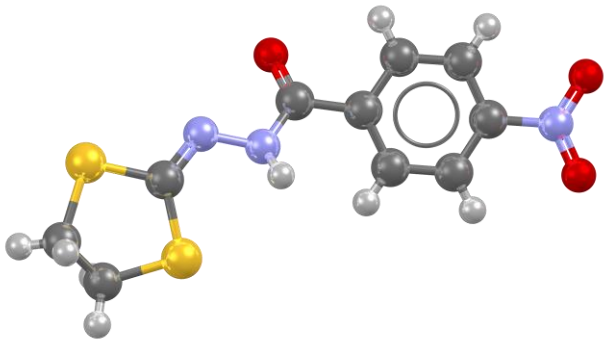
Investigating Solid Form Stability: Understanding Hydrogen Bond Propensities (MAT-007)

2023.2 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 has 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. There are several tools in the CSD-Portfolio suite that can be used in a complementary manner to explore these landscapes and gain insights into the relative stability of such polymorphs.

The Hydrogen Bond Propensity (HBP) tool in Mercury can be used to evaluate the relative likelihoods of possible hydrogen bonding networks. Mogul Geometry Check provides the user with the ability to assess molecular conformation, and the Full Interaction Maps tool can be used to assess intermolecular interactions 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.

Learning Outcomes

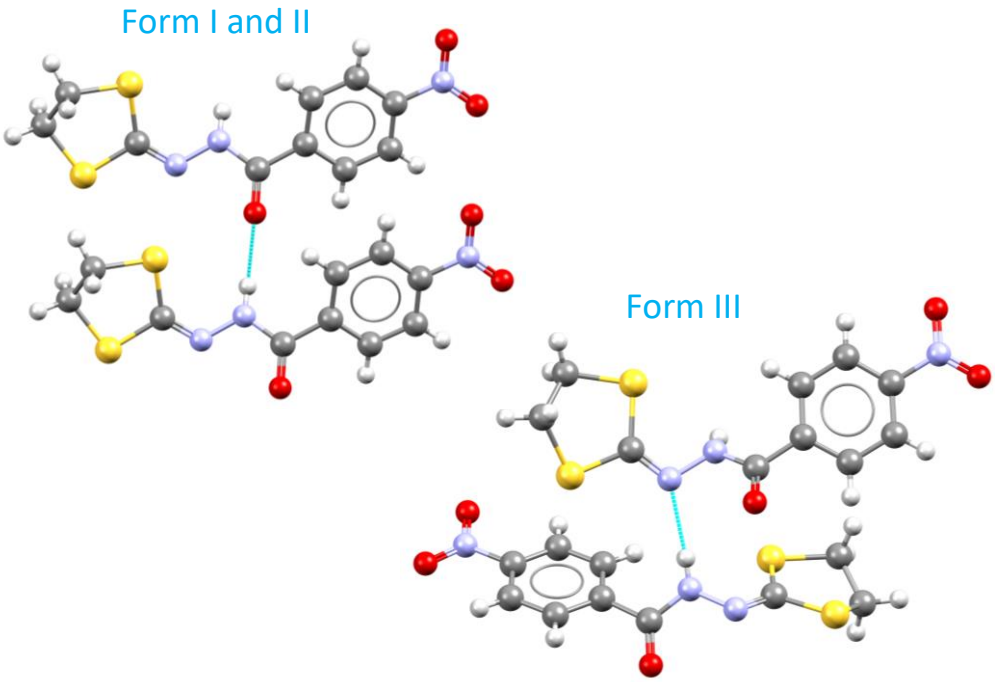
At the end of this workshop, you will:

- Be familiar with the Hydrogen Bond Propensities (HBP) tool.
- Know how to perform a HBP analysis and how to read and interpret the results.
- Understand how HBP can be used in polymorphs analysis.
- Be able to use Mogul Geometry Check and Full Interaction Maps as complementary approaches to assess solid forms.

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

Pre-required Skills

Familiarity with the Mercury interface is important; you can access the Visualization in Mercury self-guided workshop here (https://www.ccdc.cam.ac.uk/Community/educationalresources/workshop-materials/csd-community-workshops/).

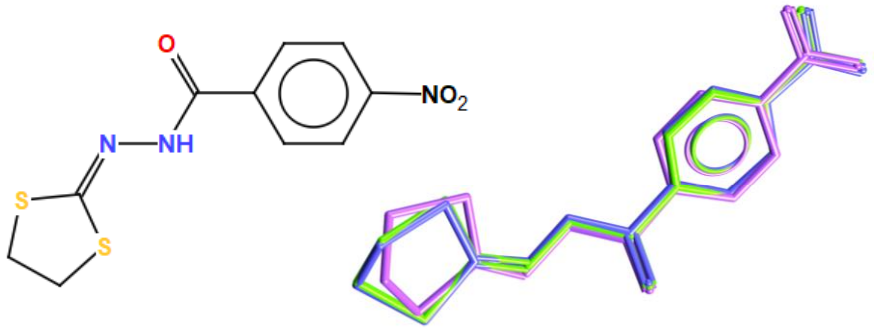


Materials

There are no additional materials required for this workshop.

Example 1. 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 *Hydrogen Bond Propensities* (HBP) tool to assess the relative likelihoods of the H-bond networks observed in the polymorphs.



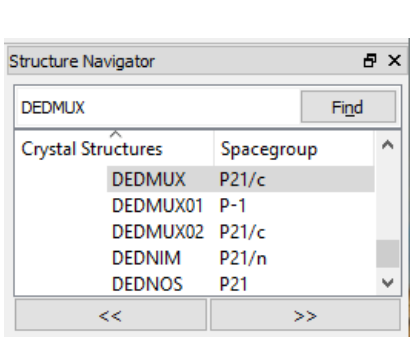
*N'-(1,3-dithiolan-2-ylidene)-4-nitrobenzohydrazide (refcode family **DEDMUX**). Form III (purple) has a different geometry.*

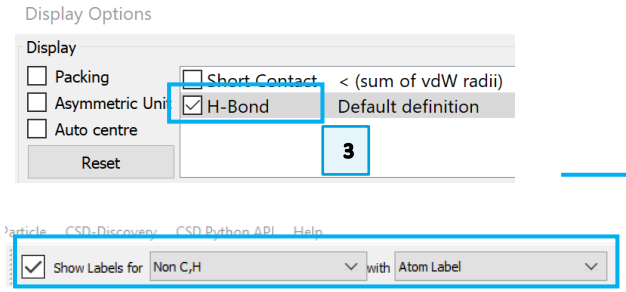
Examine the H-bonding network.

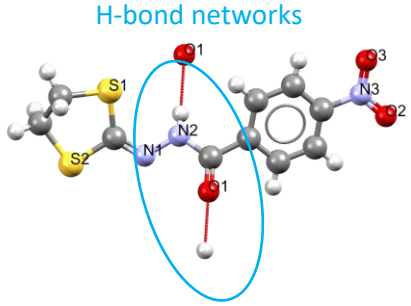
1. Start Mercury by double-clicking the icon on your Desktop or navigating from the Start Menu (*Start > CCDC > Mercury*).

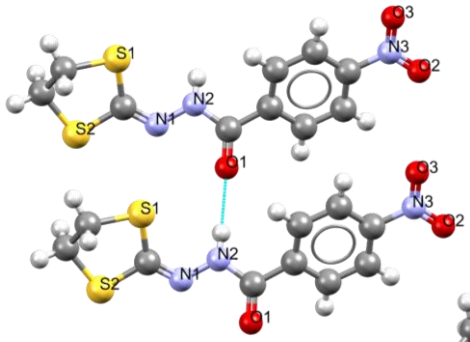
2. In the **Structure Navigator** window, type the refcode "DEDMUX", to bring up the structure of the first polymorph.

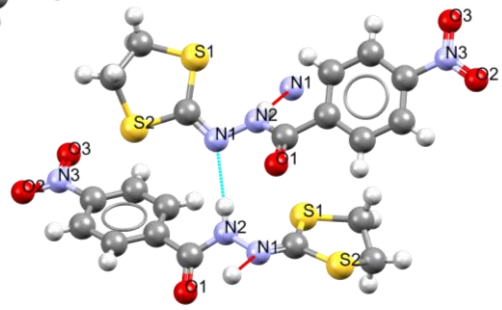
3. Ensure that the **H-Bond** check box in the *Display Options* area of the Mercury interface is toggled on and expand the contacts for Form I by clicking the red lines or atoms of the hanging contacts. Check the **Show Labels for** box from the toolbar and select *Non-C,H* from the dropdown menu. Note that the NH of the amide group (N2) acts as donor and the O atom of the carbonyl group (O1) 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 amide NH (N2) and the imine N (N1). Check all the possible *donors* and *acceptors*. How many are there?











Calculate H-bond propensity.

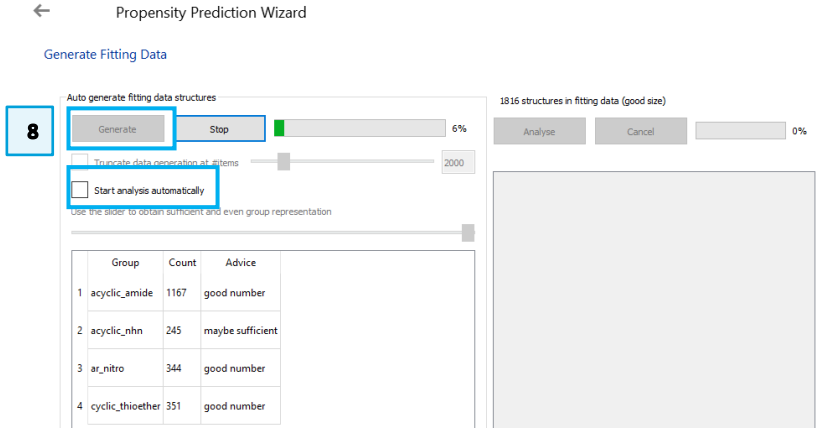
4. Ensure that "DEDMUX02" is selected in the **Structure Navigator** and press Reset in the **Display Options**.

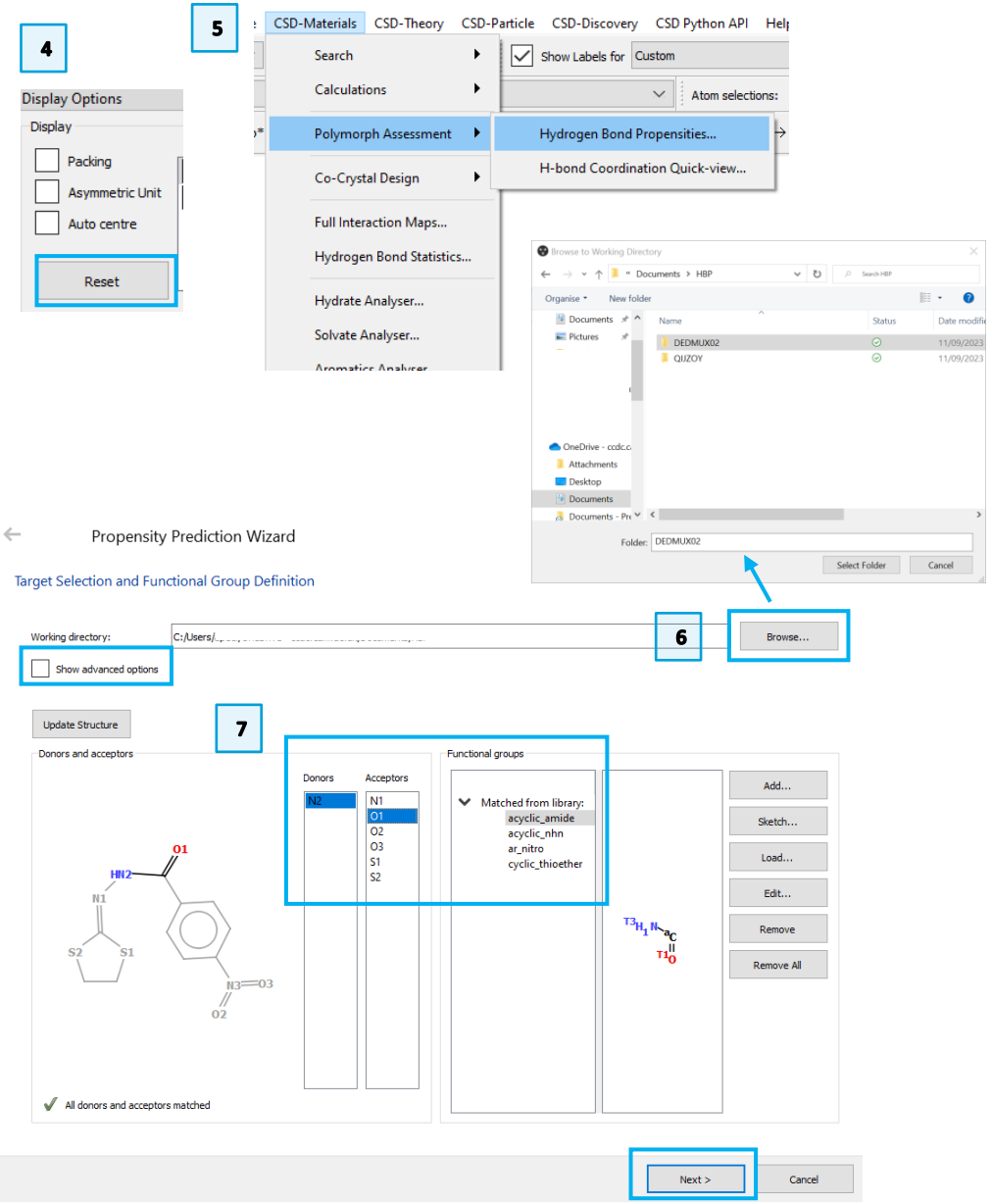
5. From the top-level menu select *CSD-Materials* > *Polymorph Assessment* > *Hydrogen Bond Propensities…*

6. 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. One donor has been identified: N2 as acyclic_amide. Six acceptors have also been identified: N1 as acyclic_nhn, O1 as acyclic_amide, O2 and O3 as ar_nitro, and S1 and S2 as cyclic_thioether.

7. 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 *Matched 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**.

8. 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.

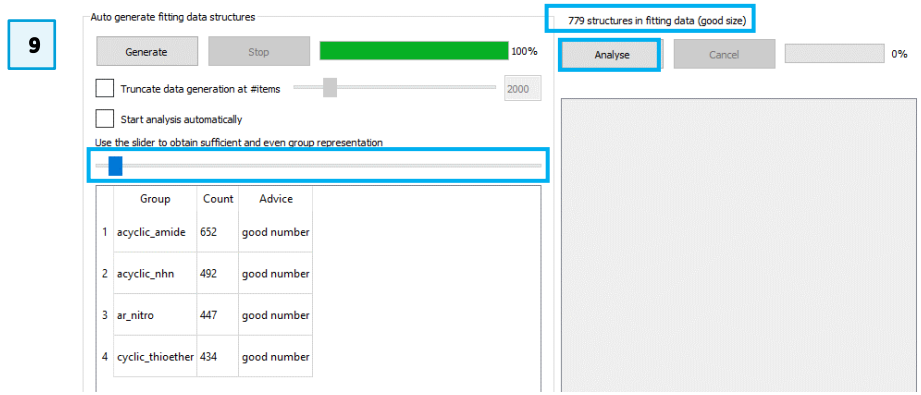




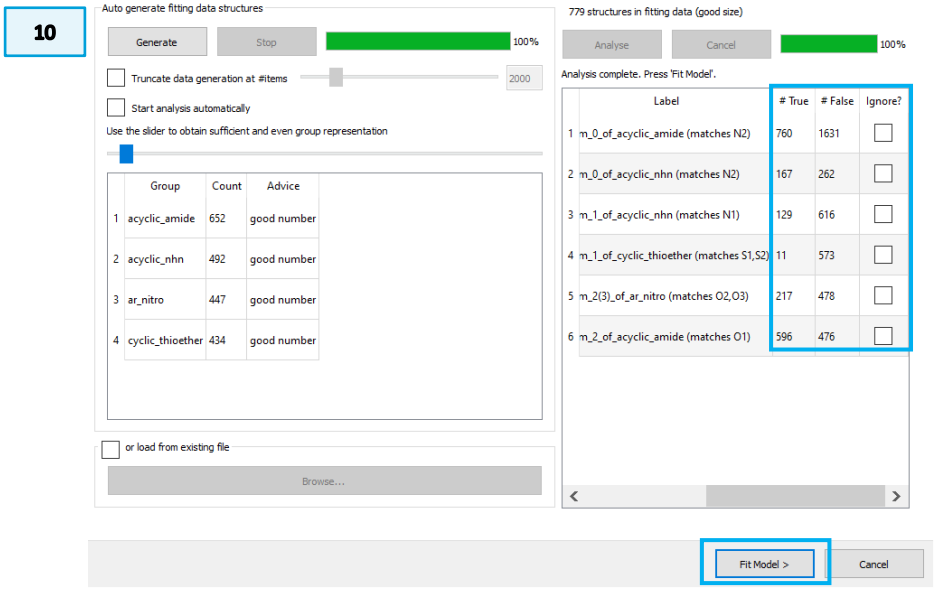
Tips and tricks

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…**

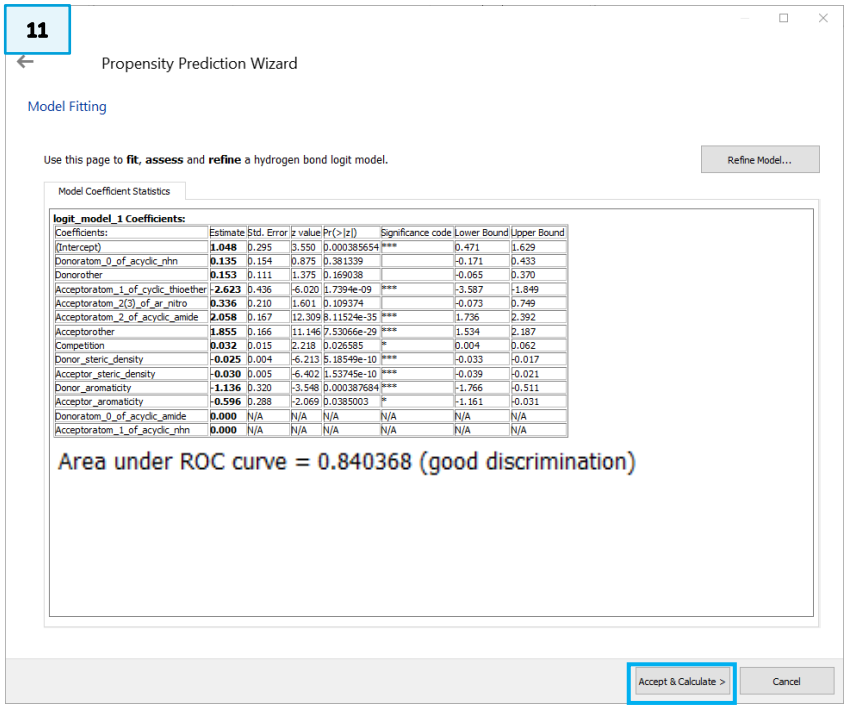
9. When the run is finished, it attempts to automatically select a sufficient number of hits (count) per functional group with fairly even representation across the groups. In general, around 300-500 structures per functional group should be enough. The group numbers can be adjusted by using the slider highlighted in blue. This allows you to remove or add structures until a more even set of data, or more appropriate number of groups, is obtained. We will leave the default settings for this example and then click **Analyse**.



10. 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, you should check that they are ticked in the **Ignore?** checkboxes. There are no very low values in this example. Click the **Fit Model >** button to continue.



11. For this example, the Area under *ROC* (Receiver Operating Characteristic) curve should be around 0.84. To achieve a good H-bond propensity calculation you should always aim for a value of around 0.75 or above. Click **Accept & Calculate** to continue.

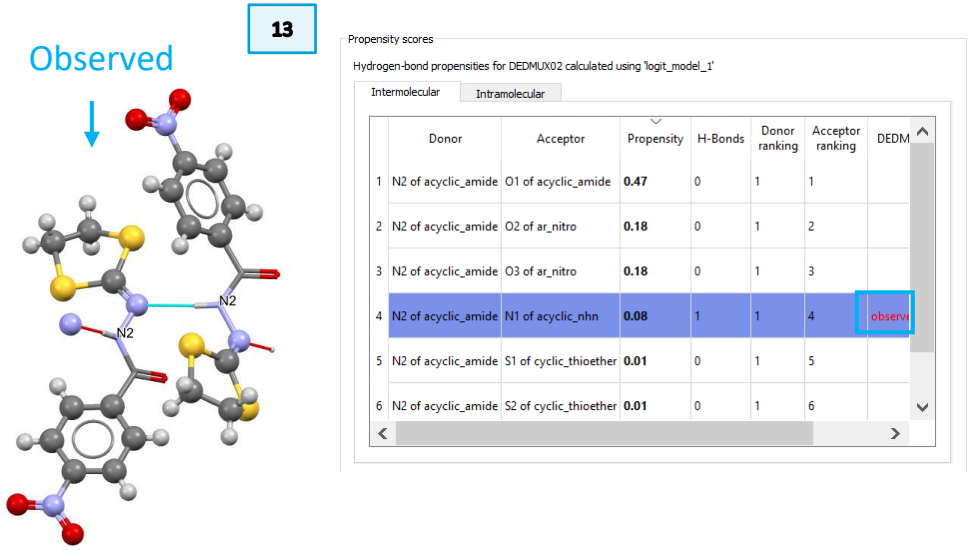


Summary of HBP results

We have now obtained the results of our HBP calculation, and we can analyse them in the graph and tables displayed. Three different layouts are available in the *Layout* section and can be selected by pressing the radio buttons; layout 1 is more convenient for viewing the graph, whereas layouts 2 and 3 are preferable for viewing the tables.

12. *The Chart:*
    - Plots *Mean H-bond Propensity* *vs* the *Mean H-Bond Co-ordination*.
    - The target structure is represented as a black circle with a white interior.
    - To zoom, use the magnifying glass icon in the lower left-hand corner of the wizard, left click and drag on the area to zoom in on it. To go back to the default view, press **Reset**.
    - To filter the chart for a given number of H-Bond pairs, use the colour legend.
    - The most likely H-bonding network is displayed towards the lower-right corner, the outcome should be read along the diagonal.
    - DEDMUX02 does not have the most likely H-bonding network.
    - Click on the points to highlight the network in the *Propensity score* table.
    - Hover over a point to display the mean propensity and mean co-ordination values.

13. *Propensity Scores Table:*
    - Select **Layout 2** or **Layout 3** to see the full Propensity Scores table. The table can be expanded horizontally if needed by dragging the double-headed arrow that appears when hovering over the border between the propensity and co-ordination tables.
    - The most likely H-bond pair will score the highest propensity.
    - The H-bonds present in the targeted structure are marked as observed.
    - The table is interactive, clicking on **observed**, which is located at the far right-hand side of the table will highlight the donor and acceptor group in the 3D visualizer, while 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 the molecule, with N2-H9···O1 giving the highest propensity. You can see this interaction is *not* observed in the DEDMUX02 structure.



14. *Co-ordination Scores Table:*
    - (d) stands for donor and (a) for acceptor.
    - =0, = 1, = 2 etc. denote 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 is red that indicates the outcome is sub-optimal.
    - For DEDMUX02, for the observed H-bond the donor N2 of the acyclic_amide (d) has optimal co-ordination (donates once, = 1). The acceptor N1 of acyclic_nhn (a) is sub-optimal as it accepts once (= 1) while there is a slightly higher likelihood for it not to accept (= 0).

15. To summarise our observations, Form III (DEDMUX02) is represented as a white 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.

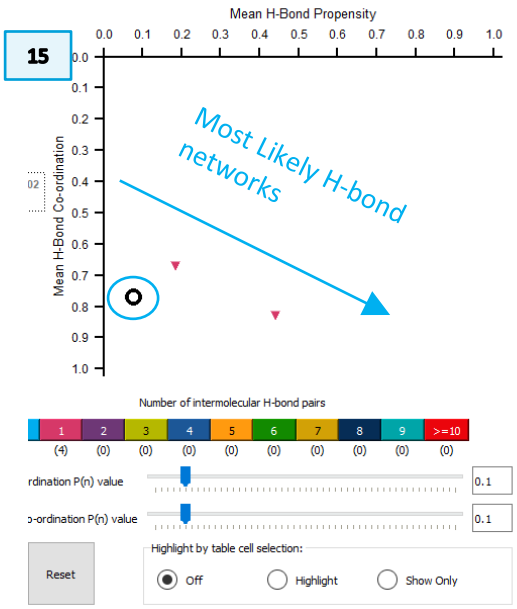
14

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

15



Propensity scores

Hydrogen-bond propensities for DEDMUX02 calculated using 'logit_model_1'

Intermolecular | Intramolecular

	Donor	Acceptor	Propensity	H-Bonds	Dono rankin
3	N2 of acyclic_amide	O3 of ar_nitro	**0.19**	0	1
4	N2 of acyclic_amide	N1 of acyclic_nhn	**0.08**	1	1
5	N2 of acyclic_amide	S1 of cyclic_thioether	**0.01**	0	1

Left-clicking table items visualizes H-bond groups/contacts and interacts with chart

Atom Highlig

Co-ordination scores

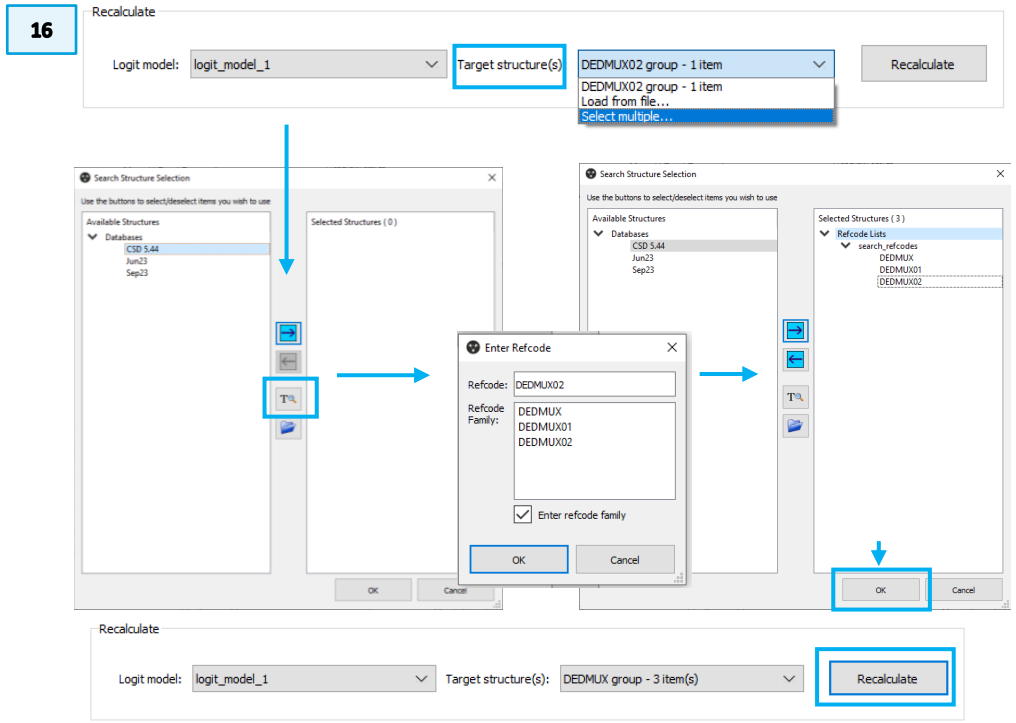
(To refresh table: left-click chart point)

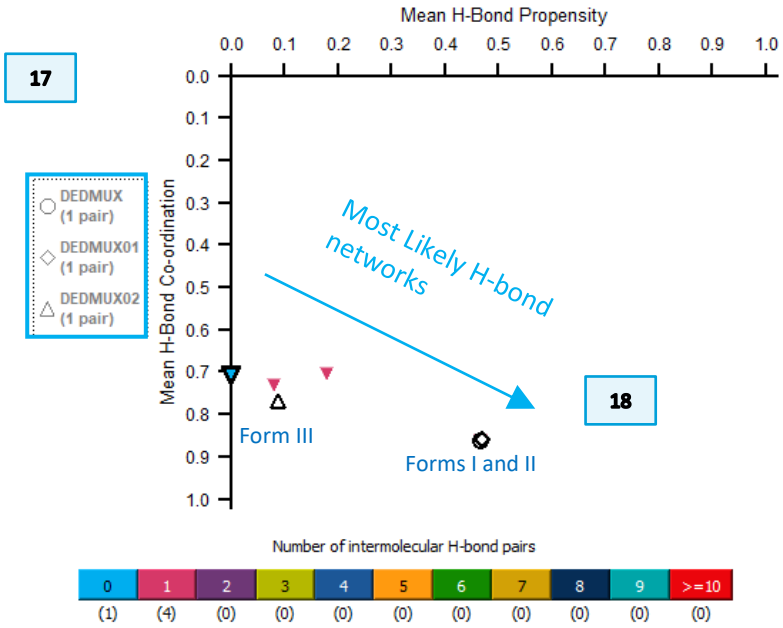
	Atom (D/A)	= 0	= 1	= 2	= 3
2	N1 of acyclic_nhn (a)	0.547	0.453	0.000	0.000
3	O1 of acyclic_amide (a)	0.319	0.660	0.018	0.002
4	O2 of nitro (a)	0.888	0.105	0.006	0.000

16. 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 Structures** pane. Click **OK**, then click **Recalculate**.

17. 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 (DEDMUX and DEDMUX01) have the same H-bond network (N-H···O) with the highest propensity and best coordination.

18. If we compare the Propensity scores of Forms I and III (DEDMUX and DEDMUX02), we can see that the highest propensity H-bond pair is used in Form I, while Form III alternatively uses a low propensity pair.





19. If we compare the Co-ordination scores of Forms I and III (DEDMUX and DEDMUX02), we can see that there are two sub-optimal acceptors for Form III. N1 donates once but slightly prefers to donate zero times, and O1 accepts zero times but prefers to accept once. In Form I the co-ordination scores for all donors and acceptors are optimal.

Conclusion

In conclusion, one of the polymorphs (Form III, CSD entry DEDMUX02) is observed to have a noticeably less likely H-bonding network than the other two experimentally observed polymorphs (Form I, CSD entry DEDMUX and Form II, CSD entry 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.

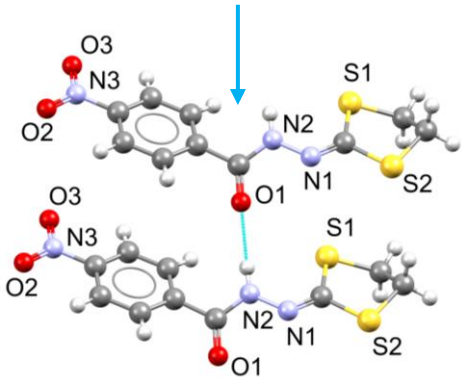
19

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)	0.105	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

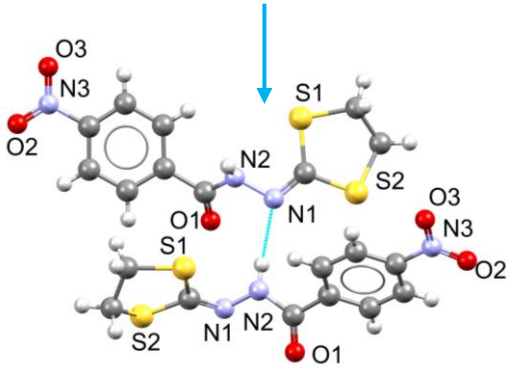


Form III

Co-ordination scores

(To refresh table: left-click chart point)

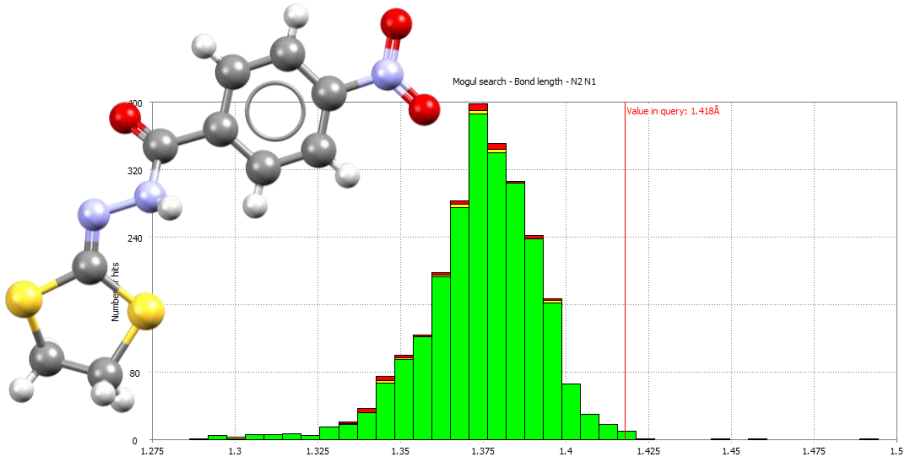
	Atom (D/A)	= 0	= 1	= 2	= 3
1	N2 of acyclic_amide (d)	0.081	0.889	0.030	0.000
2	N1 of acyclic_nhn (a)	0.547	0.453	0.000	0.000
3	O1 of acyclic_amide (a)	0.319	0.660	0.018	0.002
4	O2 of nitro (a)	0.888	0.105	0.006	0.000
5	O3 of nitro (a)	0.888	0.105	0.007	0.000
6	S1 of cyclic_thioether (a)	0.974	0.026	0.000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



Example 2 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 that a 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 correlate 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.



N'-(1,3-dithiolan-2-ylidene)-4-nitrobenzohydrazide (CSD entry DEDMUX02) and the distrubtion of N-N bond lengths identified by Mogul in similar fragments in the CSD.

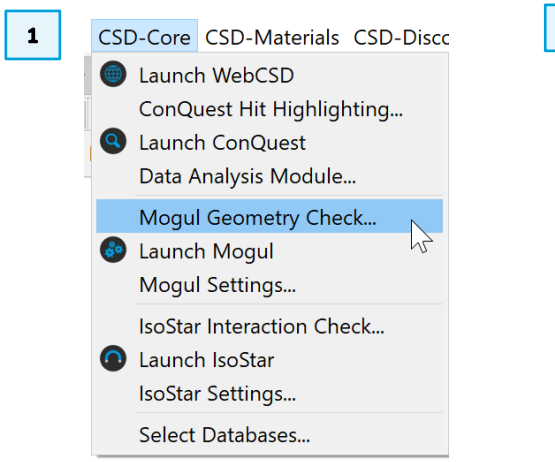
---

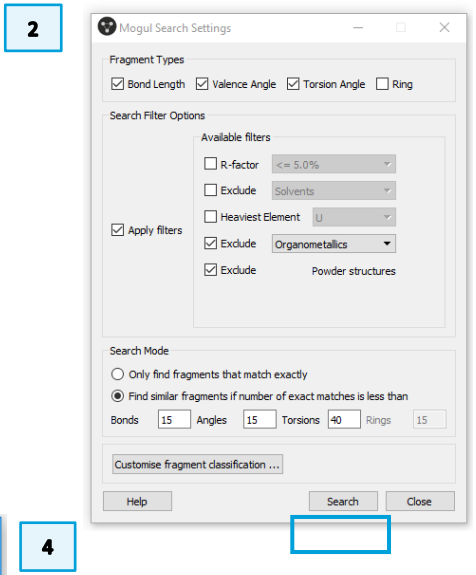
1. Close all HBP-related dialogue boxes and with DEDMUX02 (Form III) loaded in Mercury, click on CSD-Core menu and then select *Mogul Geometry Check* from the dropdown menu.

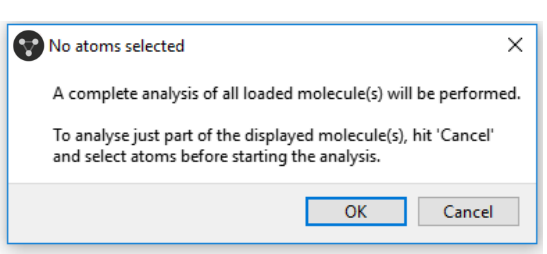
2. In the **Mogul Search Settings** window, you can typically use the default settings, but we can streamline our search by unticking the box for rings and ticking the boxes for **Apply Filters**, **Exclude Organometallics**, and **Exclude Powder** structures. Click **Search** to start.

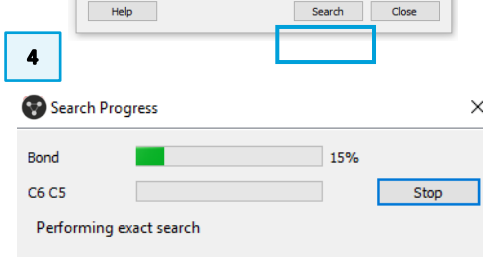
3. A dialogue box will pop up to warn you that you are going to check the entire molecule. Click **OK** to continue.

4. The search will begin to run. You can follow its progress in the *Search Progress* dialogue box.









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

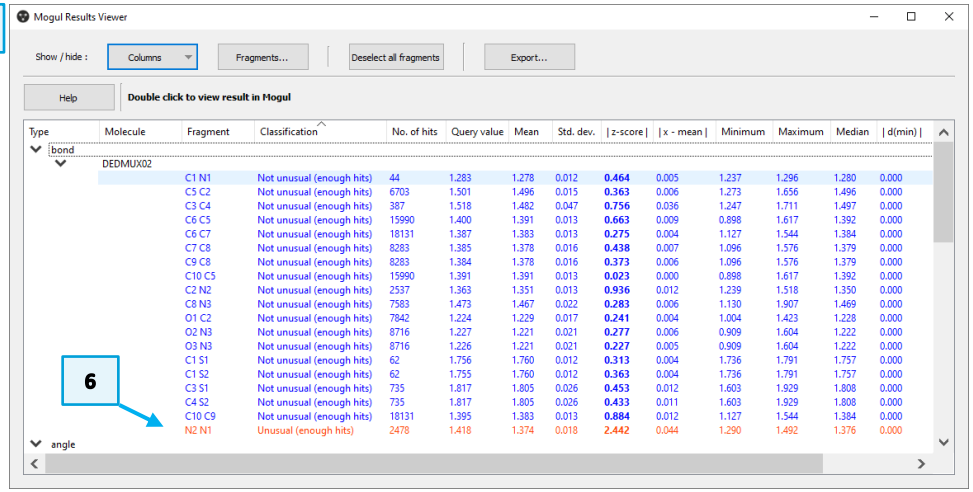
6. The results are color-coded. Unusual values are flagged in red. You can see that the N2-N1 bond and C1=N1-N2 angle are flagged in red. It is worth noting that the unusual bond and angle contain the N1 and N2 atoms which are involved in forming a more unlikely H-bond interactions, as observed in the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

7. 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.

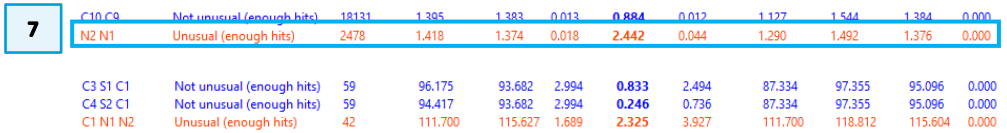
8. The red line marks the value of the bond distance from your molecule (the query). The histogram shows the data from the CSD, color coded by update. (Note, you can double-click the color swatches to change what color is shown.). To see which structures contribute to a certain bar on the histogram click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

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

5

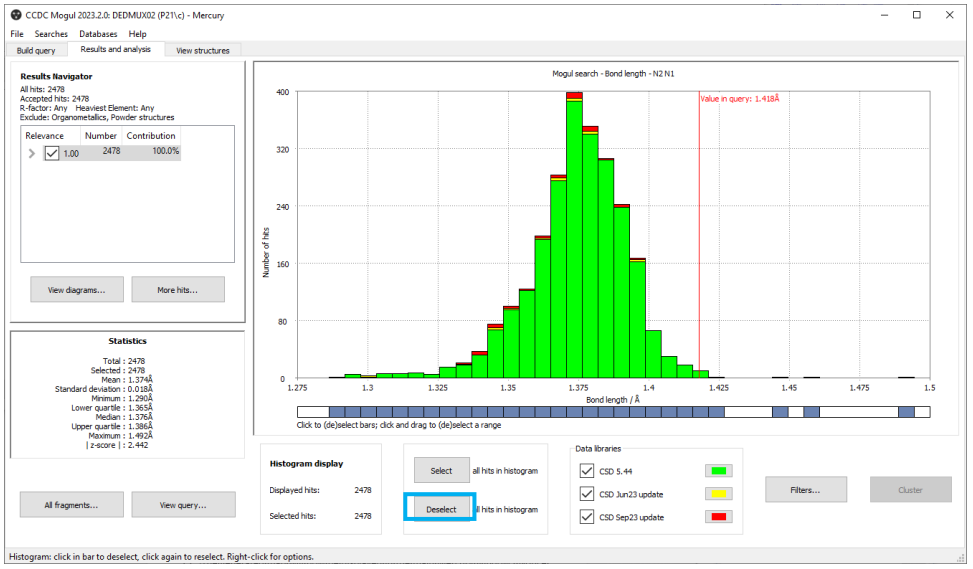


7

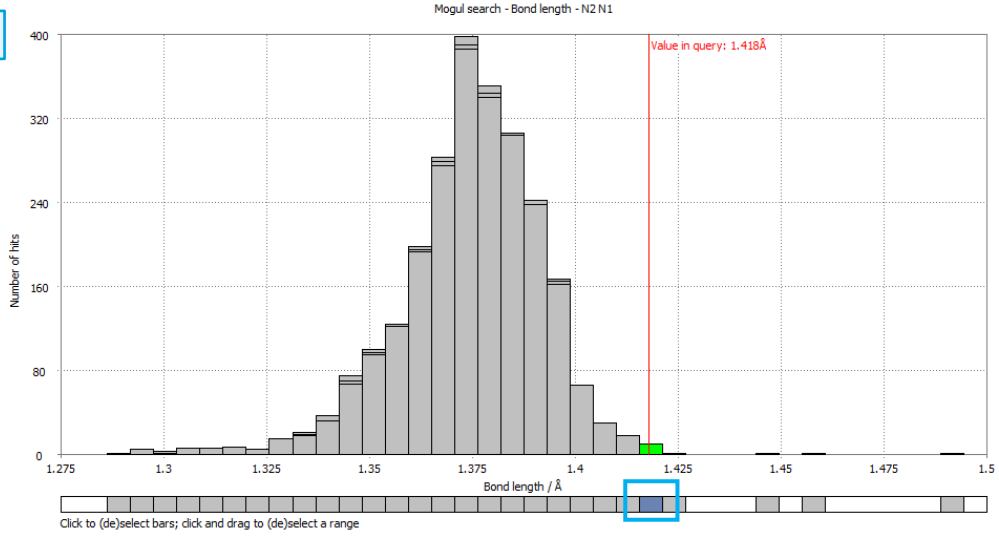


C10 C9	Not unusual (enough hits)	18131	1.395	1.383	0.013	**0.884**	0.012	1.127	1.544	1.384	0.000
N2 N1	Unusual (enough hits)	2478	1.418	1.374	0.018	**2.442**	0.044	1.290	1.492	1.376	0.000
C3 S1 C1	Not unusual (enough hits)	59	96.175	93.682	2.994	**0.833**	2.494	87.334	97.355	95.096	0.000
C4 S2 C1	Not unusual (enough hits)	59	94.417	93.682	2.994	**0.246**	0.736	87.334	97.355	95.096	0.000
C1 N1 N2	Unusual (enough hits)	42	111.700	115.627	1.689	**2.325**	3.927	111.700	118.812	115.604	0.000

8



9



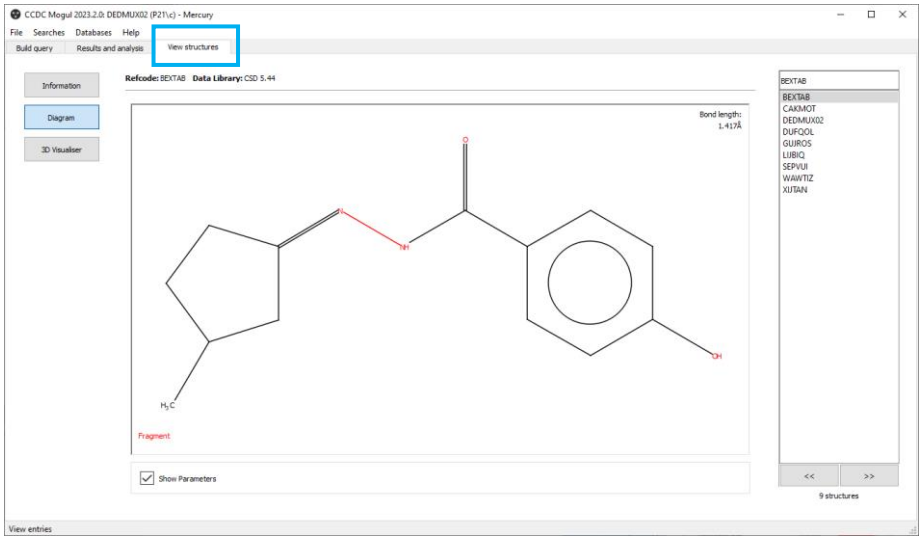
10. Now click the **View Structures** tab, near the top of the window to see a list of refcodes included in this bin. The default view for this window is the 2D diagram. Note that only 13 structures are present in CSD with this particular N1–N2 bond distance. Scroll through the refcodes on the right side of the window to view different structures.

11. 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.

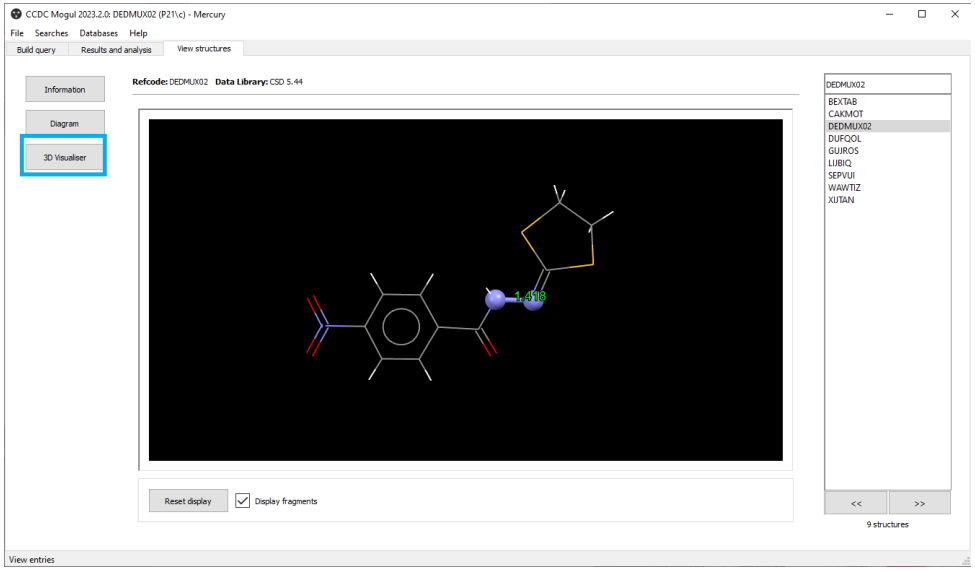
12. In the **Structure Navigator** window in Mercury, type the refcode "DEDMUX", to bring up the structure of Form I of N'-(1,3-dithiolan-2-ylidene)-4-nitrobenzohydrazide. Launch *Mogul Search Settings* and start the search as explained in **Steps 2–5**.

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

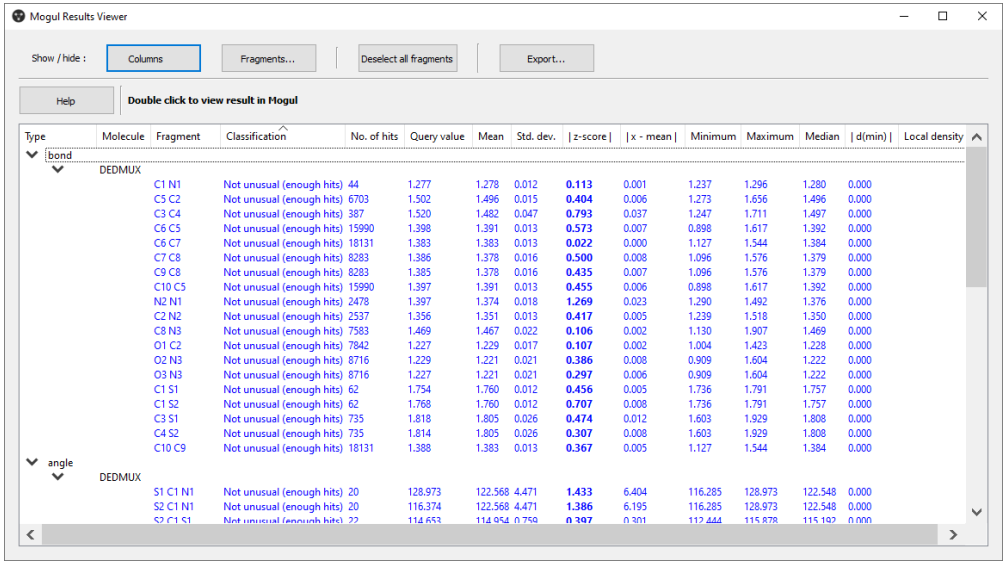
10



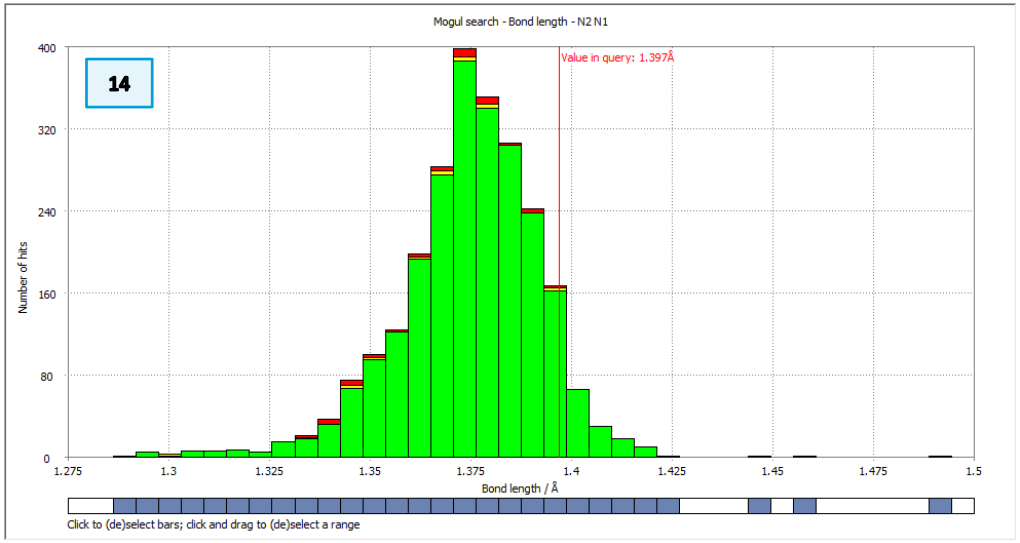
11

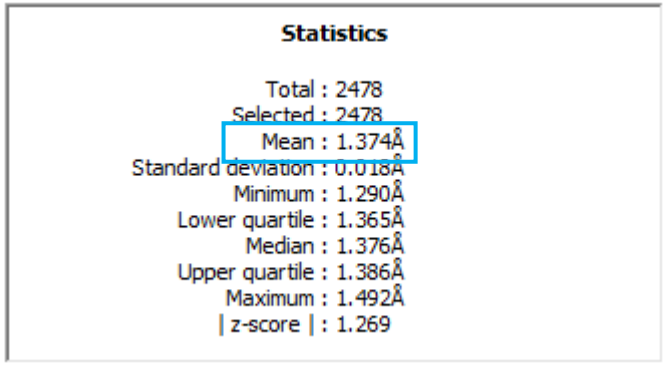


13



14. Click on **N2 N1** in the *bond* section of the **Mogul Results Viewer** to bring up the histogram of bond lengths. Notice how the query value in DEDMUX (1.397 Å), whilst still above the mean (1.374 Å)[1], is within a significantly more populated region of the histogram than DEDMUX02.





Conclusion

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.

Exercise

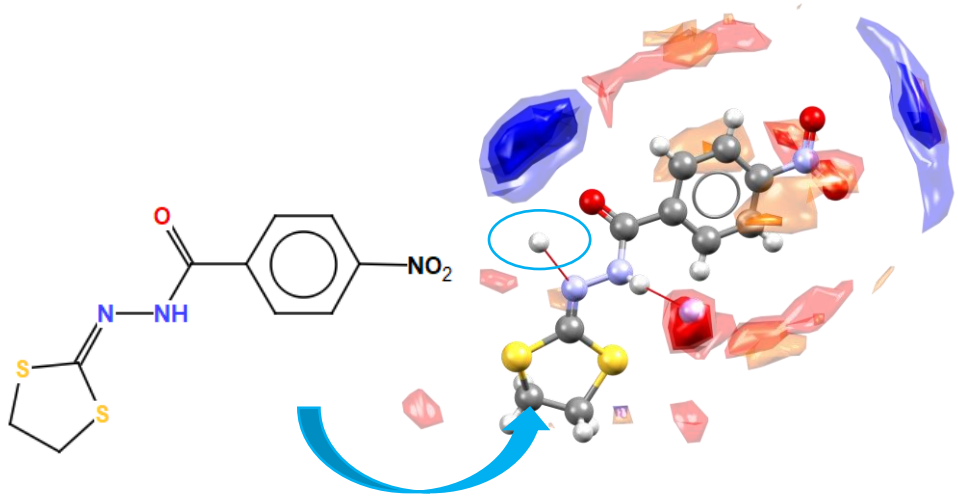
Explore the histograms and structures for the C1=N1-N2 angles in DEDMUX and DEDMUX02.

[1] Note that the mean value may change (slightly) depending on the version of the CSD you have installed.

Example 3. 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 *Full Interaction Maps* (FIMs) correlates 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.



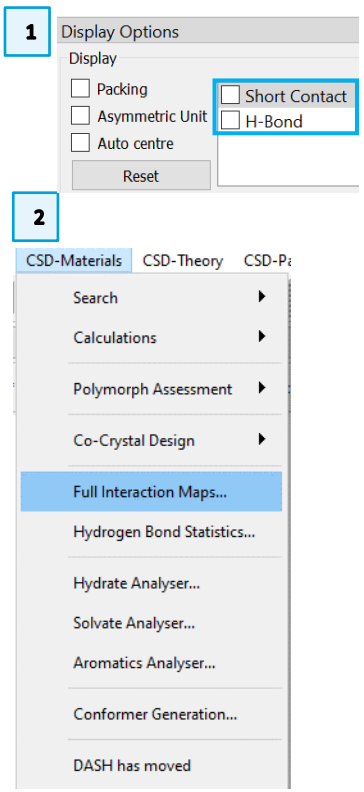
2D diagram of CSD entry DEDMUX and corresponding Full Interaction Map, with hydrogen bonds overlaid.

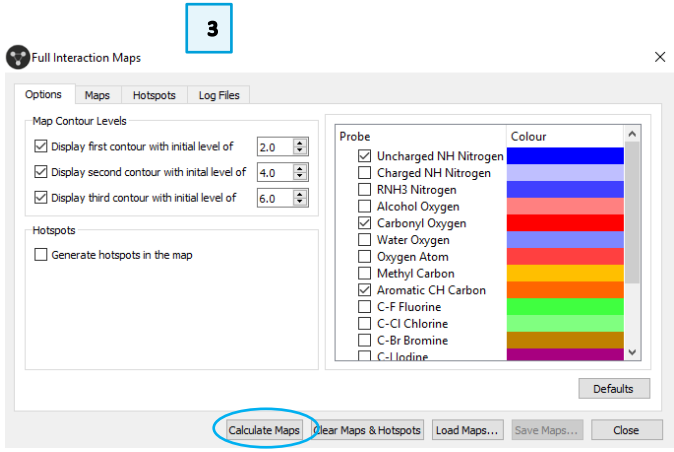
1. With DEDMUX (Form I) loaded in Mercury, check that the **H-Bond box** in the **Display Options** toolbar is not checked.

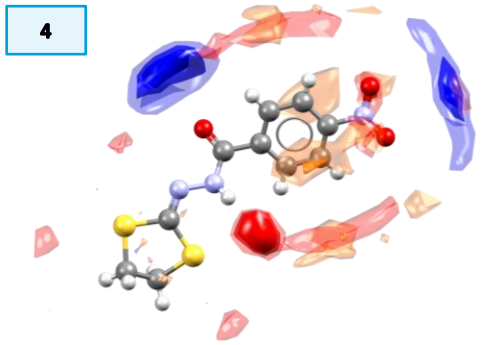
2. Click on *CSD-Materials* menu and then select *Full Interaction Maps…* from the dropdown menu.

3. In the **Full Interaction Maps** window, 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.

4. 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 areas in which there is a high probability of locating a hydrogen bond donor, and Orange regions indicate hydrophobic pockets.







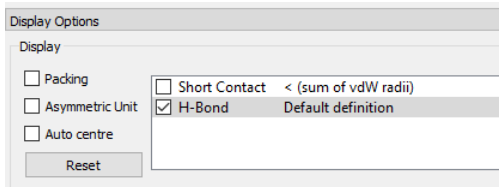
5. Now we want to see how the overall packing of this polymorph fits with the Full Interaction Map we have generated. Tick the box for **H-bond** in the **Display Options** toolbar then double click the H-bond line to open the **Define H-bonds** window

6. Use a *D-H…A angle* of 120˚ to define the hydrogen bond criteria. Then tick the box for **Require hydrogen atoms to be present**. Click **OK** to apply the change.

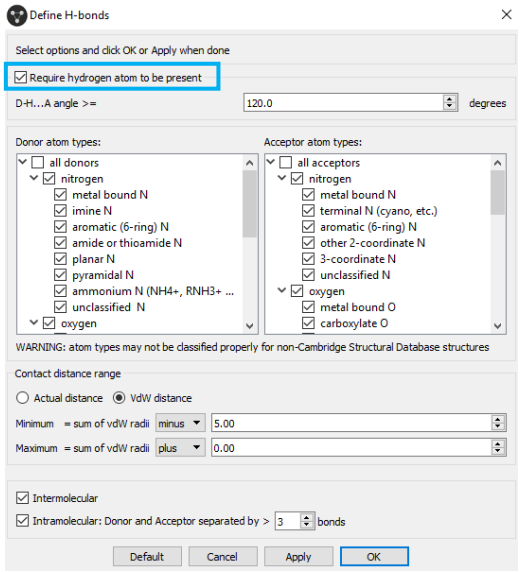
7. Now you will see dashed red lines in the Mercury window that indicate where hydrogen bonding interactions/contacts are present.

8. 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 N'-(1,3-dithiolan-2-ylidene)-4-nitrobenzohydrazide.

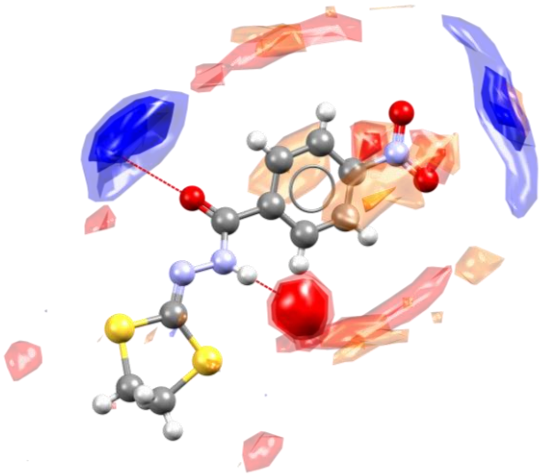
5



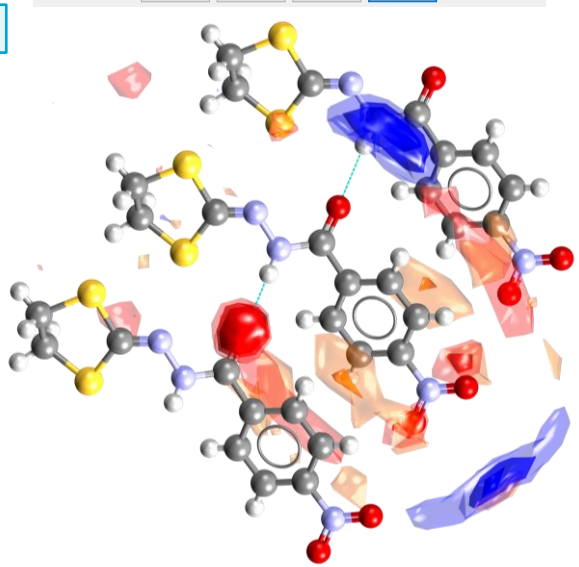
6



7



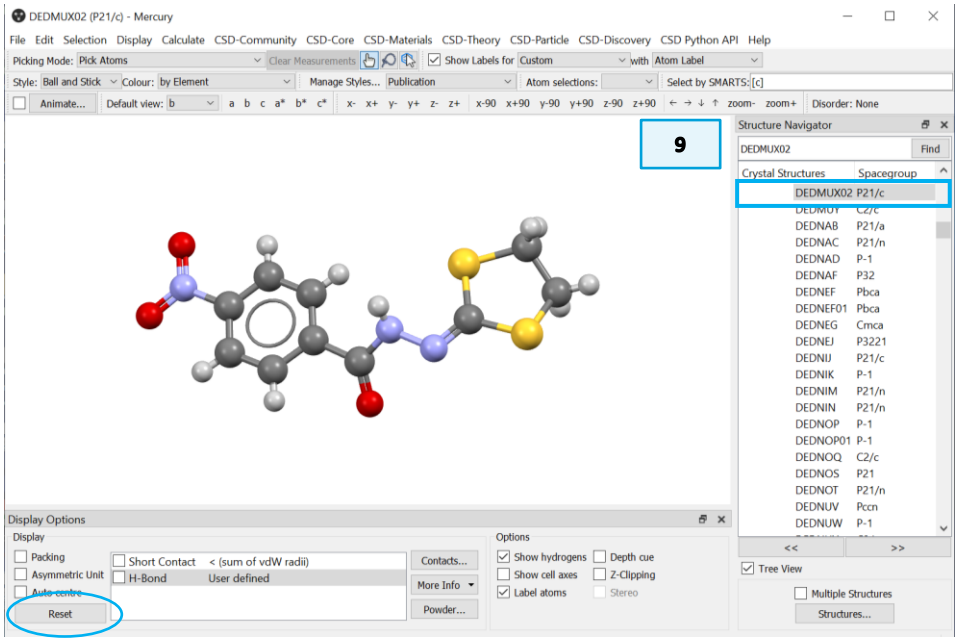
8

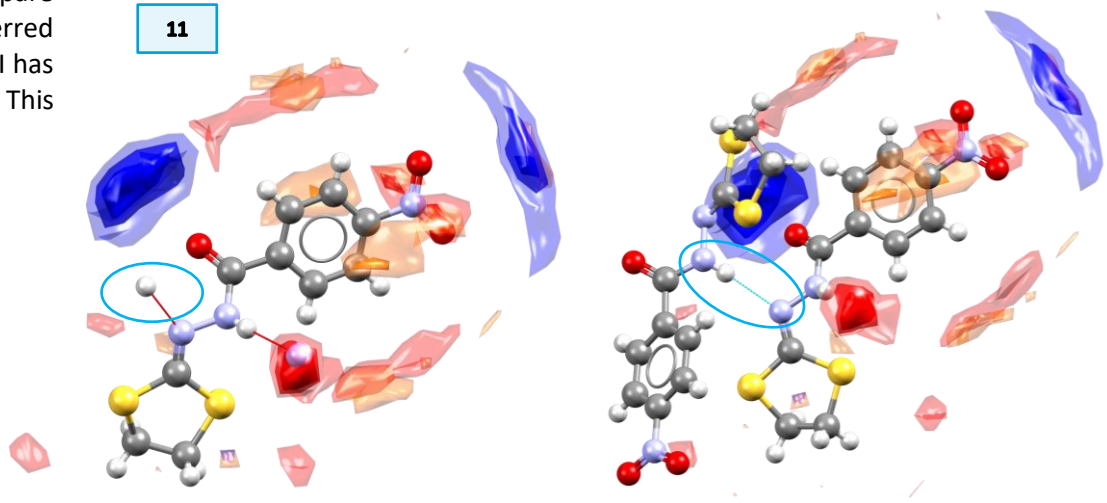


9. Now let's look at the Form III polymorph. In the **Structure Navigator toolbar**, type DEDMUX02. Click the **Reset** button in the Display Options toolbar to remove all the hydrogen bonding interactions.

10. In the **Full Interaction Maps** window click **Calculate Maps** again (see **Step 2**). You should now have a Full Interaction Map surrounding the molecule. Following **Steps 4** and **6** above, turn on the hydrogen bonding interactions.

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





Conclusion

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 analysis.

Summary

To summarise, in this workshop you have learnt how to use three complementary tools to understand polymorph stability in N'-(1,3-dithiolan-2-ylidene)-4-nitrobenzohydrazide. Using Hydrogen Bond Propensity, we performed a detailed analysis of potential hydrogen bonding networks based on the functional groups of the molecule and observed that one of the three known polymorphs exhibits an unfavourable hydrogen bonding network. Using other knowledge-based approaches, including Mogul geometry check and Full Interaction Maps, the least stable polymorph was confirmed.

You should now be able to:

- Set-up and run a Hydrogen Bond Propensity (HBP) calculation.
- Interpret the Propensity and Coordination scores.
- Compare the HBP results with reference structures.
- Run a Mogul geometry check on a structure and interpret the results.
- Generate and interpret simple Full Interaction Maps for a structure.

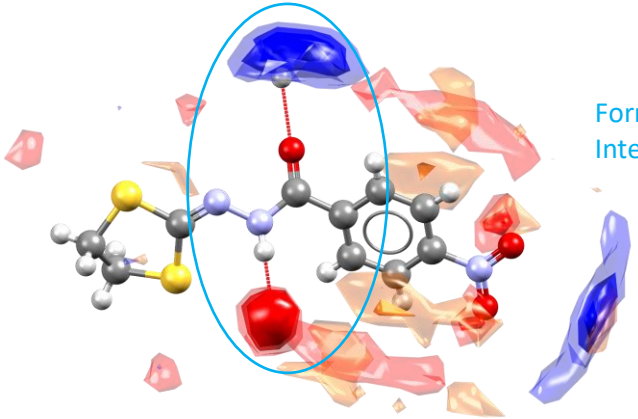
Next Steps

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).

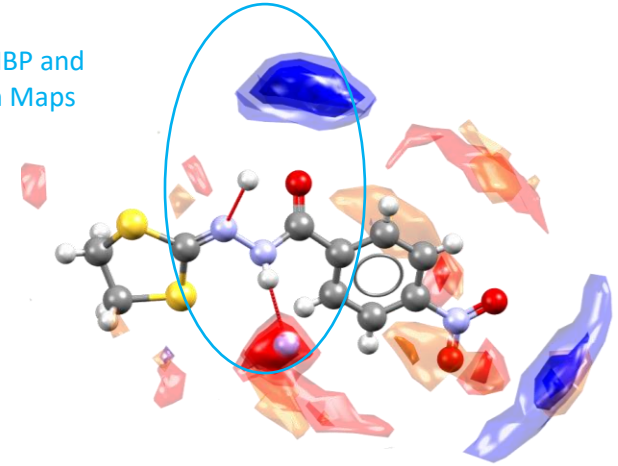
https://www.ccdc.cam.ac.uk/community/training-and-learning/workshop-materials/

Feedback

We hope this workshop improved your understanding of the *Hydrogen Bond Propensity* component 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-007 Thank you!



Form I: High HBP and Full Interaction Maps **satisfied**.

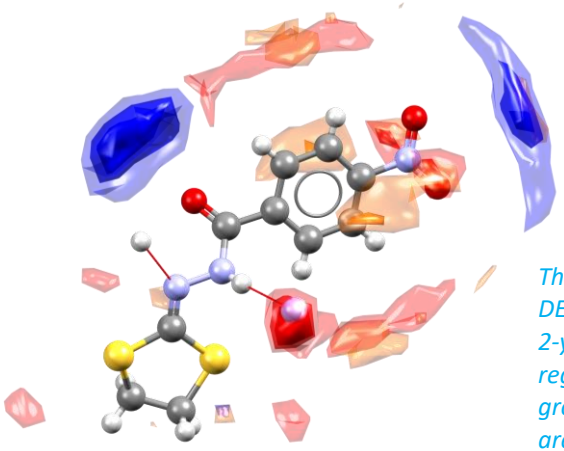


Form III: Low HBP and Full Interaction Maps **unsatisfied**.

Glossary

Full Interaction Maps

Full Interaction Maps provide a way of visualising the preferred interactions of a molecule. Regions around the molecule (maps), where chemical probe groups are likely to be found, based on pre-extracted IsoStar data from the CSD, are calculated. The program works by first identifying distinct functional groups in the molecule being studied, then finds the relevant data in IsoStar. The group-based interaction data is evaluated, taking into account the environmental effects of combinative factors and steric exclusion to create a 3D picture of molecular interaction preferences.



The Full Interaction Maps for CSD entry DEDMUX02, a polymorph of N'-(1,3-dithiolan-2-ylidene)-4-nitrobenzohydrazide. The blue regions show the likely locations of donor groups, the red regions show where acceptors are expected to be found and the brown regions indicate potential hydrophobic

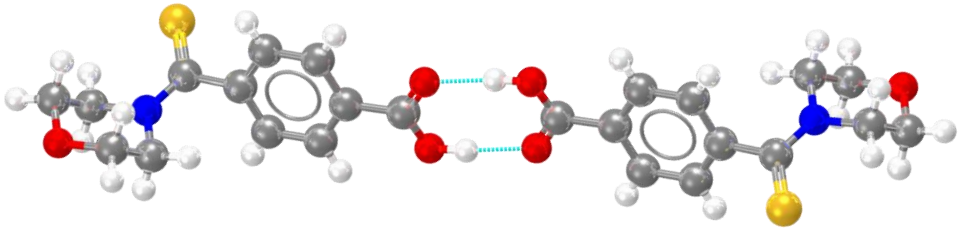
Graph Sets

Graph set analysis of hydrogen-bonding describes the pattern of the hydrogen bond chains or motif and includes the numbers of hydrogen-bond donors and acceptors.
A graph-set descriptor is written as $\mathbf{G^a_d(n)}$, in which **G** represents the type of pattern, **a** is the number of hydrogen bond acceptors involved in that pattern, **d** is the number of donors and **n** the number of atoms in the pattern. The pattern type, G, can be one of four different options: C for an infinite chain, S for an intramolecular hydrogen bonding pattern, R for an intermolecular ring and D for a discrete, finite hydrogen-bonding pattern.
Relevant bibliographic references:

• M. C. Etter, *Acc. Chem. Res.*, **23**, 120, 1990
• J. Bernstein, R. E. Davis, L. Shimoni and N.-L. Chang, *Angew. Chem. Int. Ed.*, **34**, 1555, 1995
• W. D. S. Motherwell, G. P. Shields and F. H. Allen, *Acta. Cryst. **B56***, 466, 2000

Hydrogen Bonds

Hydrogen bonding occurs between donor-acceptor interactions precisely involving hydrogen atoms. The H-bonds interactions are classified as: strong (mostly covalent), moderate (mostly electrostatic) and weak (electrostatic). Their strength is observed to be between 12 and 30 kJ/mol.



In light blue, example of hydrogen bonds for refcode MULWIC.

Hydrogen Bond Donor/Acceptor

If a typical hydrogen bond is depicted as X—H···Y—Z, where the dots denote the bond, X—H represents the hydrogen bond *donor*. The *acceptor* may be an atom or anion Y, or a fragment of a molecule, Y—Z, where Y is bonded to Z. The acceptor is an electron-rich region such as, but not limited to, a lone pair on Y or a π-bonded pair of Y—Z. [Source: E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. Dannenberg, P. Hobza, H. G. Kjaergaard, A. C. Legon, B. Mennucci and D. J. Nesbitt, *Pure Appl. Chem*., 2011, **83**, 1637–1641.]

Hydrogen Bond Propensity (HBP)

- The HBP tool in Mercury > CSD-Materials evaluates the relative likelihoods of possible H-bonding networks in any observed polymorphs of a target system.
- Probabilities for hydrogen bond pairings to form in the target system are calculated from a statistical model built from relevant structures in the CSD. The model encapsulates information regarding the environment of the functional groups, which ensures the prediction is specific to the target molecule.
- Combining probabilities of hydrogen bond formation with a statistical model that captures information regarding how often a functional group participates allows the generation of chemically sensible alternative structures.
- The view of the solid-state landscape of an active ingredient afforded through the combination of propensity and coordination addresses questions such as how likely polymorphism is and whether there is the possibility of a more stable form. Specifically, you can:
    - Predict likely hydrogen bonds for a given molecule.
    - Assess crystal forms, e.g., by identifying sub-optimal hydrogen bonding.
    - Calculate hydrogen bond propensities for individual donor and acceptor groups.
    - Perform a comprehensive analysis of hydrogen bonding on a set of structures.

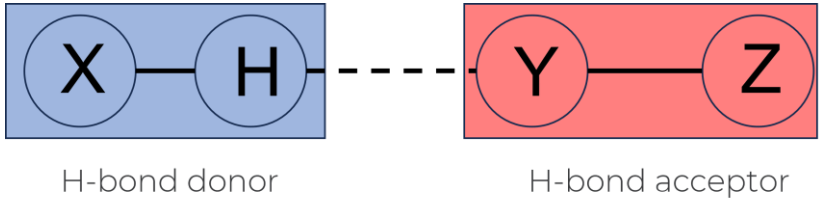
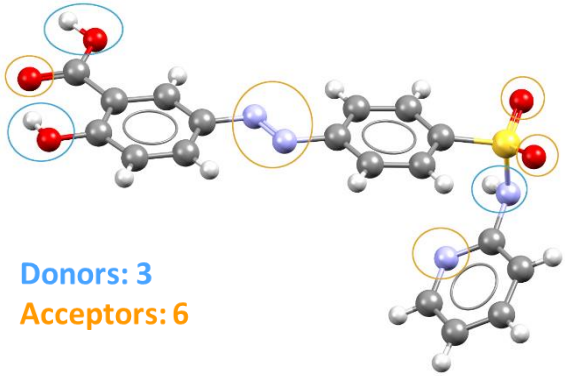


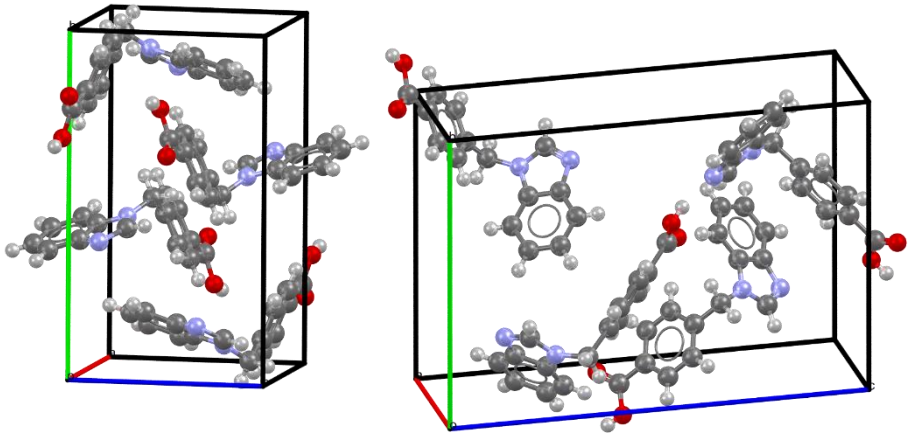
Illustration of a hydrogen bond interaction with between hydrogen bond donor X–H and hydrogen bond acceptor Y–Z.



Sulfasalazine exhibits 3 potential donors and 6 acceptors that might compete in forming H-bond interactions. HBP can be used to evaluate which of these potential interactions are more likely to form.

Polymorph

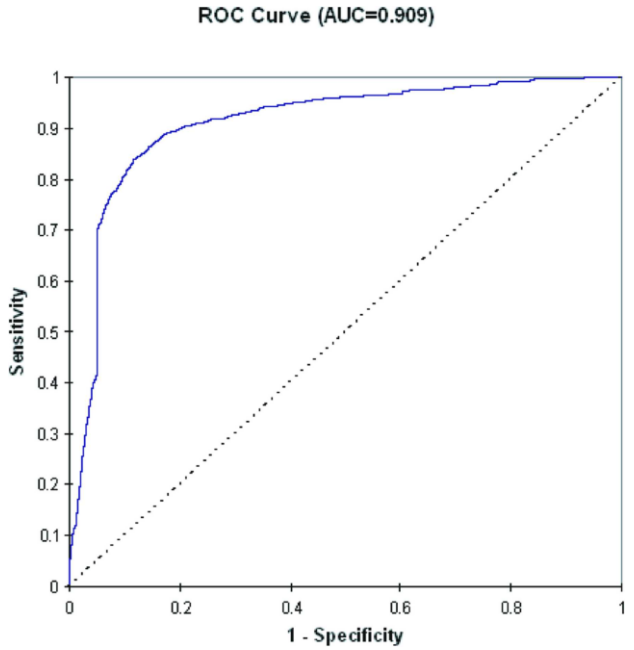
Polymorphism is the occurrence of two or more crystalline forms of the same substance. Where available, polymorph information can be displayed for Cambridge Structural Database (CSD) structures. Structures known to be polymorphic contain comments which include the word polymorph (when reported by the author), e.g., non-triboluminescent polymorph. There is also a CSD subset of polymorphic structures.



Example of polymorphic structures of 4-(1H-Benzimidazol-1-ylmethyl)benzoic acid: the monoclinic polymorph (CSD Entry ABADIS) at the top, and the orthorhombic polymorph (CSD Entry ABADIS01) on the right.

Receiver Operating Characteristics

The ROC curve (receiver operating characteristics) gives a measure of how well classified the predictions are using the training data as a test. It calculates percentage classification using a variable cut-off, either side of which a propensity is considered positive or negative. The *sensitivity* is the fraction of correct positive predictions and the *specificity* is the fraction of correct negative predictions. The diagonal dotted line is the outcome of a purely random model as there is an equal number of correct and incorrect predictions. An AUC (area under the curve) greater than 0.5 indicates the model predictions are correct more frequently than a random choice. An AUC of 1 indicates a perfect model: correct every time. An AUC above 0.8 is considered excellent and above 0.9 indicates an outstanding model. The difficult middle section around sensitivity/specificity = 0.5 needs a well discriminating model in less extreme cases, and must be well described by the model parameters in order to obtain a high AUC. One may observe that the example LHP model gives an outstanding classification of the training data and achieves an AUC of 0.909. [Reference: P. T. A. Galek, L. Fábián, W.D. Sameul Motherwell, F. H. Allen and Neil Feeder, *Acta Crystallogr. B*, 2007, **63**, 768 – 782]



ROC curve using the model to predict the training set outcomes The diagonal dotted line indicates the curve of a model with no predictive power: there is equal likelihood of a correct and an incorrect prediction.

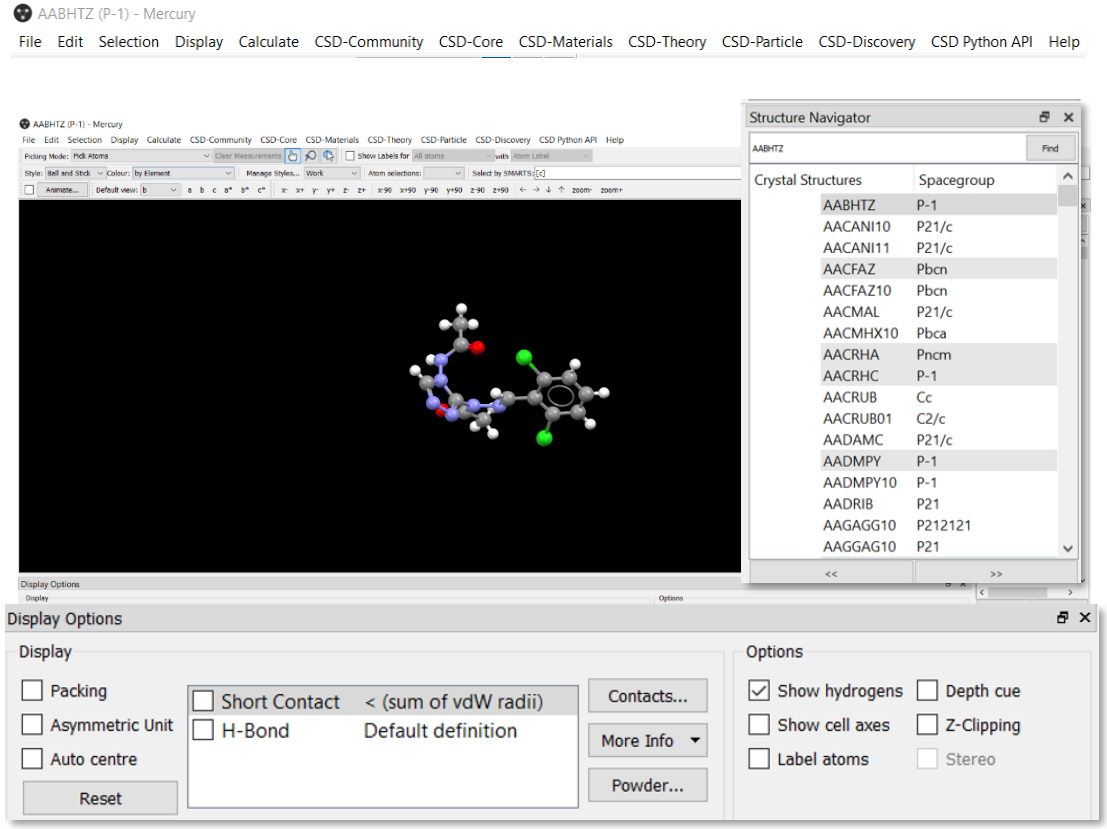
Basics of Mercury Visualization

Mercury is the CCDC's visualization software to view 3D structures of small molecules, generate images, and animations of molecules.

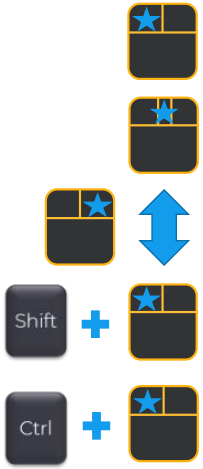
In the following we will see some of the basics of navigation and visualization in Mercury that you will find helpful to support your analysis.

In the **Mercury interface** we find:

- **At the top**: list of menus from which we can access visualization and analysis options, and other CSD components such as CSD-Materials.
- **On the right-hand side**: the **Structure Navigator**, with the database loaded (depending on your licence). The Structure Navigator allows you to select a refcode to visualize in the main Mercury window.
- Beneath the main display window: Display options toolbar. You can quickly view a packing diagram, display Hydrogen bonding and detailed information about the molecule using the More Info option.



Using the **mouse to enhance visualization**:



- Left mouse button and move – rotate molecules.
- Middle Mouse wheel – move molecules up and down.
- Right mouse button and move up and down – zoom in and out of molecules.
- Shift + Left mouse button and move - rotate in the plane molecules.
- Ctrl + Left mouse button and move - translate molecules.

Right click:

a) Near a molecule and

b) Away from a molecule