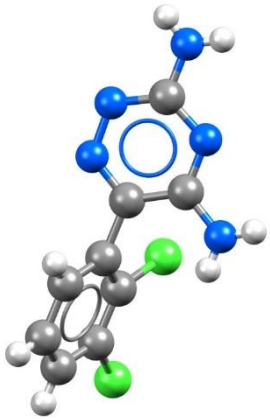
Polymorph Risk Assessment

MAT-001

2020.3 CSD Release



CCDC
advancing structural science

Table of Contents

Introduction

Molecules have the potential to adopt multiple different packing arrangements in the solid state, a phenomenon known as polymorphism, which have significant influence on a material's performance. Exploring the polymorphic landscape and understanding the relative stability of polymorphs is an important process, especially in early-stage formulation in the pharmaceutical industry.

The Hydrogen Bond Propensity (HBP) tool in Mercury can be used, to evaluate the relative likelihoods of possible H-bonding networks in any observed polymorphs of a target system.

Before beginning this workshop, ensure that you have a registered copy of CSD-Materials or CSD-Enterprise installed on your computer. Please contact your site administrator or workshop host for further information.

Objectives

- Familiarise with the Hydrogen Bond Propensities tool.
- Learn how to perform a HBP analysis and how to read and interpret the results.
- Explore how HPB can be used in polymorphs analysis.
- Explore complementary approaches to assess solid forms (namely, Mogul and Full Interaction Maps).

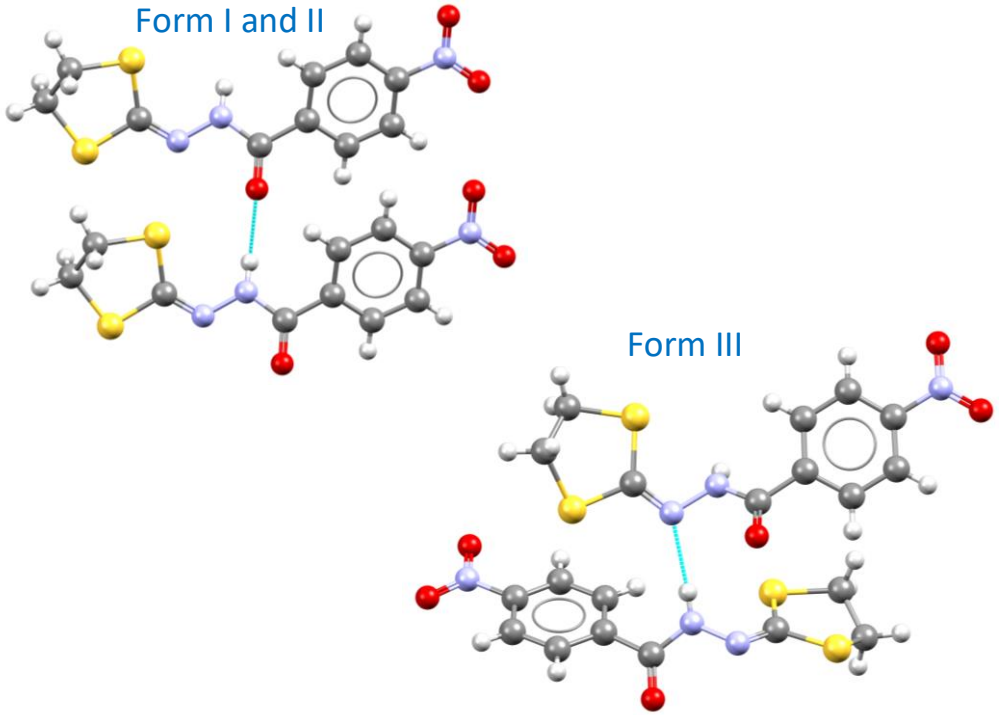
This workshop will take approximately **1-1.5** hours to be completed.

Pre-required skills

Familiarity with the Mercury interface is important; you can access the Visualization in Mercury self-guided workshop here.

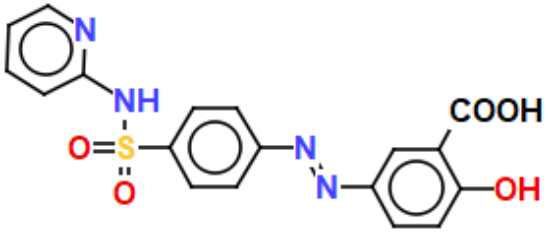
Materials

There are no additional materials required for this workshop.

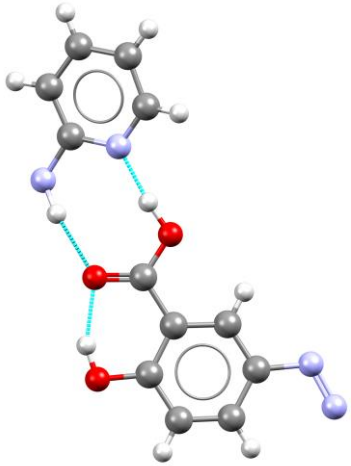


Example 1 A monomorphic system

Sulfasalazine is used to treat ulcerative colitis and Crohn's diseases. Only one polymorph has been reported so far for the amide tautomer of this compound. In this example we will investigate the polymorphic landscape of sulfasalazine and assess the potential for polymorph formation.



*Sulfasalazine (refcode **QIJZOY**)*



Examine H-bonding network

In this section we will examine the potential hydrogen bond donors and acceptors present in sulfasalazine.

1. Start Mercury by double-clicking the icon on your Desktop or navigating from the Start Menu (Start > CCDC > Mercury)

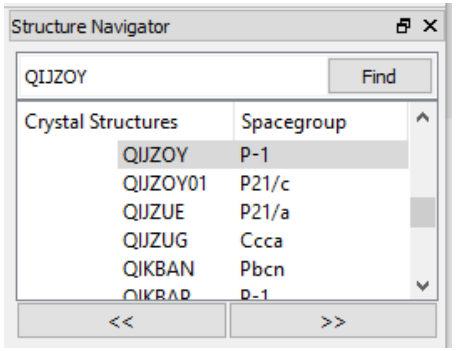
2. In the **Structure Navigator** window, type the refcode *QIJZOY*, to bring up the structure of sulfasalazine amide tautomer.

3. The structure will be displayed in the 3D visualiser. There are 3 potential donors and 6 acceptors.

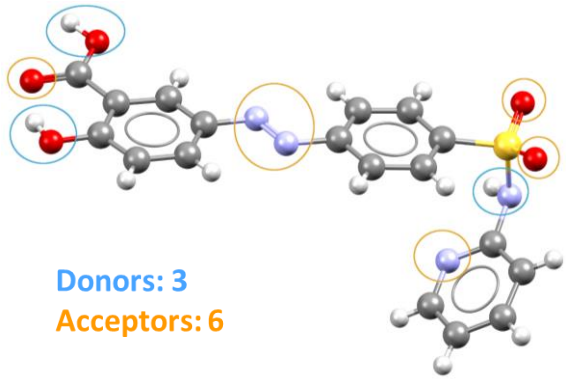
4. Toggle on the **H-Bond** check box in the *Display Options* to investigate how many of the potential donors and acceptors are utilised by sulfasalazine.

1

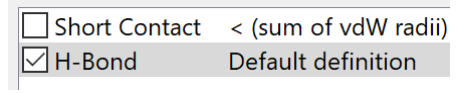
2



3



4



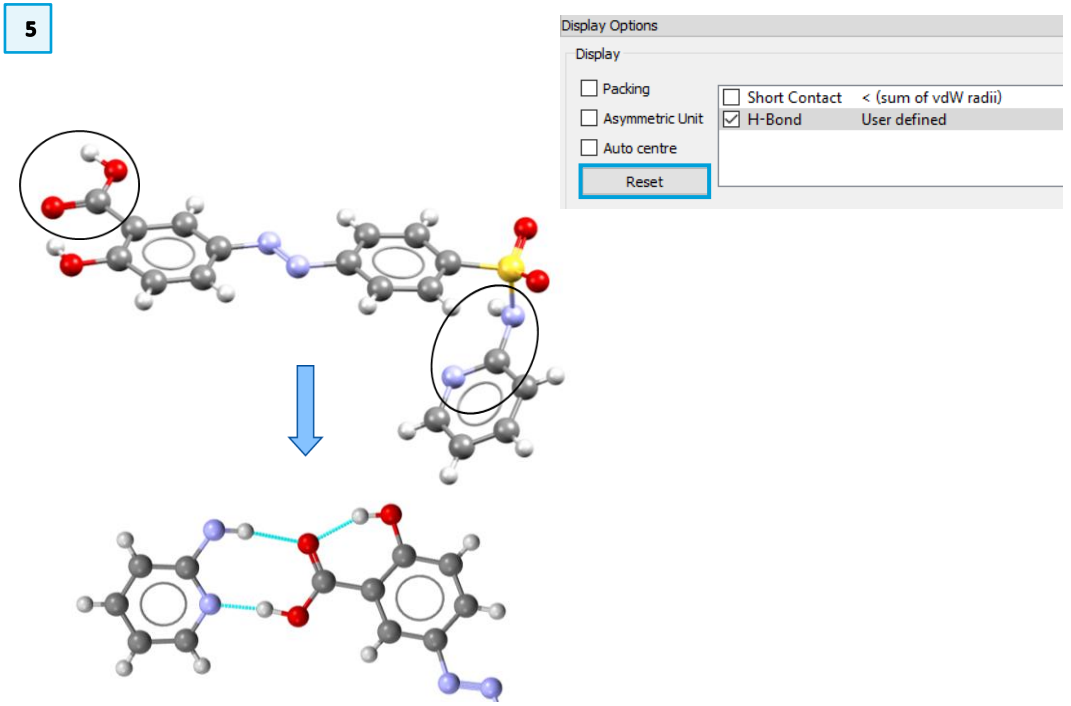
5. Two of the acceptors and two donors are used in intermolecular interactions, forming centrosymmetric dimers involving the carboxylic acid and pyridylamino functional groups. An intramolecular hydrogen bond is also formed between the hydroxyl group and the O atom of the carboxylic group. Press **Reset** button in the *Display Option* dialogue box before continuing.

Calculate H-bond propensity

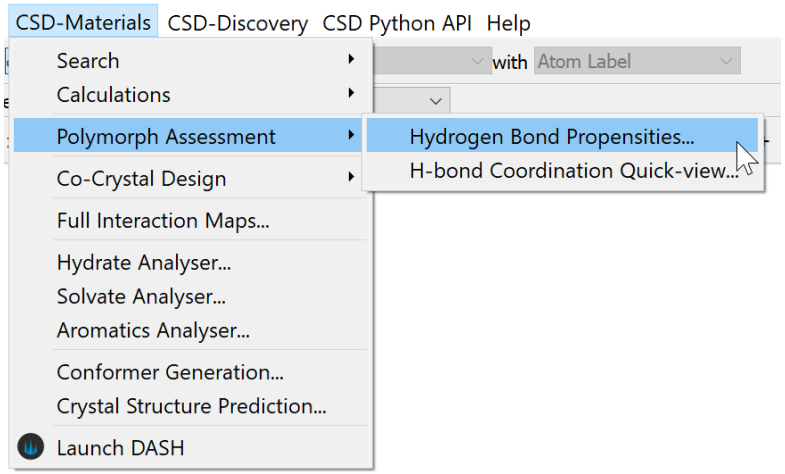
6. From the top-level menu select **CSD-Materials > Polymorph Assessment > Hydrogen Bond Propensities…**

7. In the *Propensity Prediction Wizard* select a working directory by clicking on **Browse…** The potential hydrogen bond donor and acceptor atoms are automatically identified and linked to their functional groups. Three donors have been identified: N2 as sulfonamide_1, O3 as ar_cooh_1, and O5 as ar oh. Eight acceptors have also been identified. Note that O3 and O5 are identified as both donor and acceptor as standard for a hydroxy group. If you want to adjust the atoms involved as donors or acceptors you can use the advanced settings: toggle on the **Show advanced options** check box and click **Edit…** However, for this example, we will use the default values.

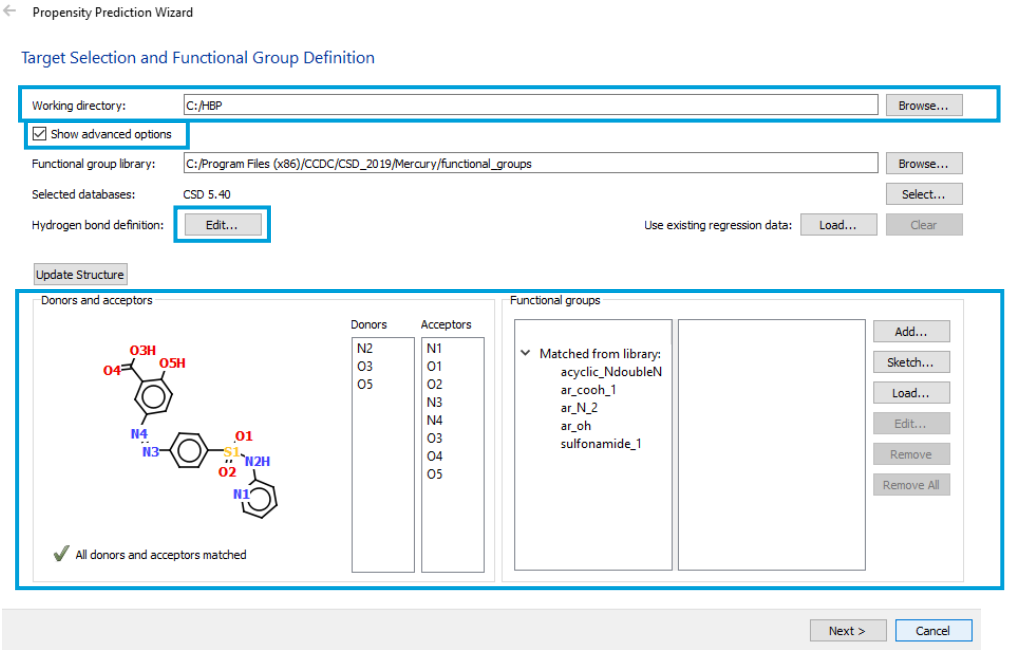
5



6



7

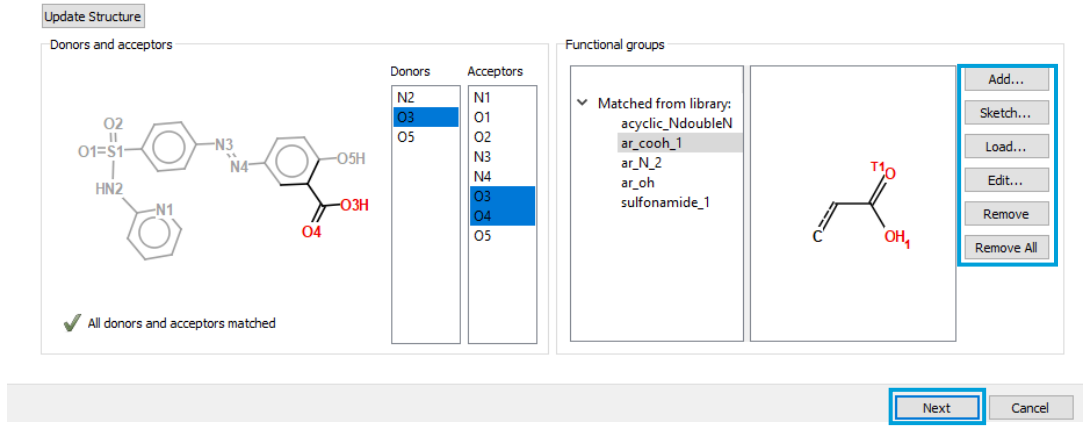


8. The *Donors* and *Acceptors* atoms can be highlighted in the 2D chemical diagram by selecting them from the list. You can also highlight a functional group from the *Match from library* list; the corresponding atoms will be automatically highlighted in the *Donors/Acceptors* lists. The functional group as defined will appear in the second window of the *Functional groups* dialogue box. You can adjust the functional groups if desired by using the buttons on the right-hand side **Add…**, **Sketch…**, etc. We will leave all the default settings for this example and click **Next**

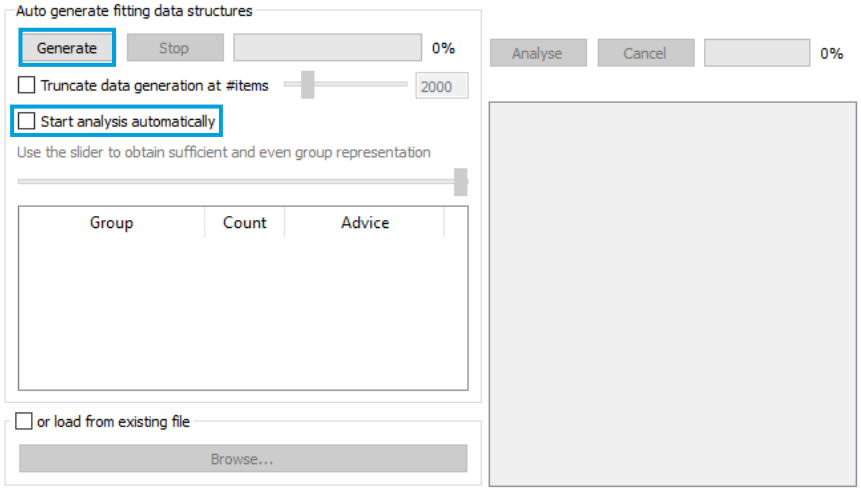
9. Ensure that the **Start analysis automatically** check box is unchecked and click **Generate**. As the training set (generated fitting data) starts to be populated with CSD structures the functional groups and an indication of their **Count** and **Advice** can be seen.

10. When the run is finished, the total number of structures found for each group is listed. The numbers for each functional group can be uneven - it is a good practice to aim for a model with groups evenly represented and around 300-400 observations per group when possible to balance enough data with chemical specificity. Be aware that your results/counts may differ based on the different data release. These results were obtained with a version prior 2020.3.

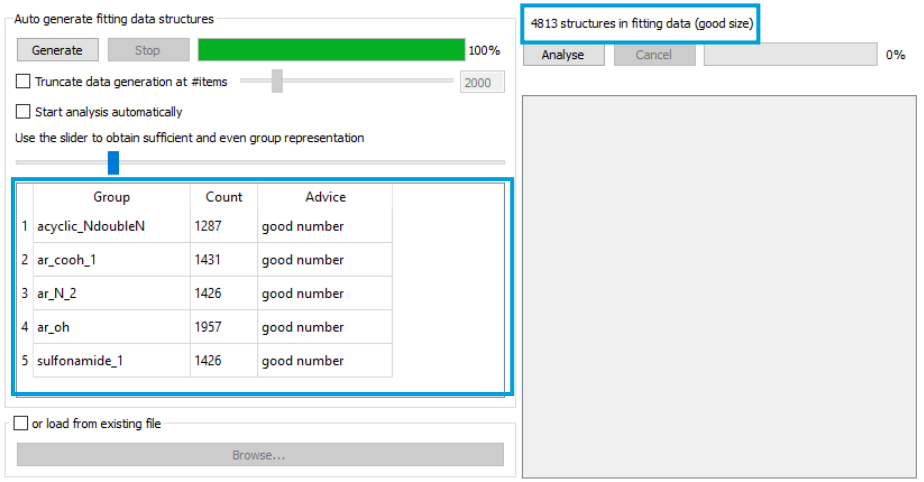
8



9



10

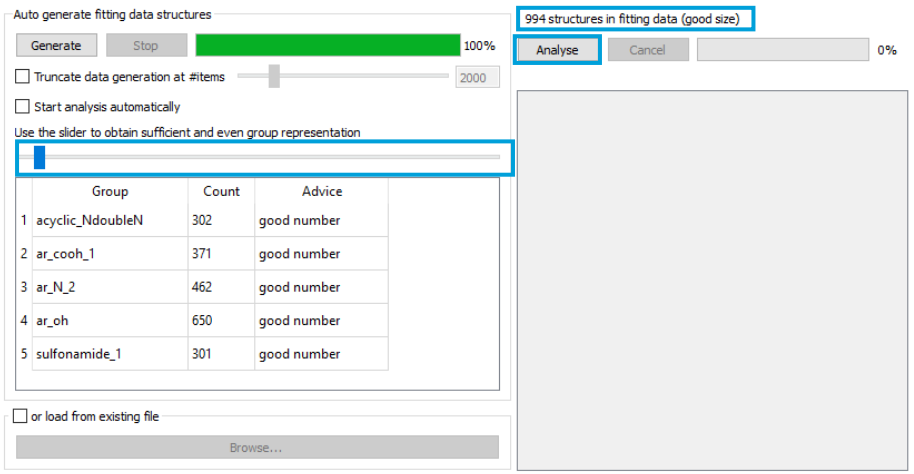


11. When the run is finished, adjust the group number by using the slider highlighted in blue. This allows you to remove or add structures until a more even set of data is obtained. In general, around 300-400 structures per functional group should be enough. Select around 800-1000 structures in total, with around 300-400 structures per functional group, then click **Analyse**.

12. When the analysis is finished the number of the True and False outcomes will be listed. If there are very low numbers for True or False, they will be automatically ticked in the **Ignore?** checkboxes. There are no very low values in this example. Click the **Fit Model >** button to continue.

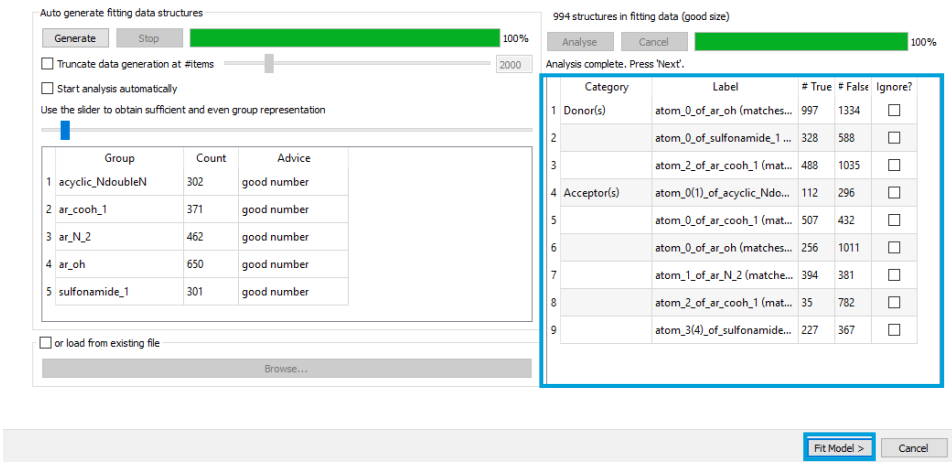
13. For this example, the Area under the ROC (receiver operating characteristic) curve (AUC) should be around 0.82. To achieve a good H-bond propensity calculation you should always aim for an AUC of around 0.75 or above. Click **Accept & Calculate** to continue.

11



	Group	Count	Advice
1	acyclic_NdoubleN	302	good number
2	ar_cooh_1	371	good number
3	ar_N_2	462	good number
4	ar_oh	650	good number
5	sulfonamide_1	301	good number

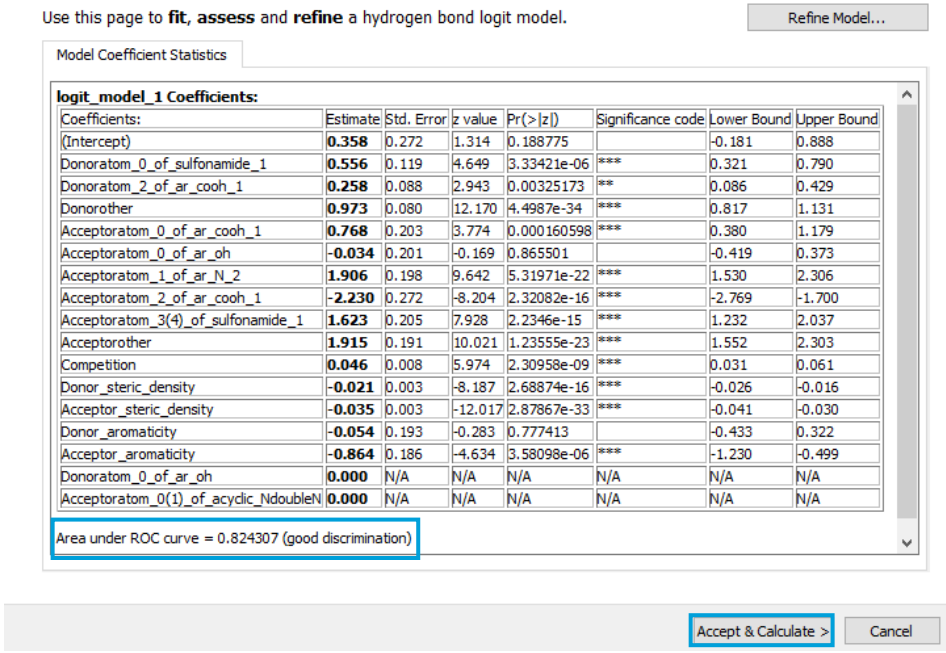
12



	Group	Count	Advice
1	acyclic_NdoubleN	302	good number
2	ar_cooh_1	371	good number
3	ar_N_2	462	good number
4	ar_oh	650	good number
5	sulfonamide_1	301	good number

	Category	Label	# True	# False	Ignore?
1	Donor(s)	atom_0_of_ar_oh (matches...	997	1334	☐
2		atom_0_of_sulfonamide_1 ...	328	588	☐
3		atom_2_of_ar_cooh_1 (mat...	488	1035	☐
4	Acceptor(s)	atom_0(1)_of_acyclic_Ndo...	112	296	☐
5		atom_0_of_ar_cooh_1 (mat...	507	432	☐
6		atom_0_of_ar_oh (matches...	256	1011	☐
7		atom_1_of_ar_N_2 (matche...	394	381	☐
8		atom_2_of_ar_cooh_1 (mat...	35	782	☐
9		atom_3(4)_of_sulfonamide...	227	367	☐

13



Coefficients:	Estimate	Std. Error	z value	Pr(>|z|)	Significance code	Lower Bound	Upper Bound
(Intercept)	**0.358**	0.272	1.314	0.188775		-0.181	0.888
Donoratom_0_of_sulfonamide_1	**0.556**	0.119	4.649	3.33421e-06	***	0.321	0.790
Donoratom_2_of_ar_cooh_1	**0.258**	0.088	2.943	0.00325173	**	0.086	0.429
Donorother	**0.973**	0.080	12.170	4.4987e-34	***	0.817	1.131
Acceptoratom_0_of_ar_cooh_1	**0.768**	0.203	3.774	0.000160598	***	0.380	1.179
Acceptoratom_0_of_ar_oh	**-0.034**	0.201	-0.169	0.865501		-0.419	0.373
Acceptoratom_1_of_ar_N_2	**1.906**	0.198	9.642	5.31971e-22	***	1.530	2.306
Acceptoratom_2_of_ar_cooh_1	**-2.230**	0.272	-8.204	2.32082e-16	***	-2.769	-1.700
Acceptoratom_3(4)_of_sulfonamide_1	**1.623**	0.205	7.928	2.2346e-15	***	1.232	2.037
Acceptorother	**1.915**	0.191	10.021	1.23555e-23	***	1.552	2.303
Competition	**0.046**	0.008	5.974	2.30958e-09	***	0.031	0.061
Donor_steric_density	**-0.021**	0.003	-8.187	2.68874e-16	***	-0.026	-0.016
Acceptor_steric_density	**-0.035**	0.003	-12.017	2.87867e-33	***	-0.041	-0.030
Donor_aromaticity	**-0.054**	0.193	-0.283	0.777413		-0.433	0.322
Acceptor_aromaticity	**-0.864**	0.186	-4.634	3.58098e-06	***	-1.230	-0.499
Donoratom_0_of_ar_oh	**0.000**	N/A	N/A	N/A	N/A	N/A	N/A
Acceptoratom_0(1)_of_acyclic_NdoubleN	**0.000**	N/A	N/A	N/A	N/A	N/A	N/A

Summary of HBP results

14. The Chart:

- plots Mean H-bond Propensity vs the Mean H-Bond Co-ordination
- target structure is represented as a magenta circle
- to zoom use the magnifying glass icon in the lower left-hand corner of the wizard, to go back to the default option press **Reset**
- the most likely H-bonding network is displayed in the lower-right corner, the outcome should be read along the diagonal
- QIJZOY has the most likely H-bonding network for sulfasalazine listed first in the lower right-hand corner
- click on the points to highlight the H-bond network in blue in the *Propensity score* table

Propensity Score Table

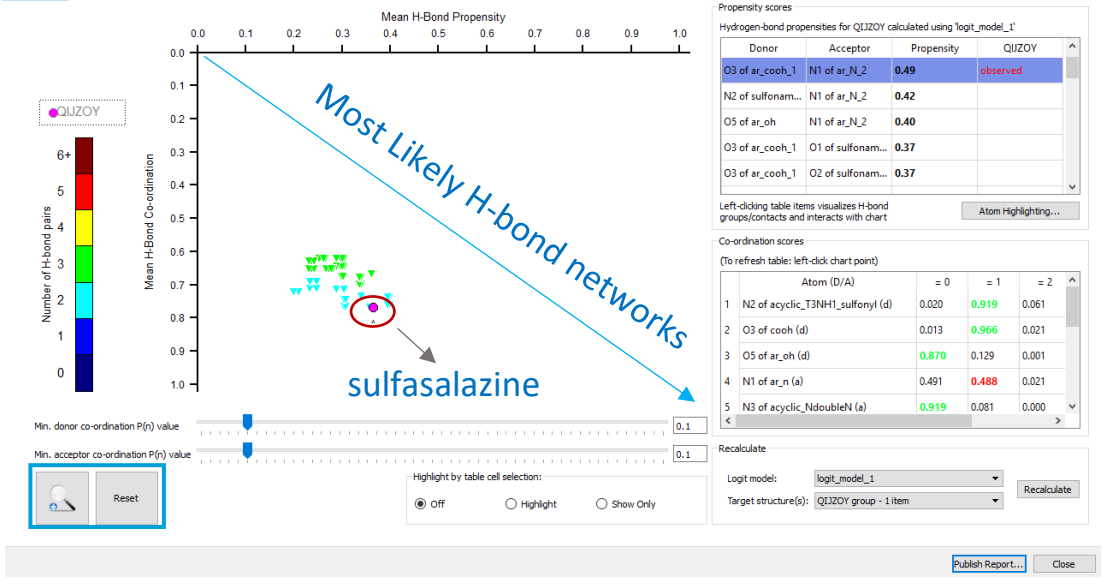
- the most likely H-bonding network will score the highest propensity and will be listed first in the table
- the H-bonds present in the targeted structure are marked as observed
- the table is interactive, clicking on **observed** will highlight the donor and acceptor group in the 3D visualizer, clicking on an atom label, in either the *Donor* or *Acceptor* columns, will highlight the functional group and label the atom in the 3D visualizer
- The *Propensity scores* table shows all possible H-bond interactions for sulfasalazine, with O3-H13…N1 giving the highest propensity. You can see this interaction is observed in the QIJZOY structure.

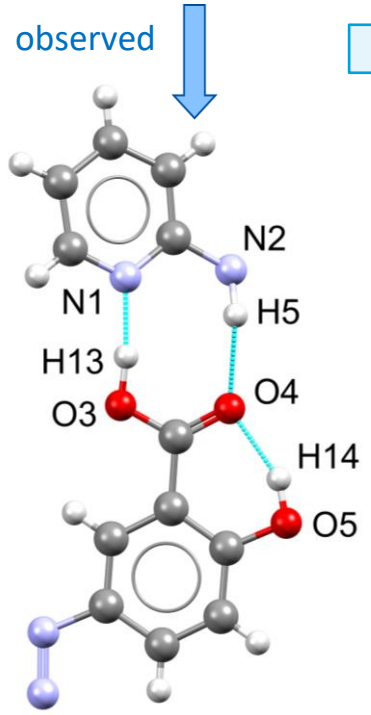
15. *Co-ordination Scores Table:*

- (a) stands for acceptor and (d) for donor,
- =0, =1, =2 denotes the number of times a functional group donates or accepts
- The numbers that are coloured relate to the outcome present in the selected H-bonding network, if this is green it indicates that the outcome is optimal, whereas if it's red that indicates the outcome is sub-optimal.
- For QIJZOY all the H-bonds present are optimal apart from N1 of the ar_n(a) group. Based on CSD data for this type of atom in this environment, it is more likely not to accept any H-bonds.

In conclusion, QIJZOY was found to be the most likely polymorph based on both propensity and coordination, and this agrees with the experiments: only one polymorph of the amine tautomer of sulfasalazine has been found so far.

14





15

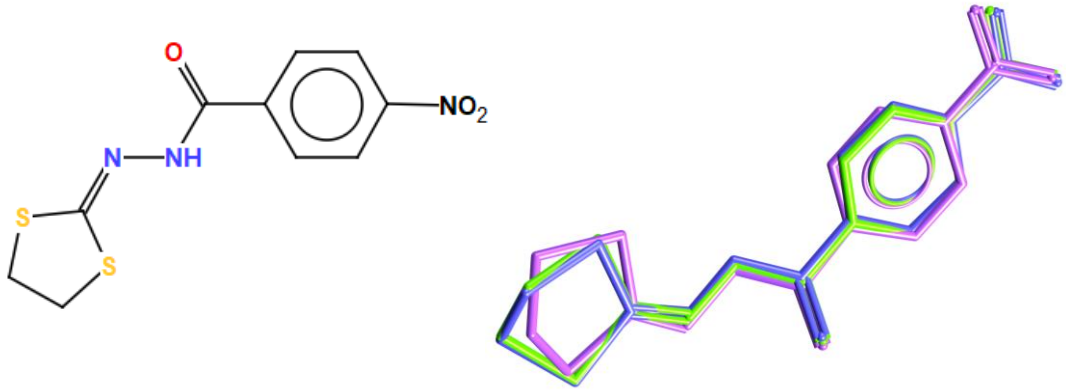
Co-ordination scores

(To refresh table: left-click chart point)

	Atom (D/A)	= 0	= 1	= 2
1	N2 of acyclic_T3NH1_sulfonyl (d)	0.020	0.919	0.061
2	O3 of cooh (d)	0.013	0.966	0.021
3	O5 of ar_oh (d)	0.870	0.129	0.001
4	N1 of ar_n (a)	0.491	0.488	0.021
5	N3 of acyclic_NdoubleN (a)	0.919	0.081	0.000
6	N4 of acyclic_NdoubleN (a)	0.904	0.096	0.000
7	O1 of acyclic_T3NH1_sulfonyl (a)	0.506	0.478	0.016
8	O2 of acyclic_T3NH1_sulfonyl (a)	0.631	0.359	0.010
9	O3 of cooh (a)	0.961	0.038	0.001
10	O4 of cooh (a)	0.464	0.532	0.004
11	O5 of ar_oh (a)	0.762	0.229	0.010

Example 2 A polymorphic system

N'-(1,3-dithiolan-2-ylidene)-4-nitrobenzohydrazide, a potentially tuberculostatic agent, is known to crystallise in three polymorphic forms. The first two polymorphs (refcodes *DEDMUX* and *DEDMUX01*) form identical H-bond networks (N-H…O) and have similar geometry, while the third polymorph (refcode *DEDMUX02*) forms a N-H…N H-bond network and the geometry of the dithiolane ring is largely different.
In this example we will use the HBP tool to assess the relative likelihoods of the H-bond networks observed in the three polymorphs.



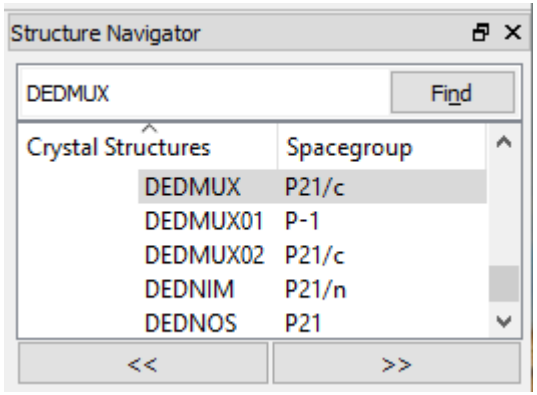
*N'-(1,3-dithiolan-2-ylidene)-4-nitrobenzohydrazide (refcode **DEDMUX**). Form III (purple) has a different geometry*

Examine H-bonding network

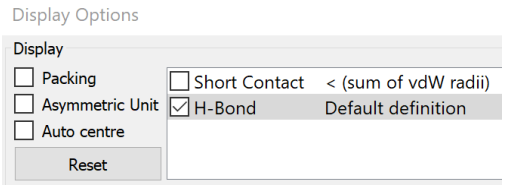
1. Start Mercury by double-clicking the icon on your Desktop or navigating from the Start Menu (Start > CCDC > Mercury)
2. In the **Structure Navigator** window, type the refcode *DEDMUX*, to bring up the structure of the first polymorph.
3. Ensure that the **H-Bond** check box in the *Display Options* area of the Mercury interface is toggled on and expand the contacts for Form I. Note that the N of the amide group acts as donor and the O atom of the carbonyl group as acceptor. The same interactions are present in Form II. You can investigate this by repeating step 2 and loading *DEDMUX01*.
   Load Form III by typing *DEDMUX02* in the **Structure Navigator** window. The H-bond interactions occurs between the N amide and the N imine.
   Check all the possible donors and acceptors. How many are there?

1

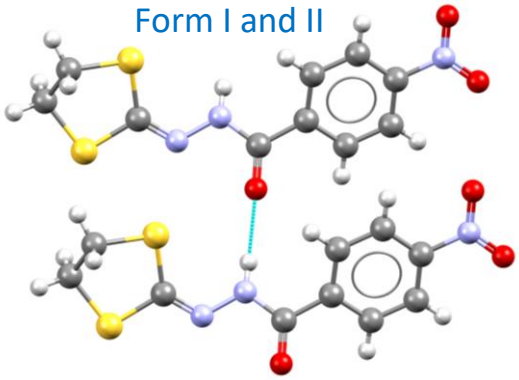
2

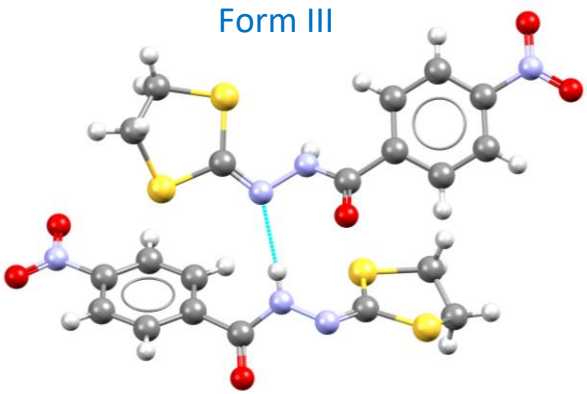


3



H-bond network

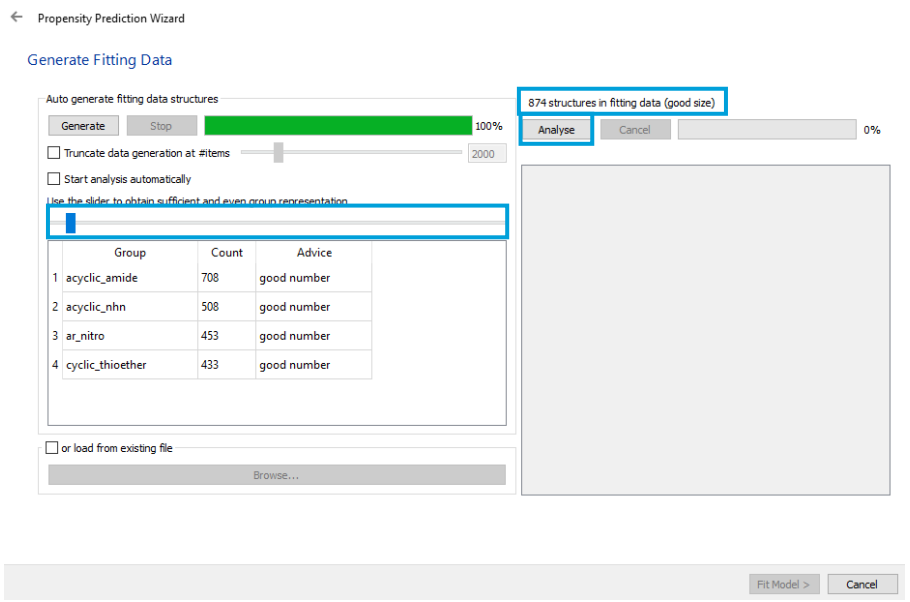




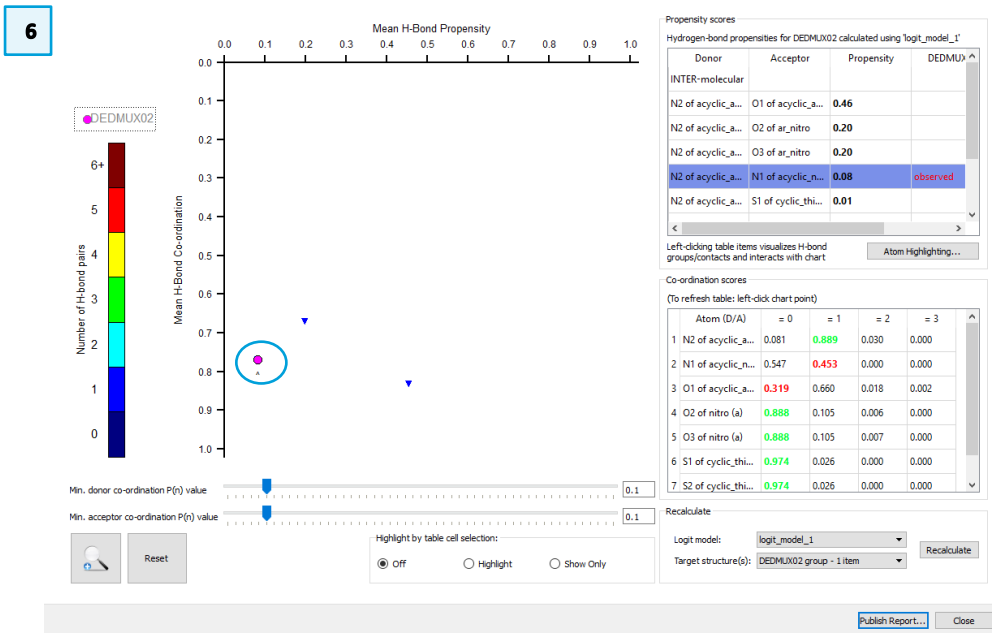
Calculate H-bond propensity

4. Repeat steps 6 to 9 from Example 1 to generate the HBP analysis for Form III.
5. Select around 800-1000 total structures using the slider and then click **Analyse**. After the analysis is finished click **Fit Model**. In the *Model Fitting* wizard click **Accept & Calculate**.
6. Form III is represented as magenta circle in the propensity chart. The N-H…N hydrogen bond interaction present in this form gives a very low propensity score (0.08). If this was the first solid form discovered, you would see that there are clearly other putative H-bonding networks that exhibit both better propensity and better coordination, so the conclusion would be that there is a significant risk of polymorphism based on H-bonding in this case.
7. To see where Forms I and II are located in the chart you can load them by clicking **Target structure(s)** drop-down menu in the *Recalculate* section and then click **Select multiple…** In the *Search Structure Section* dialog box, click the T icon, then tick the box for **Enter refcode family**, then click **OK**. You can see the three DEDMUX refcodes in the **Selected structure(s)** pane. Click **OK**, then click **Recalculate**.

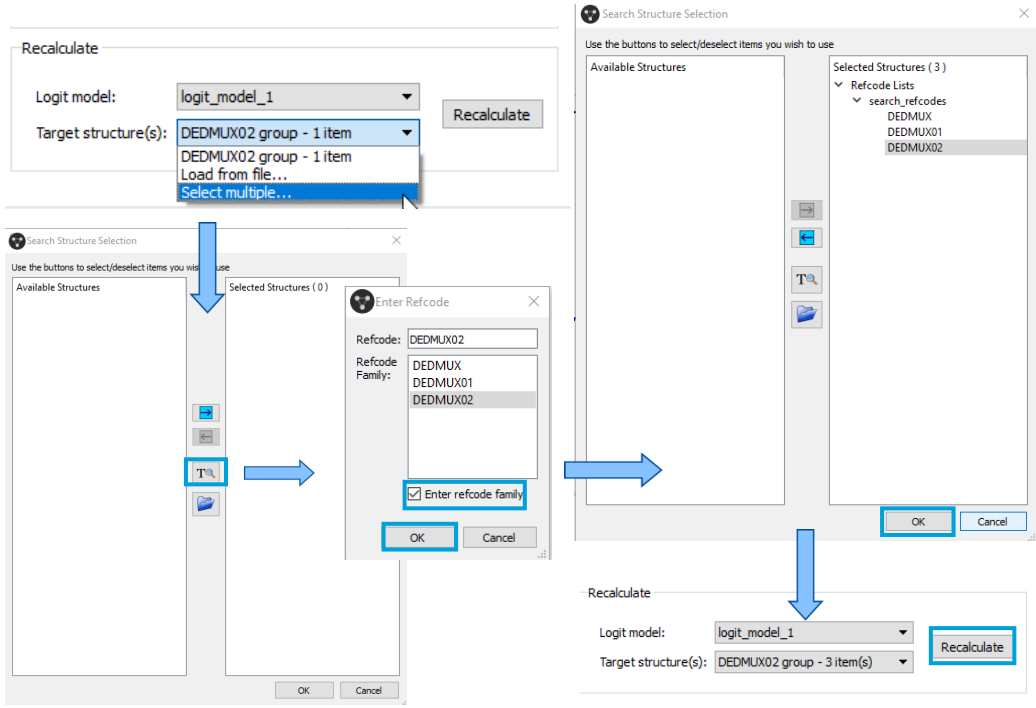
5



6



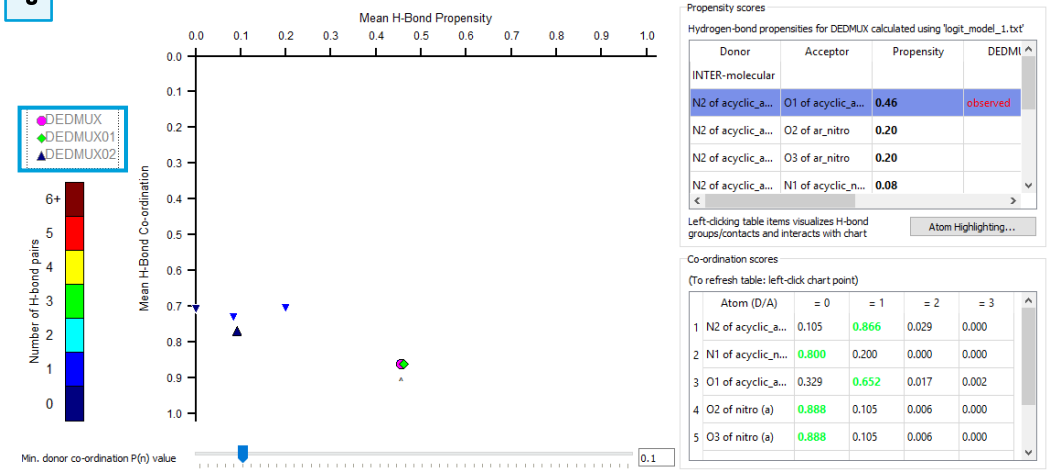
7



8. All three polymorphs are now plotted on the chart. To identify where each polymorph is represented on the chart, check the legend shown on the left-hand side of the dialogue indicating the structures displayed. You can see that Form I and II have the same H-bond network (N-H…O) with the highest propensity and best coordination.
9. If we compare the Co-ordination scores of Forms I and III we can see that there are two sub-optimal acceptors for Form III. N1 donates once but will prefer to donate zero times and O1 accepts zero times but will like to accept once. In Form I the co-ordination scores for all donors and acceptors are optimal.

   In conclusion, one of the polymorphs (DEDMUX02) is observed to have a noticeably less likely H-bonding network than the other two experimentally-observed polymorphs (DEDMUX & DEDMUX01). To evaluate the similarity of the two polymorphs with the same H-bonding network, we would follow this up by looking into the molecular conformations, packing density and the 3D geometry of the intermolecular interactions.

8



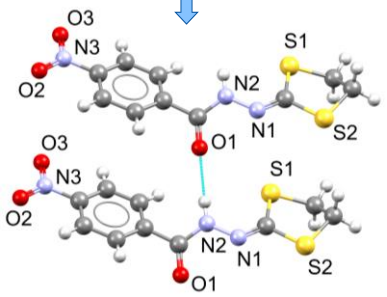
9

Co-ordination scores

(To refresh table: left-click chart point)

	Atom (D/A)	= 0	= 1	= 2	= 3
1	N2 of acyclic_amide (d)	0.105	0.866	0.029	0.000
2	N1 of acyclic_nhn (a)	0.800	0.200	0.000	0.000
3	O1 of acyclic_amide (a)	0.329	0.652	0.017	0.002
4	O2 of nitro (a)	0.888	0.105	0.006	0.000
5	O3 of nitro (a)	0.888	0.105	0.006	0.000
6	S1 of cyclic_thioether (a)	0.974	0.026	0.000	0.000
7	S2 of cyclic_thioether (a)	0.974	0.026	0.000	0.000

Form I

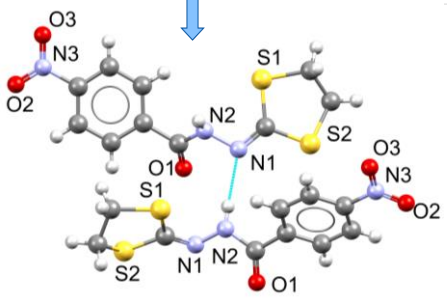


Co-ordination scores

(To refresh table: left-click chart point)

	Atom (D/A)	= 0	= 1	= 2	= 3
1	N2 of acyclic_amide (d)	0.081	0.889	0.030	0.000
2	N1 of acyclic_nhn (a)	0.547	0.453	0.000	0.000
3	O1 of acyclic_amide (a)	0.319	0.660	0.018	0.002
4	O2 of nitro (a)	0.888	0.105	0.006	0.000
5	O3 of nitro (a)	0.888	0.105	0.007	0.000
6	S1 of cyclic_thioether (a)	0.974	0.026	0.000	0.000
7	S2 of cyclic_thioether (a)	0.974	0.026	0.000	0.000

Form III

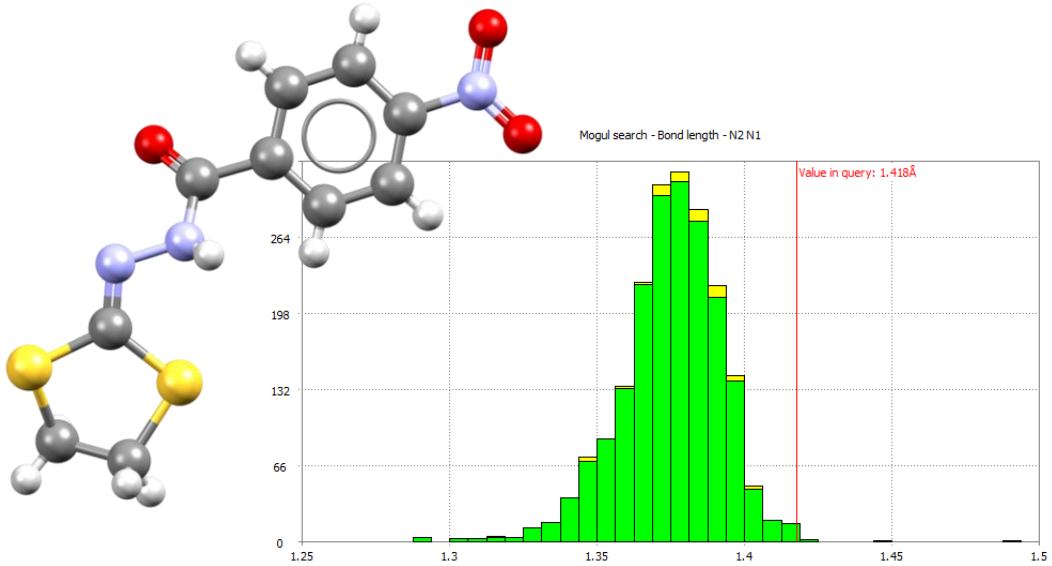


Complementary approaches to assess solid forms

Using Mogul to assess molecular conformation

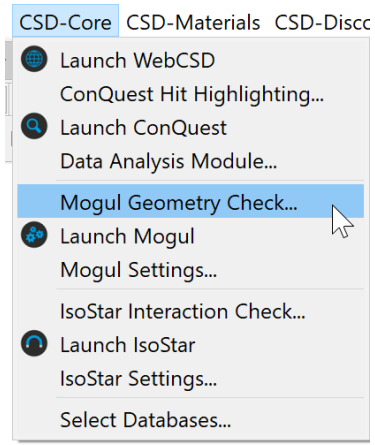
Mogul is able to provide an assessment of a given structure's conformation by comparing it to the data from the hundreds of thousands of structures already in the CSD. By using the statistical distributions of similar fragments, Mogul can confirm your 3D geometry is appropriate, or flag values that are too far outside the norm.

In this example, you will see how to use Mogul to corelate the HBP findings for N'-(1,3-dithiolan-2-ylidene)-4-nitrobenzohydrazide polymorphs with **the geometric performance.** Mogul can be run as a stand-alone application or from the Mercury interface. For this tutorial, we will use Mercury to run Mogul.

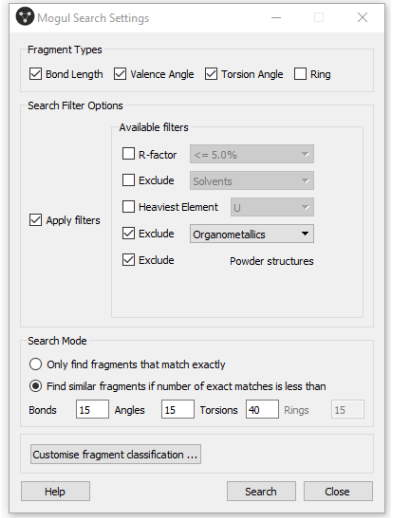


10. Close all the HBP related dialogue boxes and with *DEDMUX02* (Form III) loaded in Mercury, click on CSD-Core menu and then select *Mogul Geometry Check* from the dropdown menu.
11. In the *Mogul Search Settings* dialogue box, you can typically use the defaults in this window, but we can streamline our search by unticking the box for rings and ticking the boxes for Apply Filters, Exclude Organometallics, and Exclude Powder structures. Click **Search** to start.
12. A dialogue box will pop up to warn you that you are going to check the entire molecule. Click **OK** to continue.
13. The search will begin to run. You can follow its progress in the *Search Progress* dialogue box.

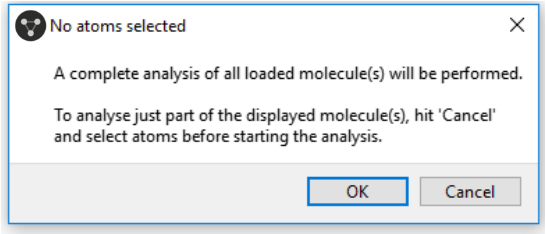
10



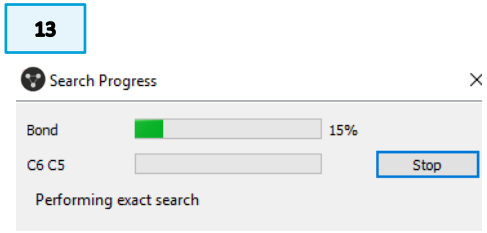
11



12



13



14. When the search is complete, your results will be displayed in the **Mogul Results Viewer.**

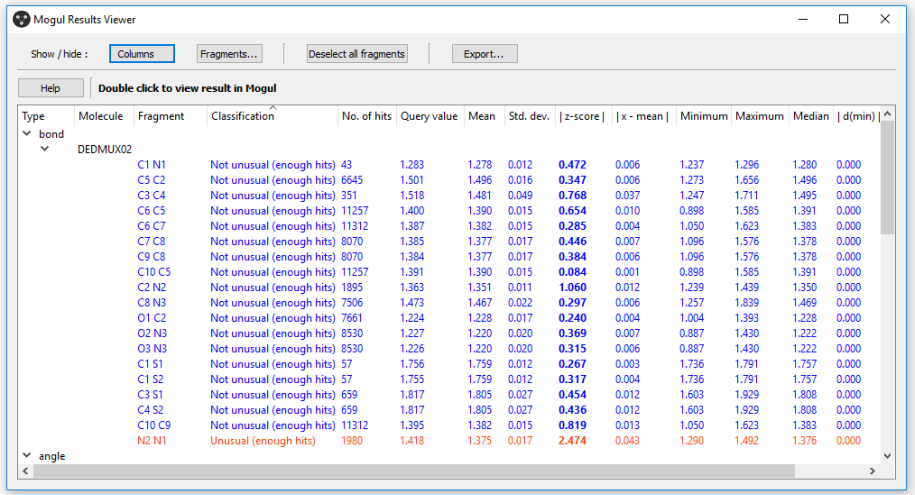
15. The results are color-coded. Unusual values are flagged in red. You can see that the N2-N1 bond and C1=N1-N2 angle are flagged in red. It is worth noting that the unusual bond and angle contain the N1 and N2 atoms which are involve in forming a more unlikely H-bond interactions, as observed in the HBP results.

16. Scroll through the results until you find the bond for N1-N2. Double-click this line to bring up the data from the Mogul library.

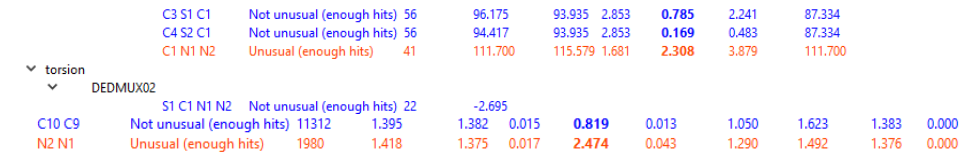
17. The red line marks the value of the bond distance from your molecule (the query). The histogram shows the data from the CSD, color coded by update. (Note, you can double-click the color swatches to change what color is shown.). To see which structures contribute to a certain bar on the histogram click **Deselect.**

18. Click the bar directly under the red query line. This will highlight that particular bar of the histogram.

14

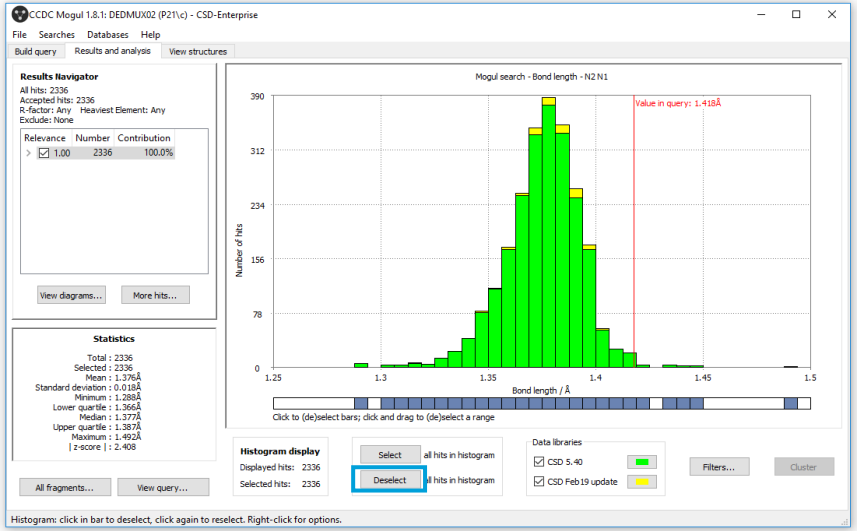


15

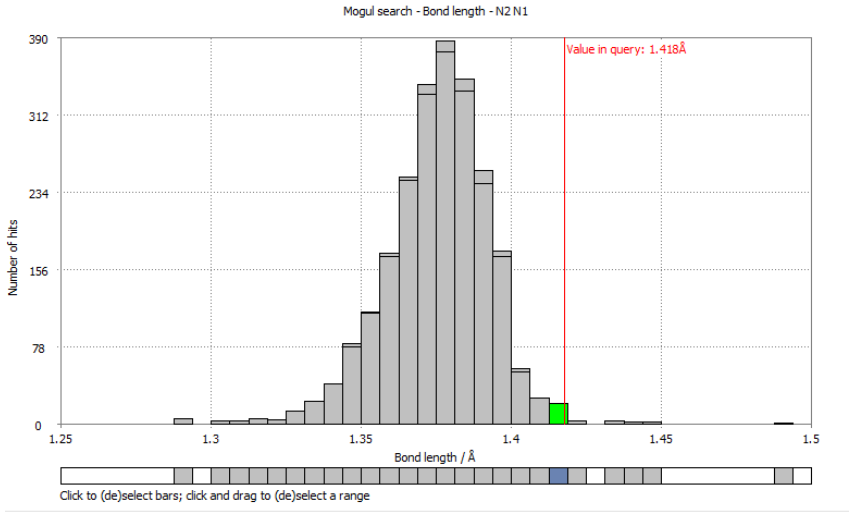


16

17



18

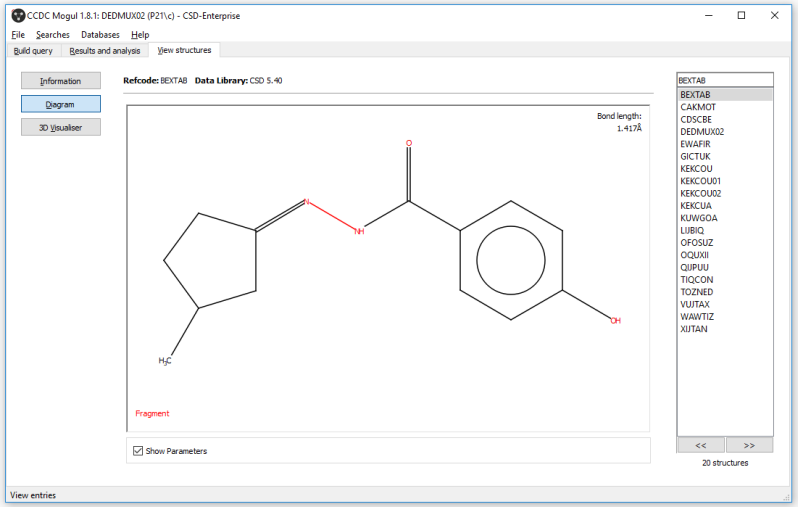


19. Now click the *View Structures* tab, near the top of the window to see a list of refcodes included in this bin. The default view for this window is the 2D diagram. Note that only 20 structures are present in CSD with this particular N1-N2 bond distance. Scroll through the refcodes on the right side of the window to view different structures.
20. Click the **3D Visualizer** button to see a 3D rotatable view of the structure. The fragment of the molecule used for comparison will be highlighted with the value displayed in green.
21. In the **Structure Navigator** window, type the refcode *DEDMUX*, to bring up the structure of Form I of N'-(1,3-dithiolan-2-ylidene)-4-nitrobenzohydrazide. Launch *Mogul Search Settings* and start the search as explained in steps 10 to 14.

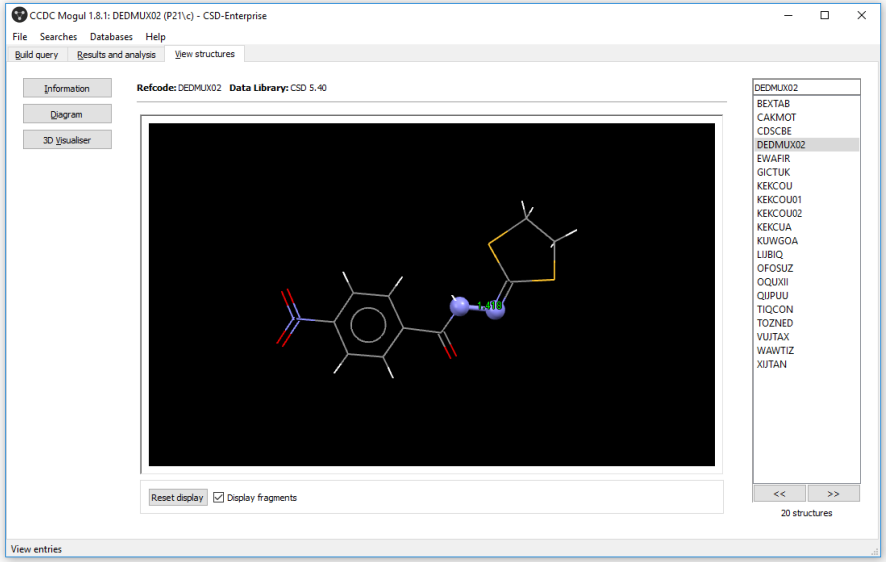
22. Scroll through the results and note that for Form I there are no red flagged bonds or angles.

In conclusion here, Mogul confirms that based on relevant structural data in the CSD, the geometry in Form I is found to be statistically usual, while Form III exhibits a conformation that is unusual. This assessment that Form I is more optimal agrees with the HBP analysis findings, the H-bonding network found in Form I is more likely based on CSD data compared with that in Form III.

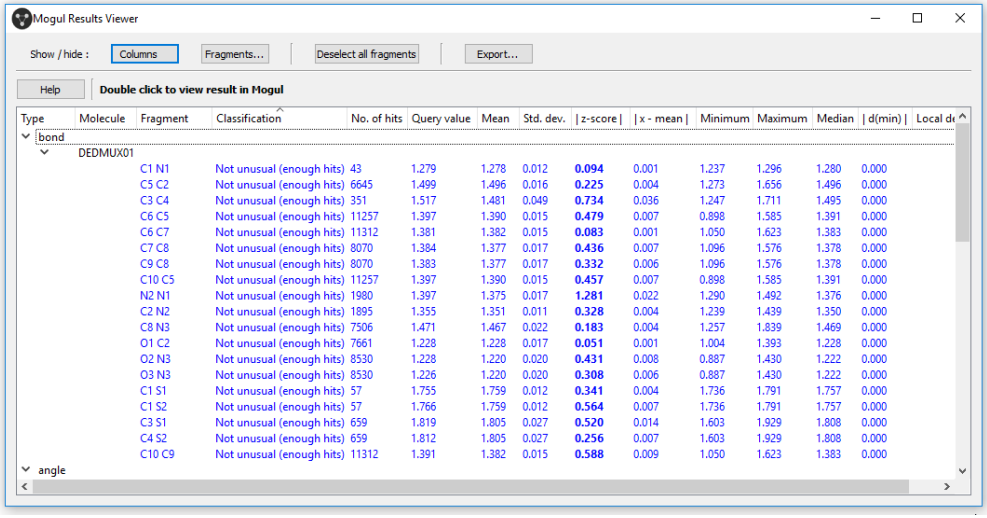
20



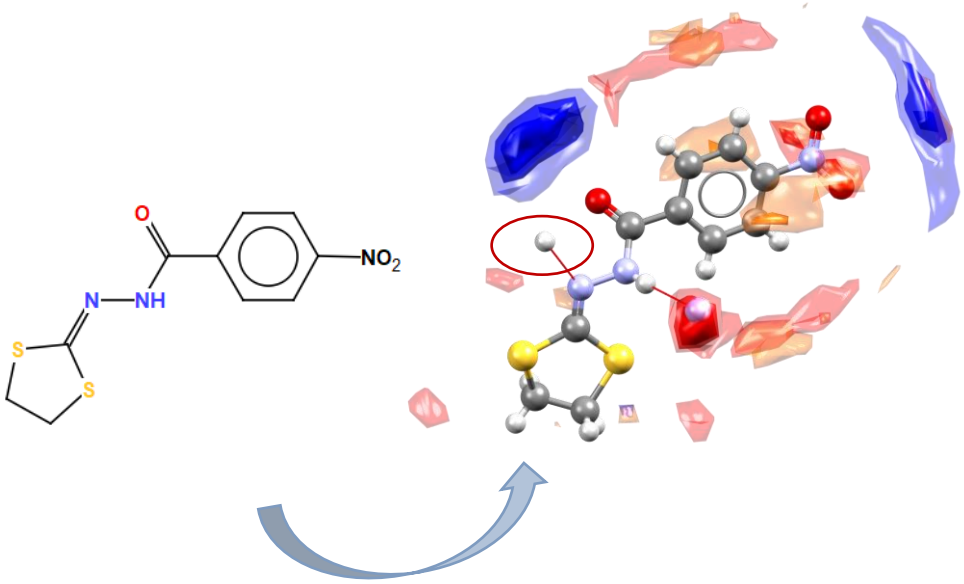
21



22



Using Full Interaction Maps to assess intermolecular interaction geometry



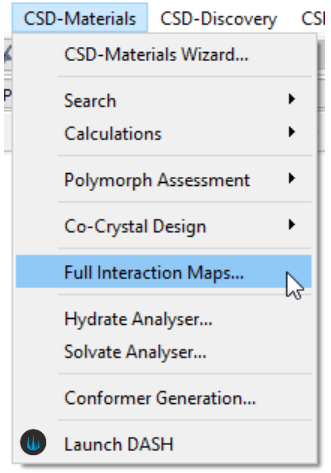
The stability of a given crystal structure is a balance between the intramolecular conformation and the intermolecular packing of the molecules in the crystalline state. One method for understanding the relative stability of crystal structures is to compare the observed intermolecular interactions with preferred geometries for that type of interaction.

In this example, you will see how FIMs corelates with HBP findings for N'-(1,3-dithiolan-2-ylidene)-4-nitrobenzohydrazide polymorphs. **How do the interactions in each polymorph compare with what is expected and observed from HBP?** You will learn how to produce Full Interaction Maps for a given structure and how to interpret these maps.

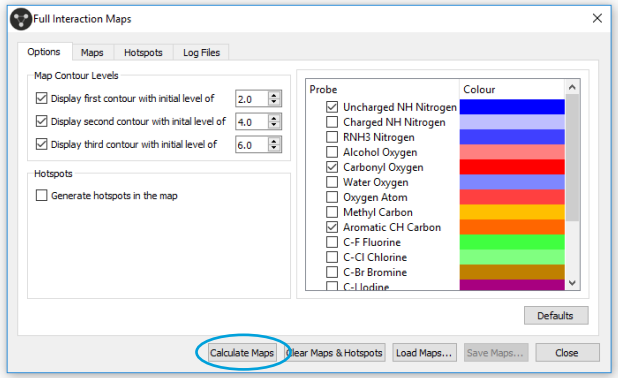
---

23. Close all the Molecular Geometry Check related dialogues. With *DEDMUX* (Form I) loaded in Mercury, click on CSD-Materials menu and then select *Full Interaction Maps…* from the dropdown menu.
24. In the *Full Interaction Maps* dialogue box, you will see several options. On the left you will find options to change the display contour levels. On the right, you will see a list of functional groups to be used as probes. For the purposes of this tutorial, we will keep the default options. These typically work well for most situations, but if you know you are looking for a specific functional group, or if you want to change the look of the map, you will want to change these settings. Click the **Calculate Maps** button to start.
25. The generated map will now be displayed in the main Mercury window. Notice the three different colors in the map. Red regions of the map denote areas in which there is a high probability of locating a hydrogen bond acceptor. Blue regions denote hydrogen bond donors, and brown regions indicate hydrophobic preferences
26. Now we want to see how the overall packing of this polymorph fits with the Full Interaction Map we have generated. Tick the box for H-bond in the Display Options toolbar.

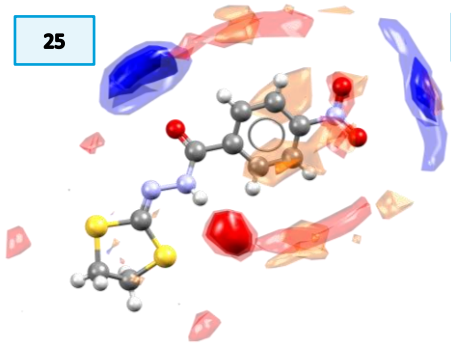
23



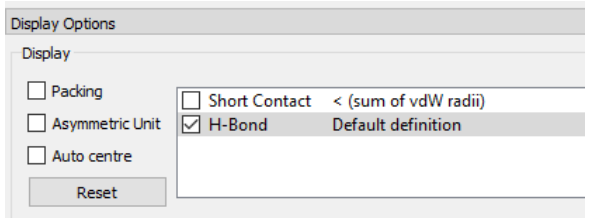
24



25



26



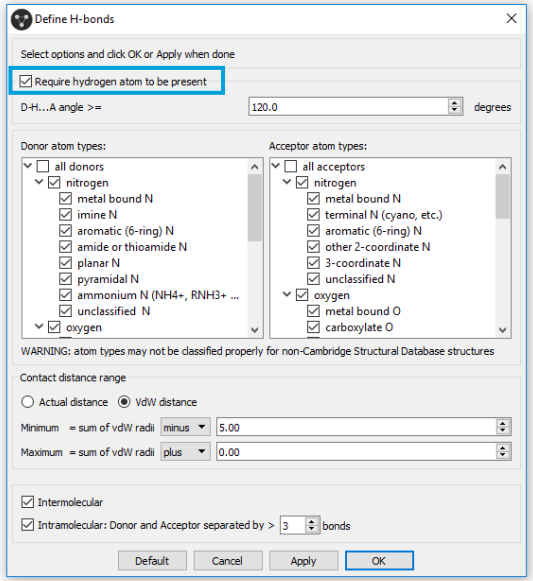
27. Use a D-H…A angle of 120˚ to define the hydrogen bond criteria. To ensure this is the case double click the H-bond line to launch the *Define H-bonds* dialogue box. Then tick the box for "Require hydrogen atoms to be present". Click **OK** to apply the change.

28. Now you will see dashed red lines in the Mercury window that indicate where hydrogen bonding interactions/contacts are present.

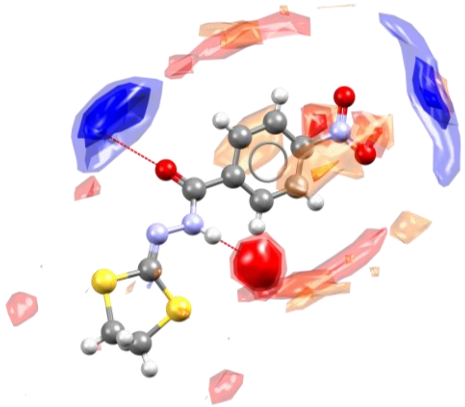
29. Click on these contacts to generate nearby molecules. You will see that in each case, the interaction falls within the contour range for the expected type. This indicates that the packing of Form I satisfies the expected interaction landscape of this conformation of sulfathiazole.

30. Now let's look at the Form II polymorph. In the Structure Navigator toolbar, type *DEDMUX02.* Click the **Reset** button in the Display Options toolbar to remove all the hydrogen bonding interactions.

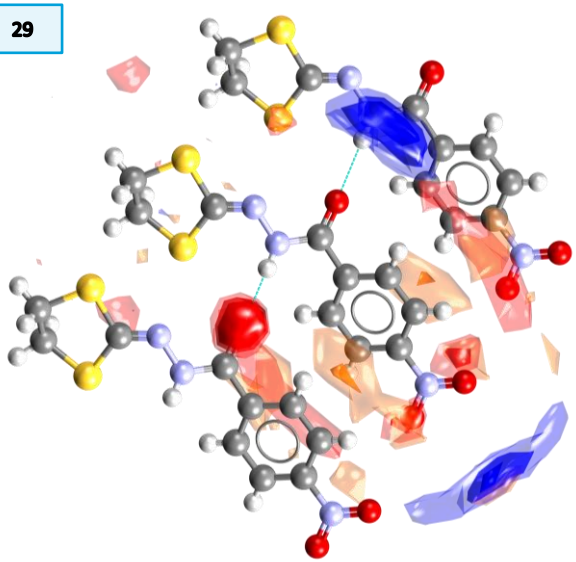
27



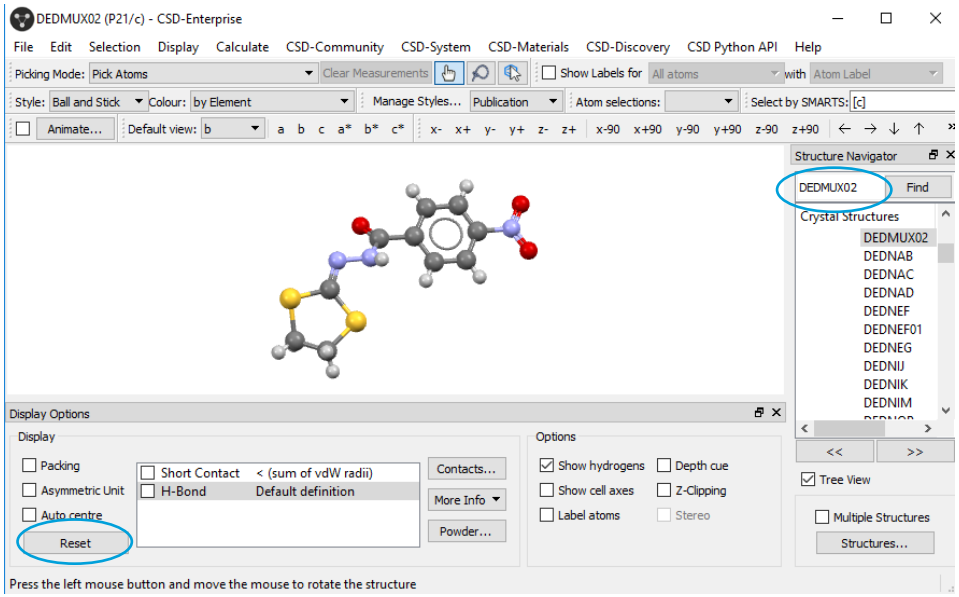
28



29



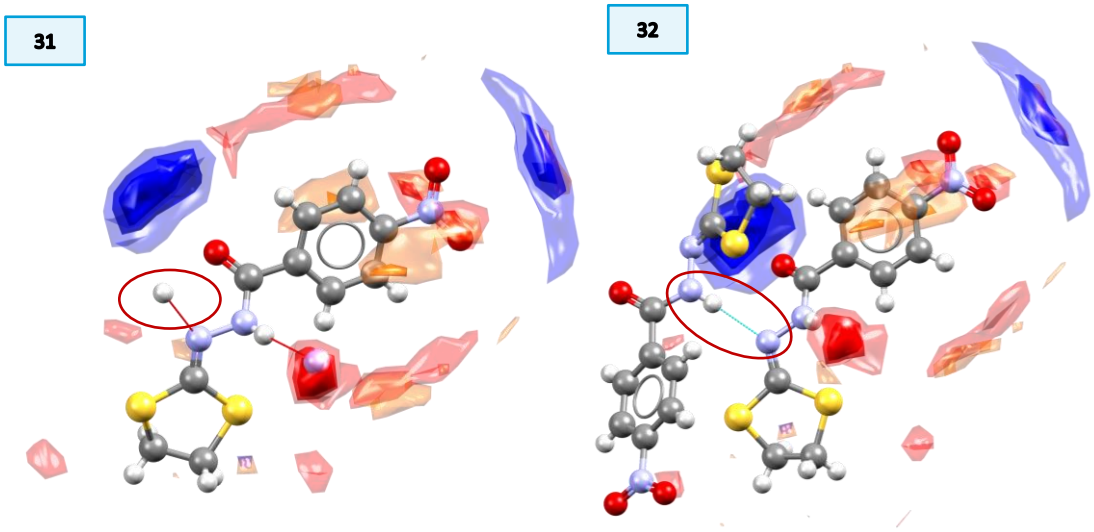
30



31. In the *Full Interaction Maps* dialogue box click **Calculate Maps** again (see step 24). You should now have a Full Interaction Map surrounding the molecule. Following steps 26-28 above, turn on the hydrogen bonding interactions

32. Click to expand the interaction around N1. Notice that one of the three interactions falls well outside the predicted region for a hydrogen bond acceptor. This suggests that this interaction has a non-ideal geometry and is likely to be significantly less stabilizing that the interactions in Form I.

In conclusion, the observed polymorphs of N'-(1,3-dithiolan-2-ylidene)-4-nitrobenzohydrazide exhibit different H-bonding interactions as well as noticeably different molecular geometry. We can use knowledge-based approaches to compare the observed intermolecular interactions in two polymorphs with the preferred geometries for these interaction types. Full Interaction Maps indicate that Form I has interactions which are near to ideal, whereas Form III has non-ideal interactions. This agrees with the HBP and molecular geometry assessment findings.



Next Steps

The last part of this workshop showed complementary methods to HBP to perform a polymorph risk assessment analysis. To learn more about these methods, you can try the Mogul workshop (available in the CSD-Core workshop area on our website) and the Full Interaction Maps workshop (available in the CSD-Materials workshop area on our website).

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

Feedback

We hope this workshop improved your understanding of the *Solvate Analyser* and you found it useful for your work. As we aim at continuously improving our training materials, we would love to hear your feedback. Click on this link to a survey (link also available from workshops webpage), 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 MAT-001. Thank you!