

CSD workshop - exercises

Mercury exercises

1. Overlay of polymorphic crystal structures
2. Making a cif-file with a disordered structure ready for input into a modelling program.
3. Assessing the stability of paracetamol

Conquest exercises

1. Combining searches
2. Searching with geometrical constraints
3. Influence of substitutions in heterocyclic γ -aminoacids on torsion angles
4. Exploring the landscape of σ -hole interactions



MERCURY exercises

1. Overlay of polymorphic crystal structures

Objectives

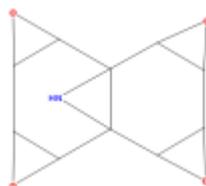
- To use the overlay functionality in Mercury to investigate the differences in the extended crystal structures of a pair of polymorphs.

Steps Required

- Read in a structure.
- Display additional information about the structure.
- Perform an initial inspection of the hydrogen bonds.
- Add missing hydrogens
- Display a second crystal structure simultaneously.
- Overlay the two structures and compare their H-bonding motifs.

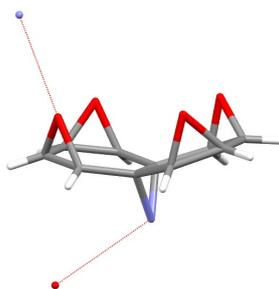
Example

- The understanding of polymorphism (i.e. why a given molecule adopts a particular crystal packing arrangement) is of extreme importance in the pharmaceutical industry, particularly for bioavailability and process problems, and could develop knowledge in various fields such as nucleation and crystal growth.
- This example investigates differences in the hydrogen bonding of the orthorhombic and monoclinic polymorphs of naphthalene tetroxide imine:



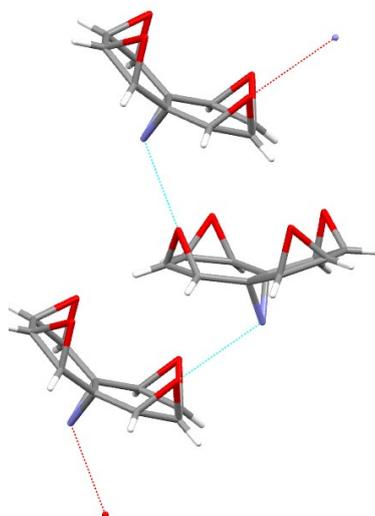
1. Load the required crystal structure (BAZYAC).
 - Check that you have loaded the orthorhombic polymorph by displaying the additional information about the structure: click on the **More Info** button (bottom centre of the main Mercury window) and then select **Structure Information...** from the resulting menu.
2. Investigate the hydrogen bonding.
 - Switch on the tick box next to the words **H-Bond Default definition** in the list box underneath the main display area:
 - No hydrogen bonds are found. However, the structure does contain an NH donor (the hydrogen atom is missing). We can try the effect of increasing the maximum allowed length of hydrogen bonds by changing the default H-bond definition.
 - Double-click on the words **H-Bond Default definition** in the list box underneath the main display area, this will open the *Define H-bonds* dialogue.

Select the **Actual Distance** radio button, change the maximum value to 3.2Å then hit **Apply**:



Two hanging hydrogen bonds are now displayed. The NH hydrogen atom coordinates are not available for this structure. However, we can infer that one hydrogen bond involves the NH donating its hydrogen. Also, the epoxy oxygen atom can be seen to accept a hydrogen bond.

- Click on the hanging H-bond from the nitrogen (not the oxygen) to add another molecule to the display. Do the same from the hanging contact from the nitrogen of the new molecule.
- We should now have a hydrogen-bonded chain consisting of three molecules:



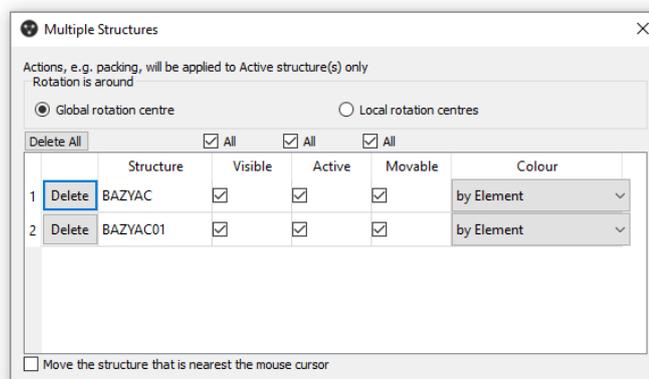
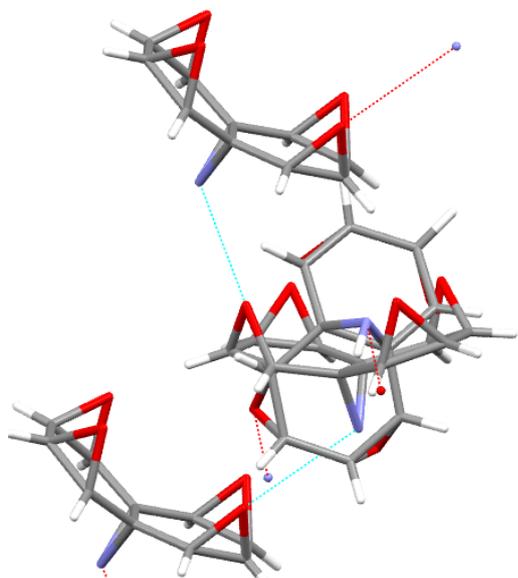
- Add the missing hydrogen on the nitrogen atom using Edit -> Auto edit structure

3. Load the second polymorph of naphthalene tetroxide imine.

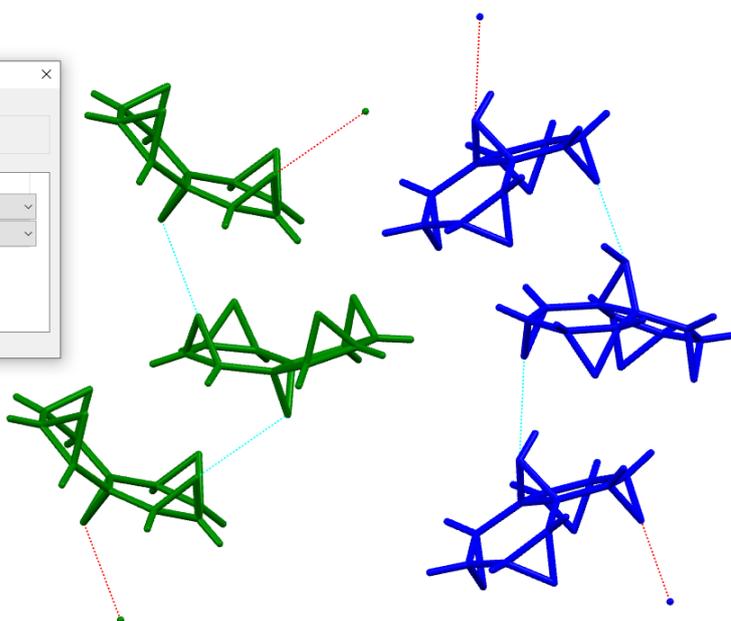
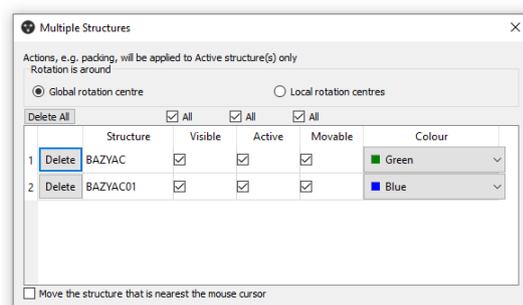
- To enable the display of multiple crystal structures, switch on the **Multiple Structures** tick box located under the list of structures on the right-hand side of the main Mercury window.
- Load the structure by typing *BAZYAC01* in the entry box above the scrolling list at the right of the Mercury window. In structure information spreadsheet window notice that *BAZYAC01* is the monoclinic polymorph.

4. Compare the two crystal structures.

- Hit the **Structures...** button located under the list of structures on the right-hand side of the main Mercury window. This will open the *Multiple Structures* dialogue. The two structures currently being displayed are listed:



- Play with the checkboxes and colours until you get something like this:

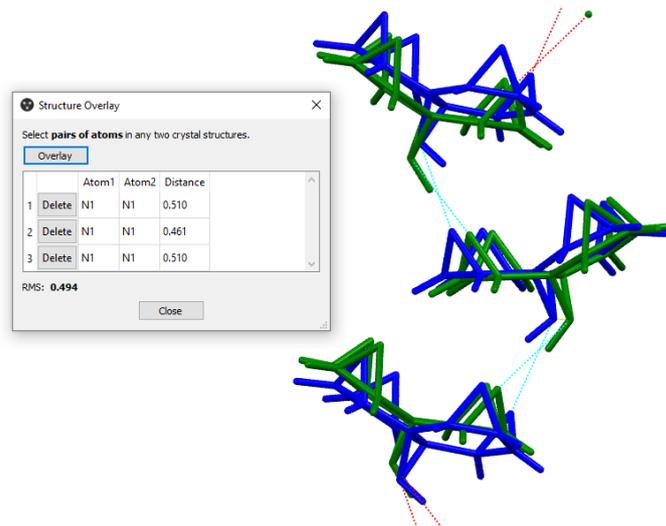


5. Overlay the structures.

- Select **Calculate** from the top-level menu followed by **Structure Overlay...** from the resulting pull-down menu. This will open the *Manual Crystal Overlay* dialogue.
- Pairs of atoms in *BAZYAC* and *BAZYAC01* must be selected in order to overlay the two crystal structures. These pairs of atoms will be used to perform the least-squares overlay.
- Ensure the picking mode is set to *Pick Atoms*; by default this can be selected using the top left toolbar in the Mercury display, if not the Picking Toolbar can be selected from the **Display** top-level menu followed by **Toolbars**. Select the nitrogen

atom in the first molecule in the chain of *BAZYAC* (coloured yellow) followed by the corresponding nitrogen atom in *BAZYAC01*. A broken yellow line will be displayed between the selected pair. Repeat this procedure for the two remaining pairs of nitrogen atoms in the display.

- In the end you should find something as shown here below



- Inspect the structures more closely. Notice that both structures show similar but not identical H-bonded chains of molecules along the crystallographic *b* axis. In particular, one of the hydrogen bonds forms to a different oxygen atom in *BAZYAC* than in *BAZYAC01*.
- This is common in polymorphism. Structures often show a common motif, but the chains or ribbons or molecules connected by the motif are packed differently in the third dimension.
- The example used is from Gervais and Coquerel, *Acta Cryst.* (2002). **B58**, 662-672.

2. Making a cif-file with a disordered structure ready for input into a modelling program.

Objectives

- To use the subgroup and edit functionality in Mercury to isolate one part of a disordered structure

Steps Required

- Read in a cif with a disordered structure with disorder related by symmetry.
- Verify the positions of the symmetry elements
- Lower the symmetry of the space group as to eliminate the symmetry elements relating the disordered parts
- Eliminate one part by using the 'delete atom' option of the edit functionality.
- Export one part of the disordered structure to a new cif-file

Example

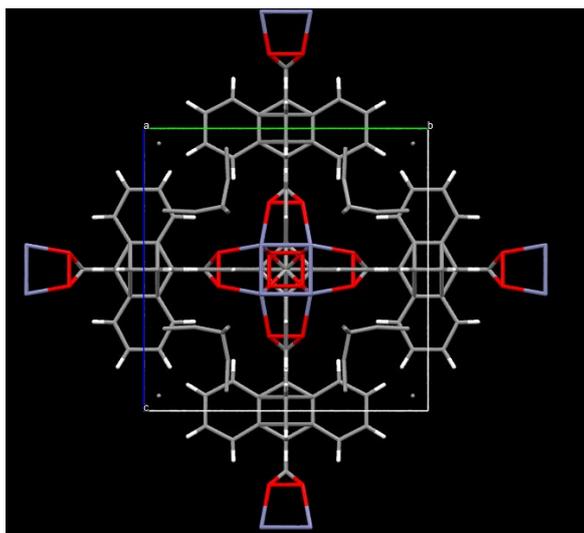
Recent cif files may contain detailed information about the disorder present in a structure. This information can be used to display only one part of the structure. Modelling and DFT program mostly require non-disordered structures for starting the calculations and it becomes thus necessary to create such files using the tools available in Mercury. At this moment Mercury does not allow – as is possible in other programs – to rapidly select a specific part of a disordered structure using the disorder-specific information present in the cif-file. This is foreseen in an upcoming update of Mercury.

Older cif-files do not have disorder-specific information integrated in the cif-file and a tool for selecting a specific part of the structures can thus not be developed for such files. In this case one disordered part – if it is sufficiently small and not related by symmetry to the other part – can be eliminated by using the ‘edit’ tool available in Mercury, followed by deleting the atoms of the unwanted part.

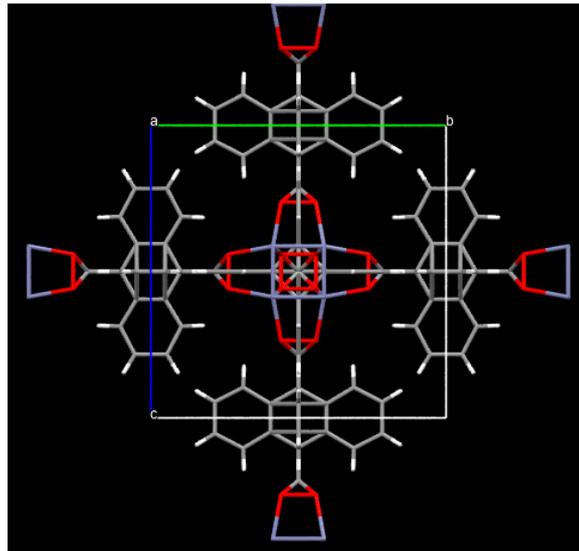
It is more difficult when the disordered parts are connected by symmetry elements belonging to the space group symmetry. The best way to pursue is by firstly eliminating the symmetry elements connecting the disordered parts and then eliminating the atoms of the unwanted part using the ‘edit’ tool.

The case to be explored is EDUTOM, an isorecticular MOF published in Science in 2002. The cif-file that can be found in the CSD is overcomplete, i.e. it contains also the atoms that are generated by symmetry and it contains massive symmetry-related disorder not specified in the cif-file. It is not necessary to explicitly take care of the overcomplete cif-file, since Mercury does this automatically when it loads the file. However, it may be necessary to export the asymmetric for use in modelling programs, since they may be confused by the presence of atoms in the cif-file that are generated by symmetry.

1. Open Mercury and load EDUTOM, perform a simple packing of the unit cell and rotate it slightly to see the triphenyl groups disordered over two perpendicular position as well as the central metal center with surrounding ligands. Note also disordered solvent fragments in the cavities of the MOF.



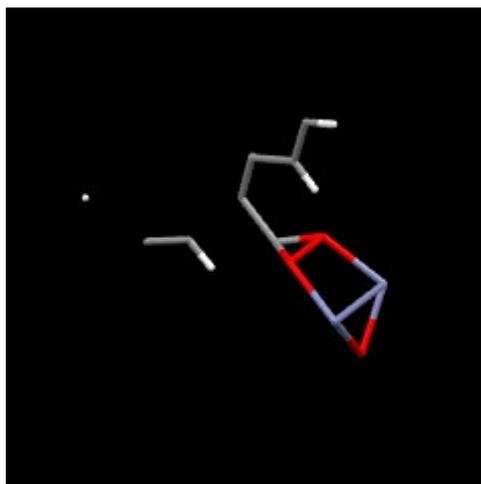
2. For the modelling studies the disordered solvent fragments are not necessary, so delete them by choosing in the top menu ‘Edit’ -> ‘Edit structure’ -> ‘Remove Atoms and Bonds’. Be sure that you are in ‘Pick Atoms’ mode and that you have displayed only the asymmetric unit of the unit cell. Once finished display again a unit cell packing:



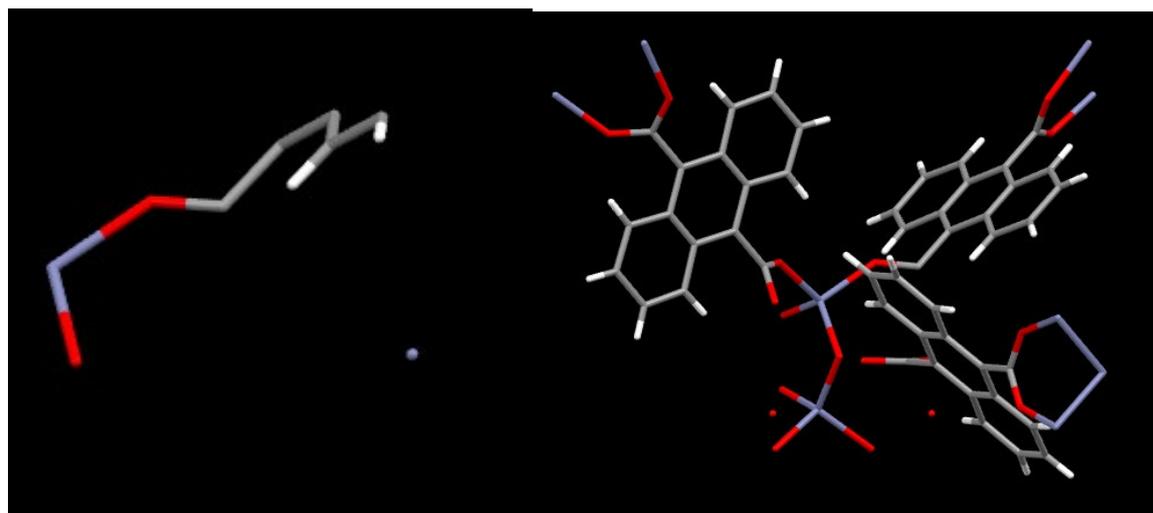
3. In order to check how the disordered groups are symmetry-related to each other, use the 'Display' -> 'Symmetry elements' and see by deselecting one by one the symmetry elements, how the triphenyl groups are related to each other.

4. We can clearly see a diagonal mirror plane that relates the two perpendicular triphenyl groups, so it will be sufficient if we could remove all mirror-planes in the space group symmetry. The space group symmetry is $Pm\bar{3}m$ – this can be checked in by following 'Display' -> 'More information' -> 'Structure information'.

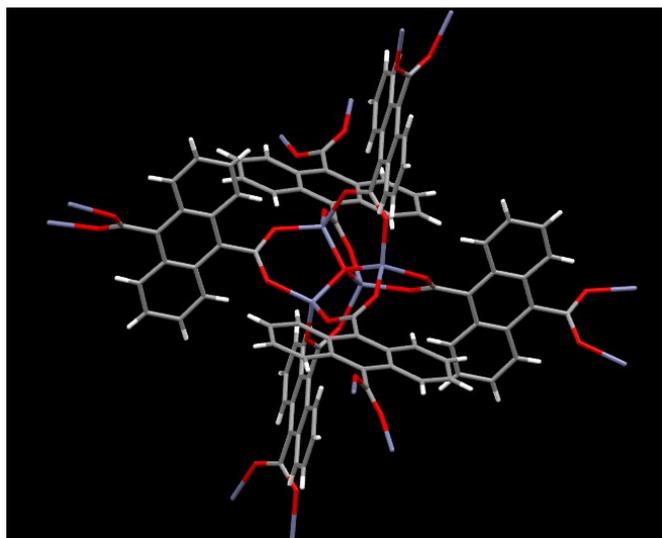
5. Disable the display of symmetry elements by unchecking 'Show symmetry elements', display only the structure in the asymmetric unit, and choose a proper subgroup in 'Edit' -> 'Change spacegroup to subgroup'. Choose in the list of available subgroups a cubic subgroup, e.g. $P23$. $P432$ may work as well. Mercury generates some extra atoms that are required by the loss of the mirror planes.



6. Now start carefully deleting atoms from one triphenyl group. The best is to switch regularly between an asymmetric unit view and a full molecular view. You will see that when you delete one atom, that another one will pop up elsewhere. This is an atom that is generated by Mercury by applying the symmetry operators of the new space group P23. You should delete the newly generated atoms also until no more atoms pop up. You can also decide to remain in molecular view to delete unwanted parts. The risk is that the asymmetric unit fragment is not completely connected anymore, or even the molecular unit as shown here:



This does not mean that the generated structure is wrong, only that the atoms of the asymmetric unit do not form a fully connected set and that they become only a fully completed set if packed in the unit cell. Note also that the obtained structure may not correspond to what you actually wanted to achieve. In the structure below all phenyl groups are perpendicular to each other, but you may want to have triphenyl which are aligned within the plane. In that case you need to do the delete process differently.



6. As a final step you need to export the generated structure to a new cif file. Display first the asymmetric unit and then do 'File'->'Save as' (cif format) and the asymmetric unit is saved within a new file together with the symmetry operators of space group P23. This file should be used as input file for your modelling program.

7. As an extra exercise continue the process to generate a similar structure but with biphenyl groups.

3. Assessing the stability of paracetamol

Objectives

- To use the motif search wizard to set up a search for specific hydrogen-bonded motifs in order to obtain occurrence frequency tables for these motifs.

Steps Required

- Follow the motif search wizard and analyze the data

Example

Paracetamol (HXACAN) has three different polymorphs, all of them having similar hydrogen-bonded patterns. Why is this? The procedure followed here could also be used for selecting molecules which could have a high propensity to co-crystallize with a target molecule.

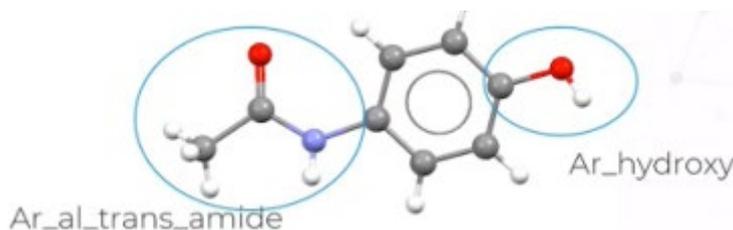
1. Open Mercury and –optionally – load the structure of paracetamol (HXACAN).

HXACAN14 P21/a
 HXACAN15 P21/a
 HXACAN16 P21/a
 HXACAN17 P21/a
 HXACAN18 P21/a
 HXACAN19 P21/a
 HXACAN20 P21/n
 HXACAN21 Pcab
 HXACAN22 Pcab
 HXACAN23 Pcab
 HXACAN24 Pcab
 HXACAN25 Pbca

Many different structures are reported, but most of them are actually the same (same spacegroup in a different setting or measured at another temperature). Check HXACAN20 and 21 which crystallize in different space groups. What can you say about the hydrogen-bond patterns?

2. Start the motif search wizard: upper menu -> CSD_Materials -> Search -> Motifs

3. Create new motifs -> Add -> select C, H, O, N and then select ar_al_trans_amide and ar_hydroxy as functional groups for the motif search query.



4. Add the following contacts – ar_al_trans_amide <-> ar_al_trans_amide ar_hydroxy <-> ar_al_trans_amide, ar_al_trans_amide <-> ar_hydroxy, and ar_hydroxy <-> ar_hydroxy. Check the guesses the wizard makes about the donor and acceptor groups. If the wizard does it wrong you can click on the donor or acceptor groups to get it right.

5. In the types of contacts dialog, choose 'inter'

6. Skip the 'advanced options' by clicking on 'next'

7. Continue until the select structures dialog.

Select the main database and eventual updates you installed.

8. Give a name for the search and start it.

9. This takes quite a while (10-15 minutes), but you see already the numbers being updated in the 'Searches' window.

motif	# structures	% frequency
> inter_A-A- (...	584	12.9
> inter_A-B- (A...	21	61.8
> inter_A-B- (A...	5	14.7
> inter_A-B- (A...	201	32.4

10. You can click on each found structure and check the hydrogen-bond interaction. You can also check negative results. What is the conclusion?

CONQUEST exercises:

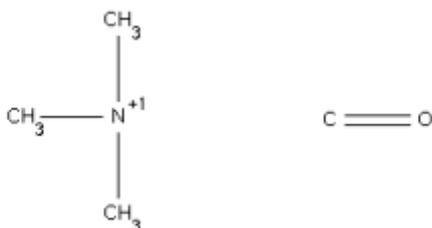
1. Combining searches

Find structures that:

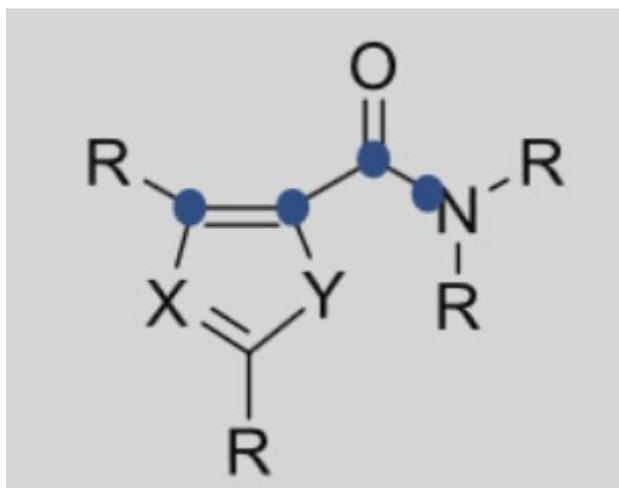
1. contain 1 or 2 osmium atoms and >6 oxygen atoms;
2. do not belong to space group P21/c;
3. were determined either by neutron diffraction or with an R-factor ≤ 0.06

2. Searching with geometrical constraints

Search for pharmacophores where the distance between the cationic nitrogen and the center of the carbonyl bond is in the range 5.0 - 5.8Å. This is the distance range required between these two functional groups when acetylcholine and its analogues bind to the enzyme acetylcholinesterase.



3. Influence of substitutions in heterocyclic γ -aminoacids on torsion angles



Search for structures with this motif and track the indicated torsion angle C β -C α -CO-N. X, Y could be O, S, N(H). No cyclicity between R's.

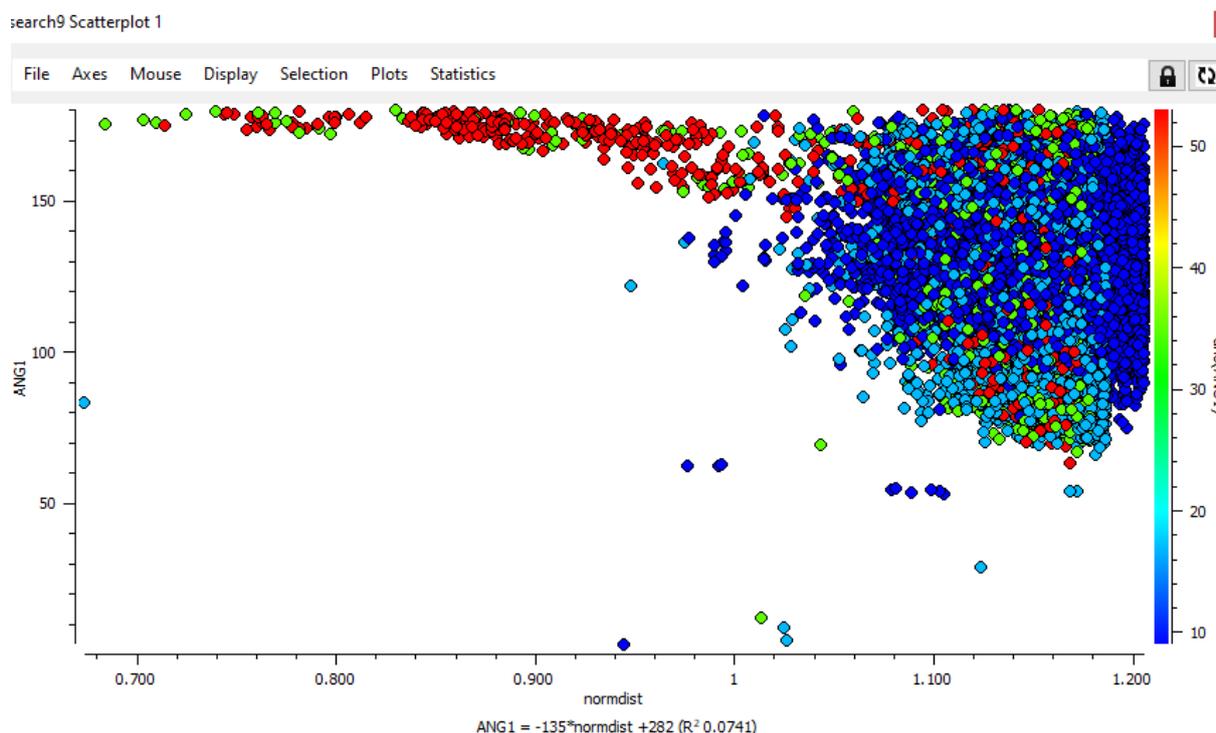
Example is from Chaise, Maillard, Legrand (2021).

4. Exploring the landscape of Sigma-hole Interactions

Goals

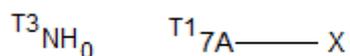
Search for s-hole interactions R-X...NR₃ between sp³ nitrogen and X=F, Cl, Br, and I. Analyze the data in Mercury by plotting the correlation between the normalized distance X...N between N and the halogen and the angle R-X...N. Color the data points per halogen. 10996 hits.

You should get the following plot:

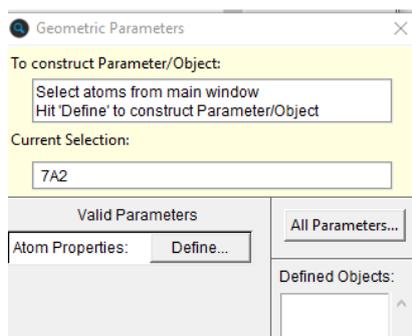


Hints

In Conquest draw your interaction in this way :



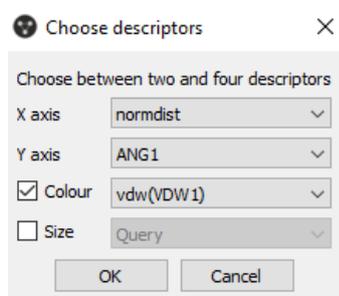
Add 3D parameters for the distance X...N between N and the halogen and for the angle R-X...N. Add also information about the nature of the halogen atom, at least its Van der Waals radius. This will be useful for the normalization of the distances. The van der Waals radius can be added by defining a new '3D' parameter and selecting just the halogen atom 7A.



Then go to 'Define Atom properties' and add at least the Van der Waals radius.

Do the Conquest search – add filters if you want (e.g. only organics) - and export the results to a .c2m file, or click directly on File -> Analyse data in Mercury.

In Mercury add a column to the spreadsheet by tools->calculator and add the normalized distance $D_{norm} = \text{dist1}/(\text{RvdW}(7A) + \text{RvdW}(N))$. Then make a scatterplot while selecting as color descriptor the van der Waals radius of the halogen.



Something different: tautomerism stories (see also the MOGUL slide on how to make a distinction between tautomeric forms)

The most famous story about tautomers in the history of science occurred in the early 1950s in Cambridge. Watson and Crick were trying to propose a structure for DNA, but had been failing for some time. They were, however, fortunate enough to be sharing an office with the American theoretical chemist Jerry Donahue. One Wednesday afternoon they discussed the possible tautomeric forms of the bases in DNA. Jerry Donahue told Watson and Crick that the literature was likely to be wrong and what the most probable tautomers for G, C, A and T were. When Jim Watson came in to work at 9.30 am on Saturday morning he had cardboard models for the four bases in the 'correct' tautomeric forms and, by the time that Francis Crick arrived for work at 10.30 am, Jim had worked out the classical G–C, A–T base pairing (J. Watson seminar, LMB, Cambridge, 9th June 2016). As they subsequently wrote in their famous paper

If it is assumed that the bases only occur in the most plausible tautomeric forms ... it is found that only one specific pair of bases can bond together