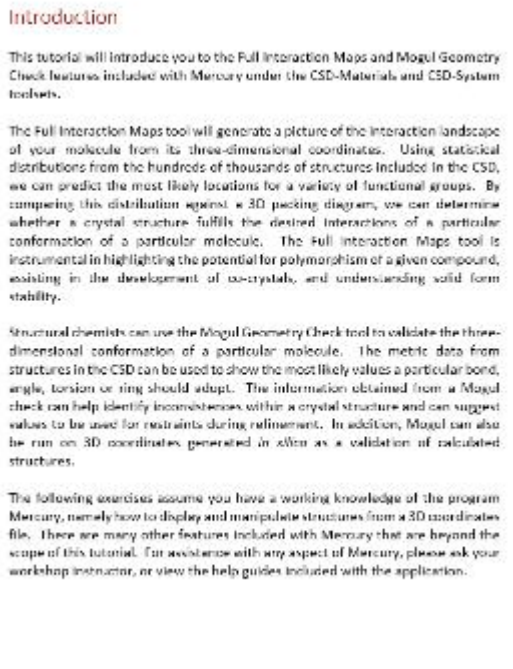
Combining plots into Full Interaction Maps

- Molecule is broken down into fragments
- IsoStar maps for each fragment in the molecule & for multiple probe types are combined together to produce Full Interaction Maps



Further tutorials

- If you have time and are interested, we have some additional tutorials on assessing *intra*molecular conformations and analysing *inter*molecular interactions
  - Mogul
  - Full interaction maps
  - Packing features



Introduction

This tutorial will introduce you to the Full Interaction Maps and Mogul Geometry Check features included with Mercury under the CSD-Materials and CSD-System toolsets.

The Full Interaction Maps tool will generate a picture of the interaction landscape of your molecule from its three-dimensional coordinates. Using statistical distributions from the hundreds of thousands of structures included in the CSD, we can predict the most likely locations for a variety of functional groups. By comparing this distribution against a 3D packing diagram, we can determine whether a crystal structure fulfils the desired interactions of a particular conformation of a particular molecule. The Full Interaction Maps tool is instrumental in highlighting the potential for polymorphism of a given compound, assisting in the development of co-crystals, and understanding solid form stability.

Structural chemists can use the Mogul Geometry Check tool to validate the three-dimensional conformation of a particular molecule. The metric data from structures in the CSD can be used to show the most likely values a particular bond, angle, torsion or ring should adopt. The information obtained from a Mogul check can help identify inconsistencies within a crystal structure and can suggest values to be used for restraints during refinement. In addition, Mogul can also be run on 3D coordinates generated in xWin as a validation of calculated structures.

The following exercises assume you have a working knowledge of the program Mercury, namely how to display and manipulate structures from a 3D coordinates file. There are many other features included with Mercury that are beyond the scope of this tutorial. For assistance with any aspect of Mercury, please ask your workshop instructor, or view the help guides included with the application.