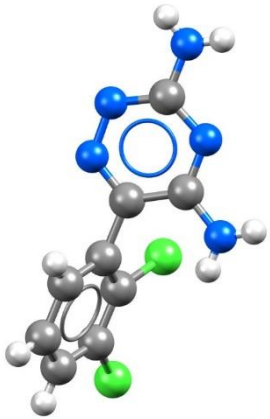
Hydrogen Bond Propensity

Version 1.1 – March 2019
CSD v5.39



The Cambridge Crystallographic Data Centre

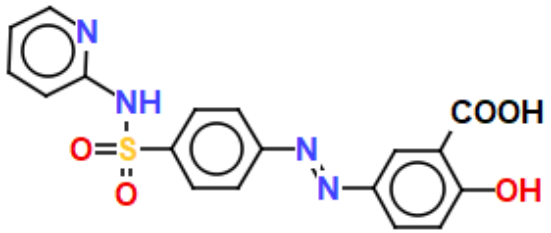
Table of Contents

Polymorph Assessment

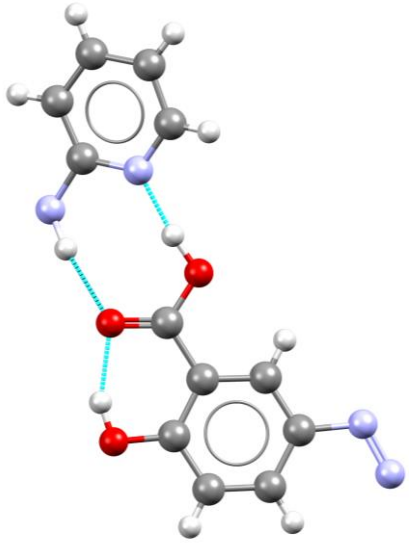
Various molecular arrangements that occur in the solid state, a phenomenon known as polymorphism, can have significant influence on a material's performance. Therefore, exploring the polymorphic landscape and identifying the most stable polymorph is an important process especially in early stage formulation in the pharmaceutical industry.

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

5



*Sulfasalazine (refcode **QIJZOY**)*



Example 1 A monomorphic system

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

Examine the Hydrogen bonding network

In this section we will examine the potential 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)

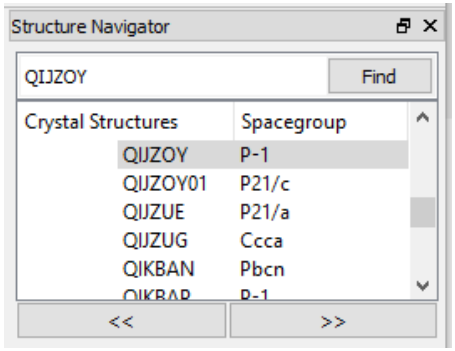
2. In the **Structure Navigator** window, type the refcode *QIJZOY*, to bring up the structure of sulfasalazine amide tautomer.

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

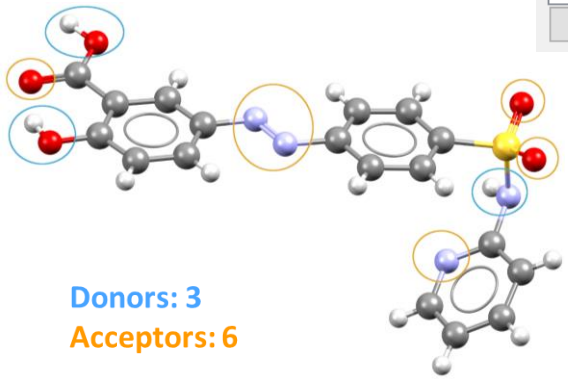
4. Toggle on the **H-Bond** check box in the *Display Options* to investigate how many of the potential donors and acceptors are utilised by sulfasalazine.

1

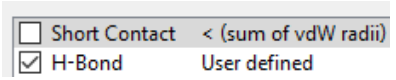
2



3



4

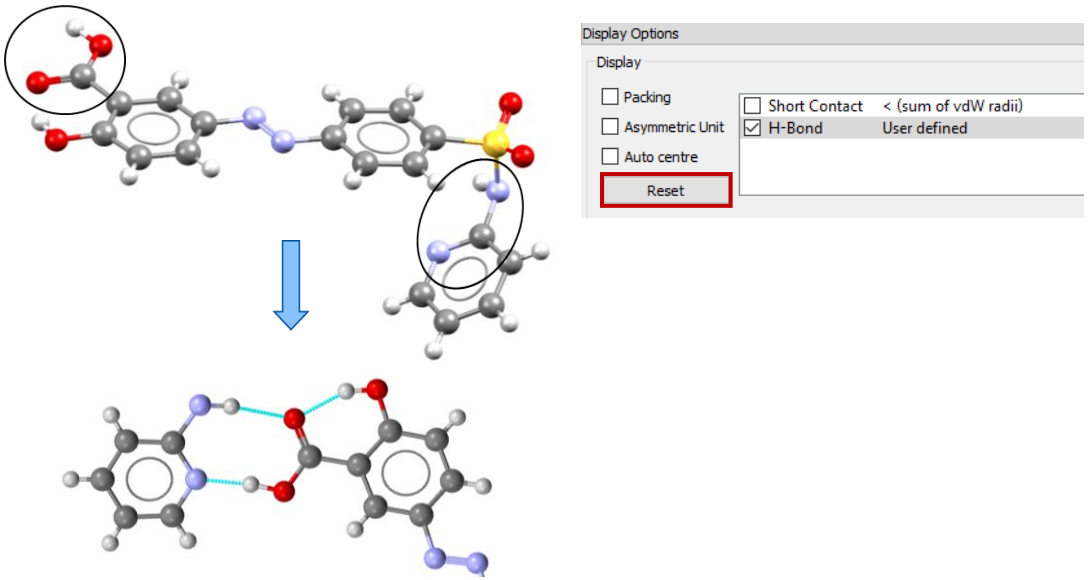


5. Two of the acceptors and two donors are used in intermolecular interactions, forming centrosymmetric dimers involving the carboxylic acid and pyridylamino functional groups. An intramolecular hydrogen bond is also formed between the hydroxyl group and the O atom of the carboxylic group. Press **Reset**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

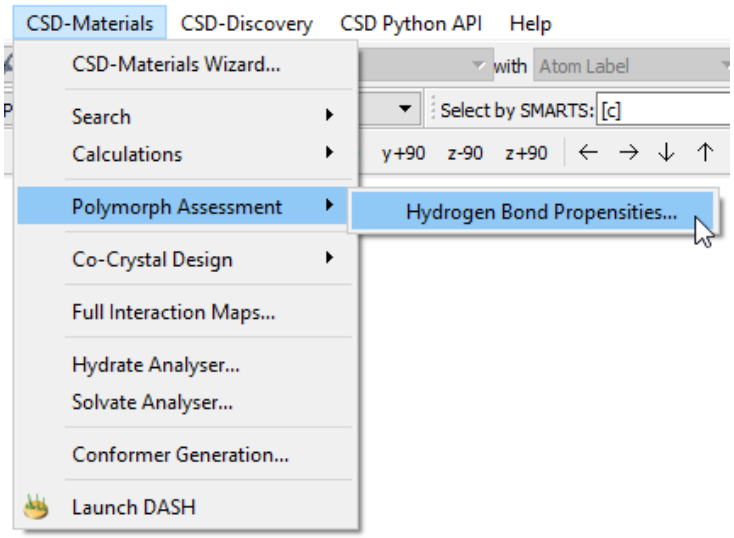
Calculate H-bond Propensity

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

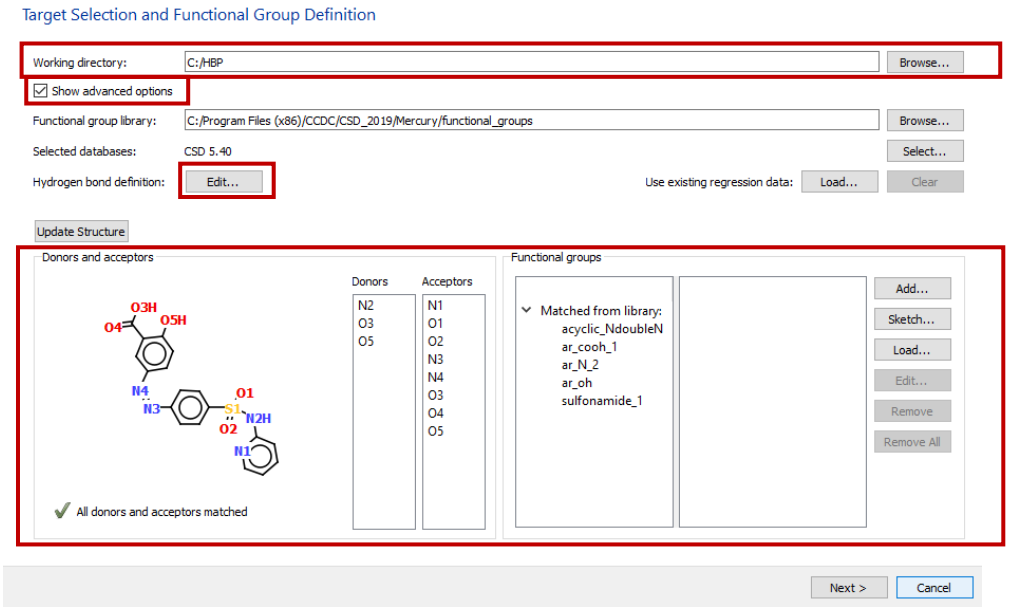
7. In the *Propensity Prediction Wizard* select a working directory by clicking on **Browse…** The potential hydrogen bond donor and acceptor atoms are automatically identified and linked to their functional groups. Three donors have been identified: N2 as sulfonamide_1, O3 as ar_cooh_1, and O5 as ar_oh. 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.



6



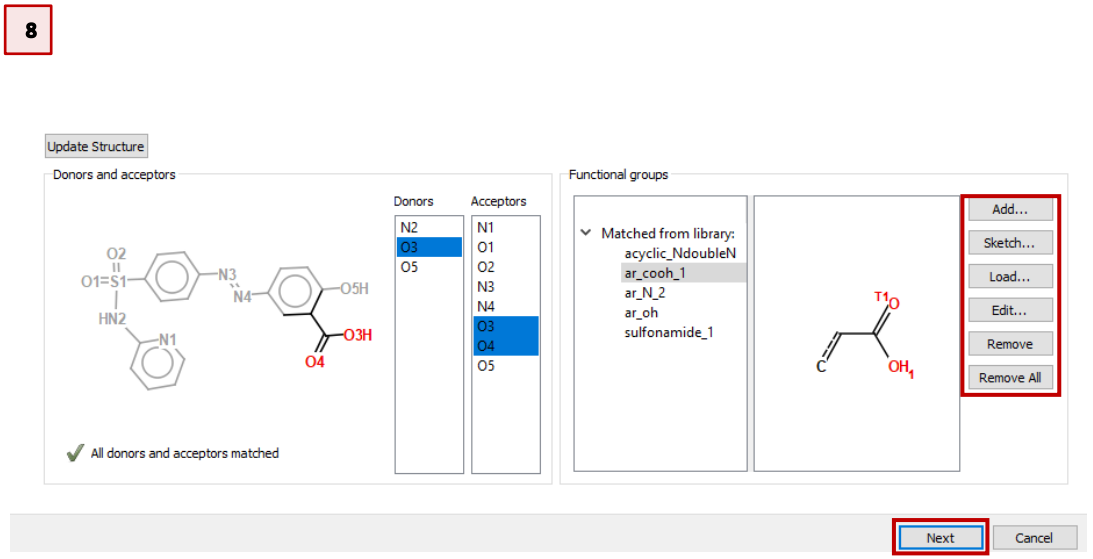
7

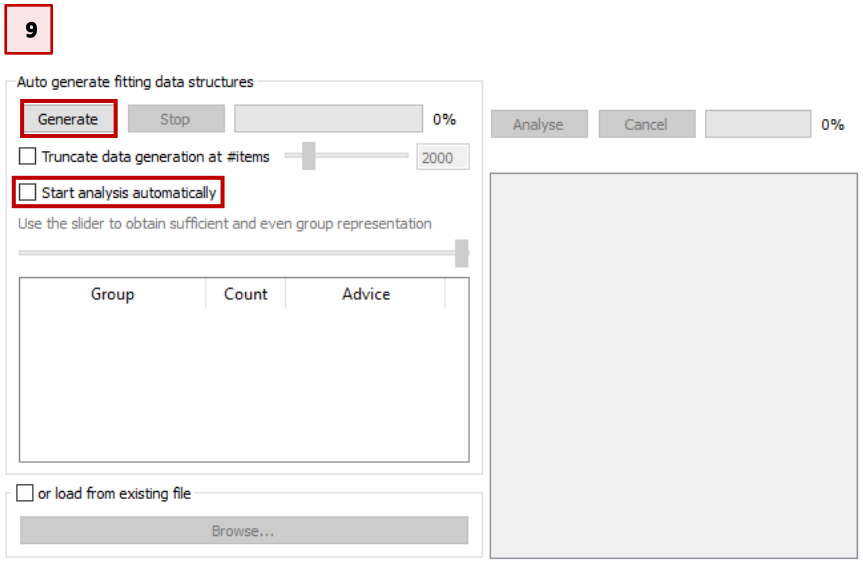


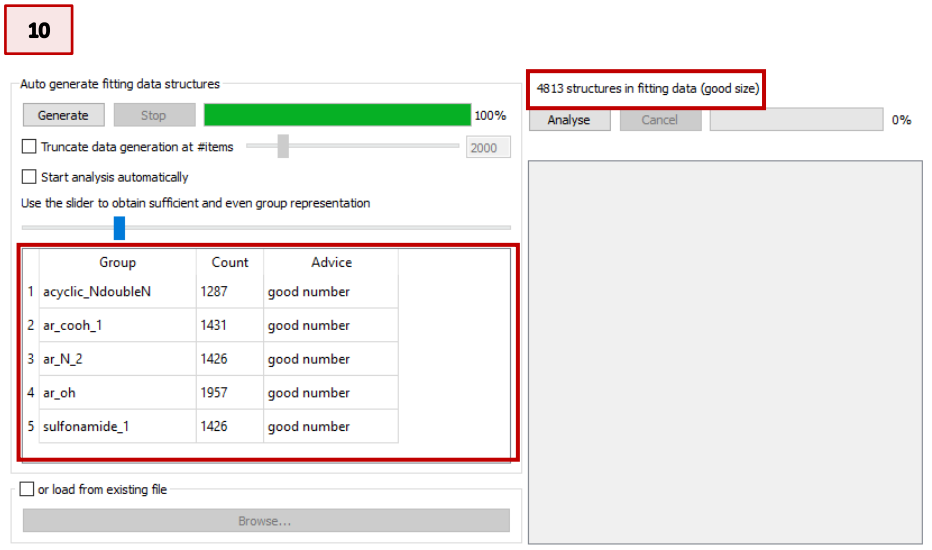
8. The *Donors* and *Acceptors* atoms can be highlighted in the 2D chemical diagram by selecting them from the list. You can also highlight a functional group from the *Match from library* list; the corresponding atoms will be automatically highlighted in the *Donors/Acceptors* lists. The functional 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 can be uneven, yet it is a good practice to aim for a model with groups evenly represented.





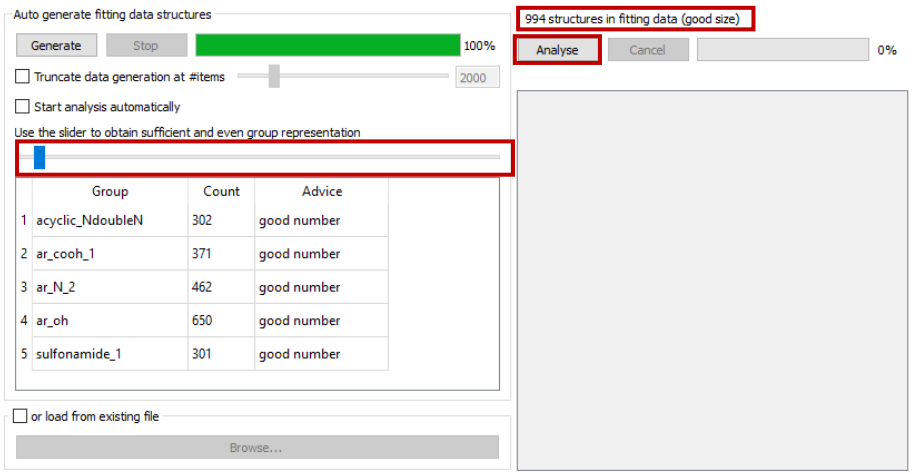


11. When the run is finished, adjust the group number by using the slider highlighted in blue. This allows you to remove or add structures until a more even set of data is obtained. In general, around 300-400 structures per functional group should be enough. Select around 800-1000 structures in total,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

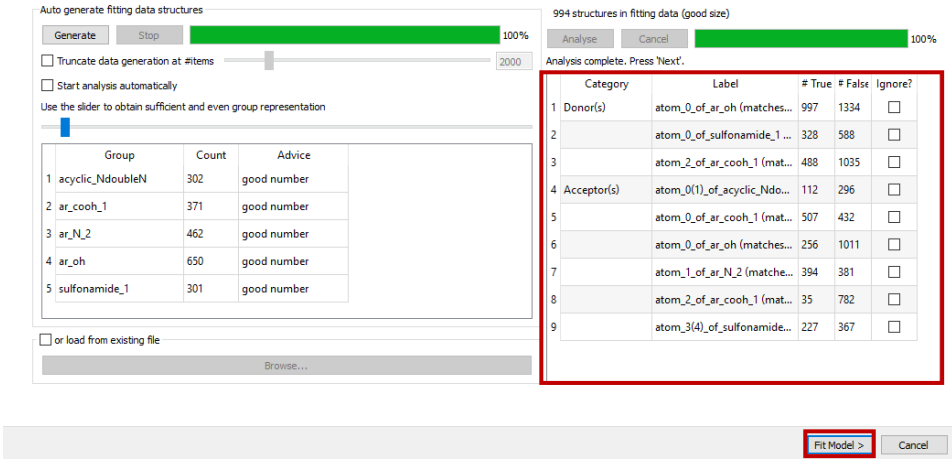
13. For this example, the Area under the ROC (receiver operating characteristic) curve (AUC) should be around 0.82. To achieve a good H-bond propensity calculation you should always aim for an AUC of around 0.75 or above. Click **Accept & Calculate** to continue.

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

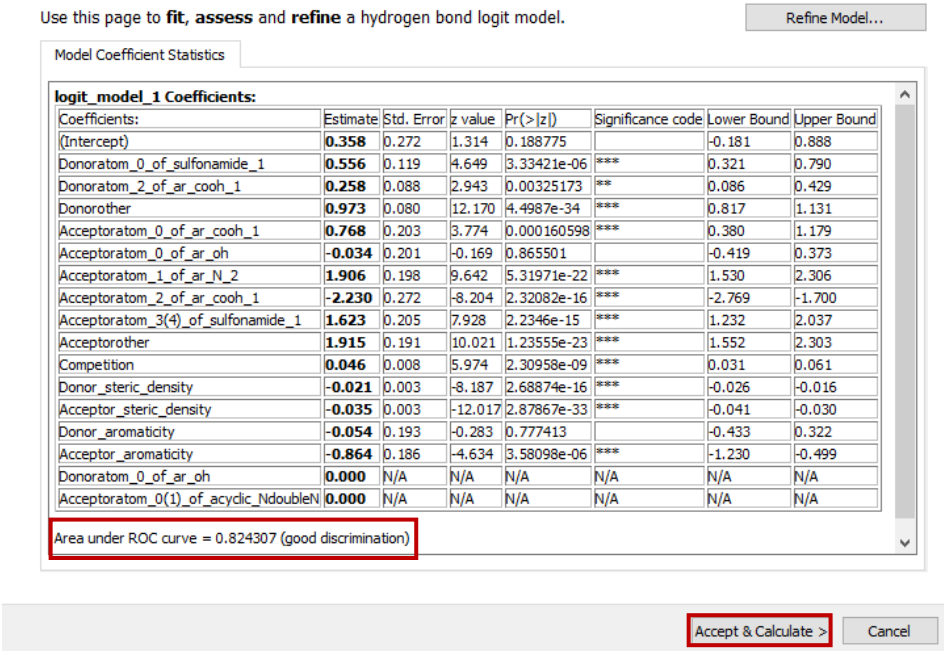
12



	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



logit_model_1 Coefficients:

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

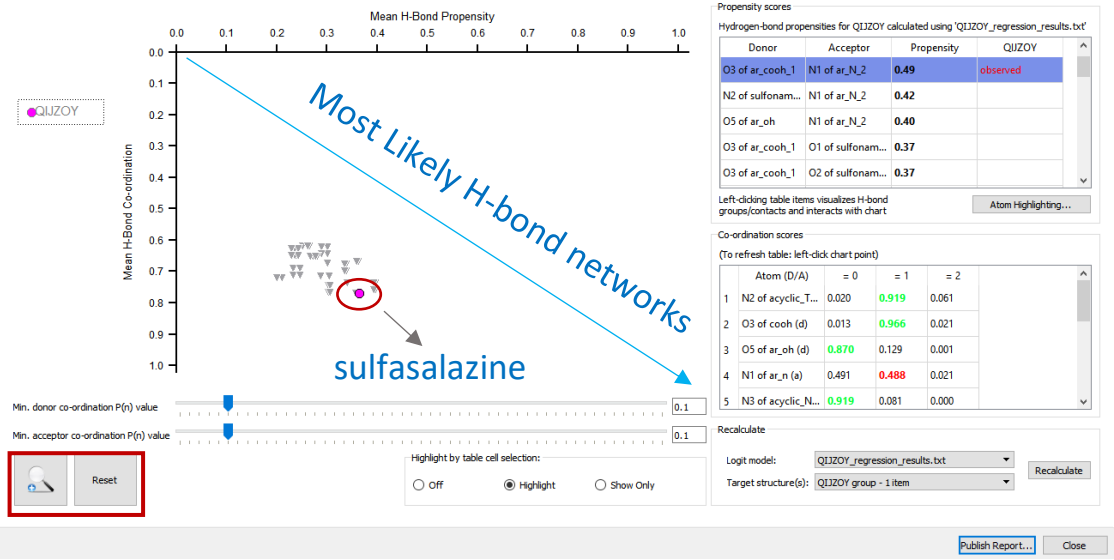
- the most likely H-bonding network will score the highest propensity and will be listed first in the table
- the H-bonds present in the targeted structure are marked as observed
- the table is interactive, clicking on **observed** will highlight the donor and acceptor group in the 3D visualizer, clicking 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.

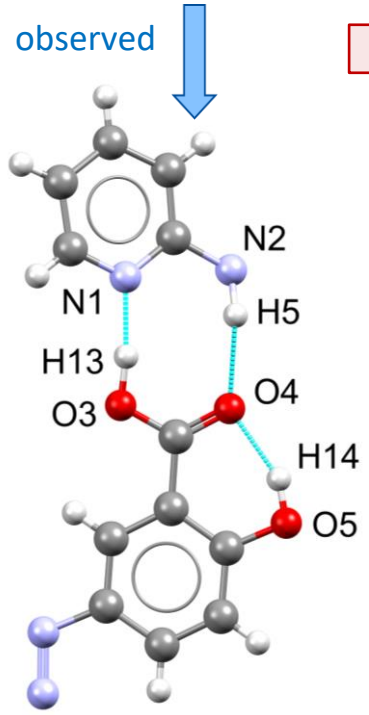
15. *Co-ordination Scores Table:*

- (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.

14





15

Co-ordination scores

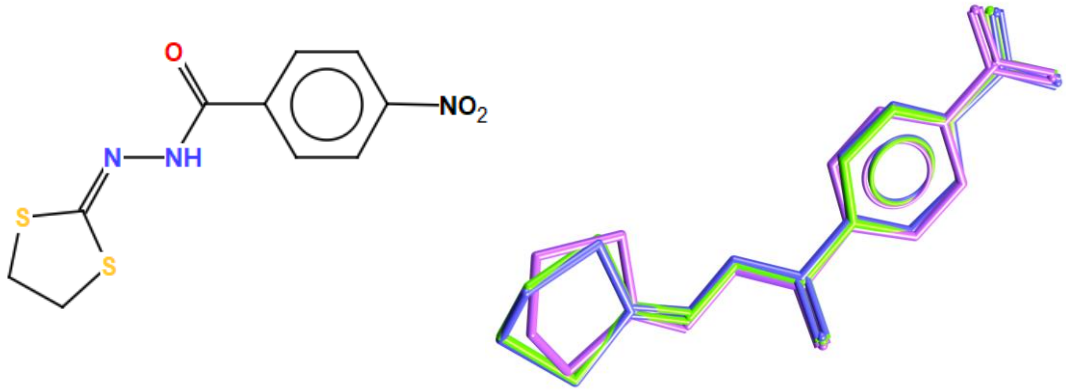
(To refresh table: left-click chart point)

	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 Two polymorphic system

N'-(1,3-dithiolan-2-ylidene)-4-nitrobenzohydrazide, a potentially tuberculostatic agent, is known to crystallise into three polymorphs. 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.



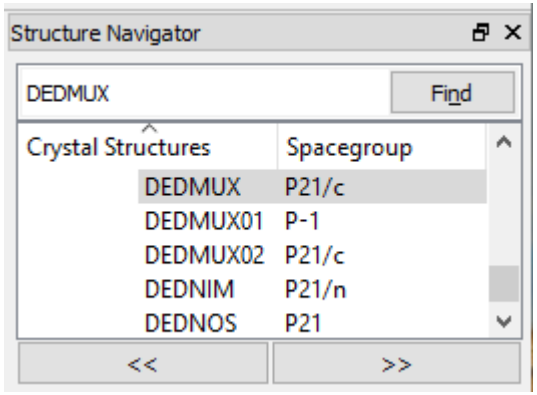
*N'-(1,3-dithiolan-2-ylidene)-4-nitrobenzohydrazide (refcode **DEDMUX**).*
Form III (purple) has a different geometry

Examine H-bond network

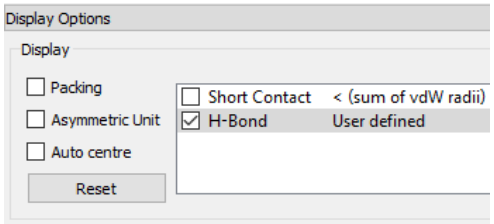
1. Start Mercury by double-clicking the icon on your Desktop or navigating from the Start Menu (Start > CCDC > Mercury)
2. In the **Structure Navigator** window, type the refcode *DEDMUX*, to bring up the structure of the first polymorph.
3. Toggle on the **H-Bond** check box in the *Display Options* check box 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 there are?

1

2

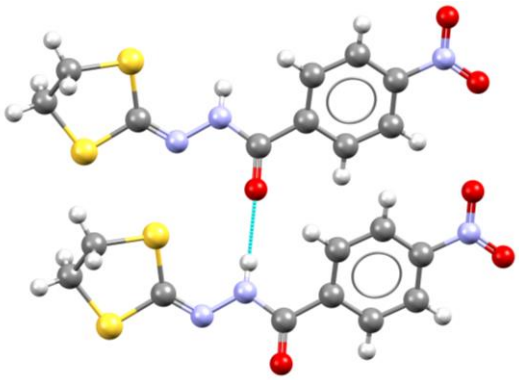


3

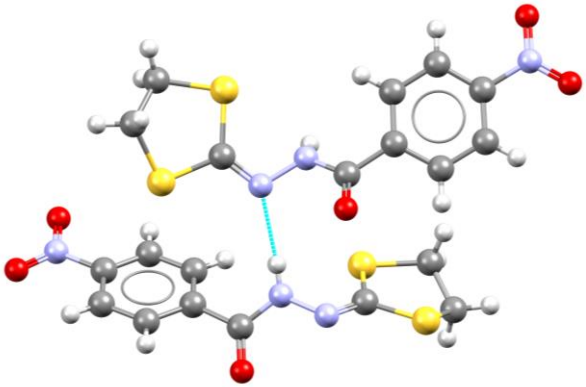


H-bond network

Form I and II



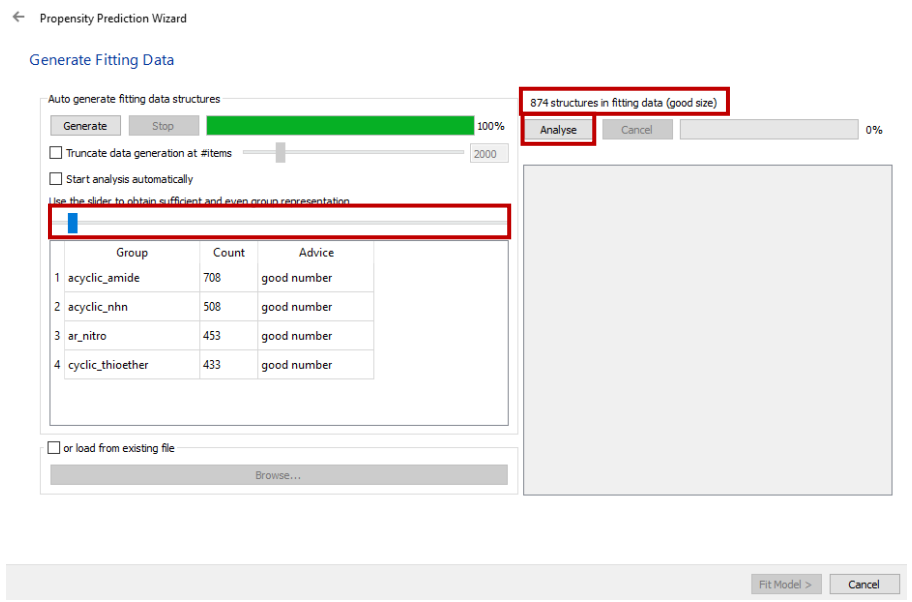
Form III



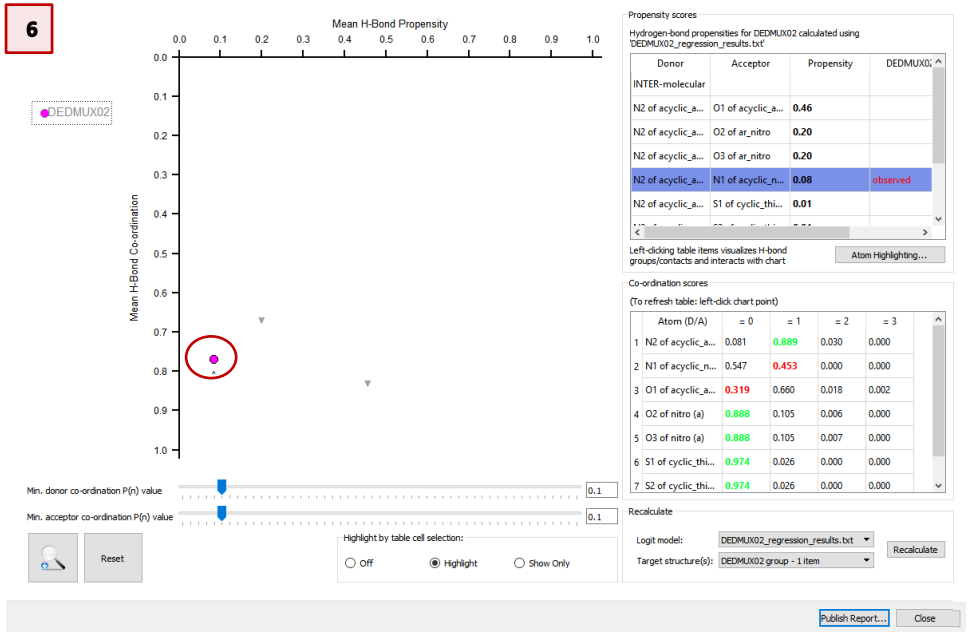
Calculate H-bond Propensity

4. Repeat step 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 Form 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 **Entire refcode family**, then click **OK**. You can see the three DEDMUX refcodes in the **Selected structure(s)** pane. Click **OK**, then click **Recalculate**.

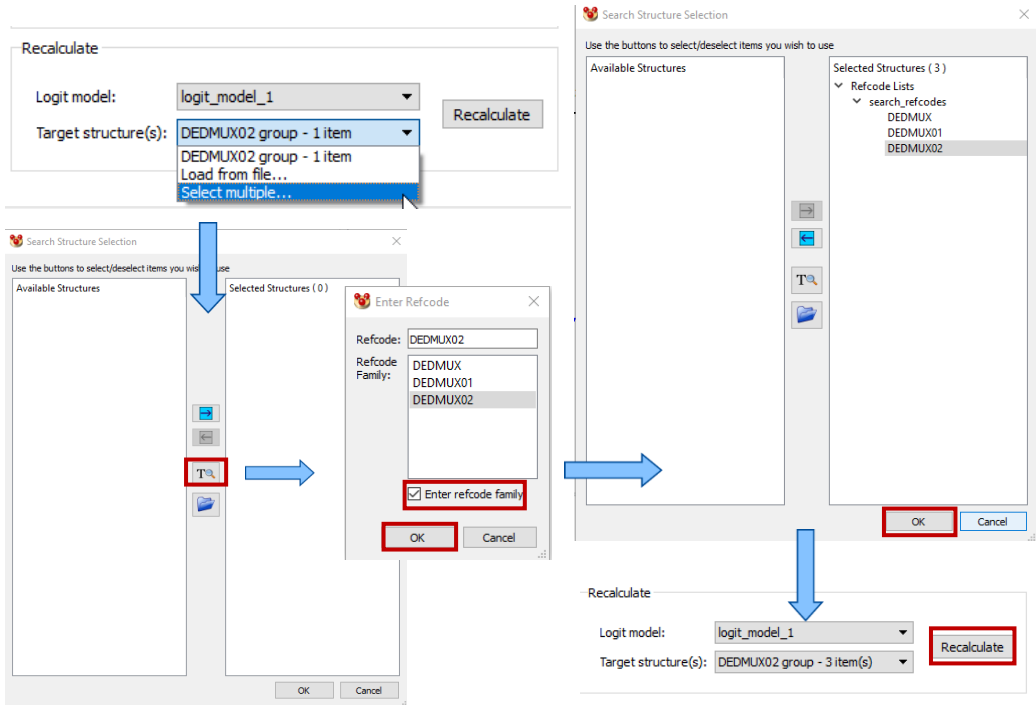
5



6



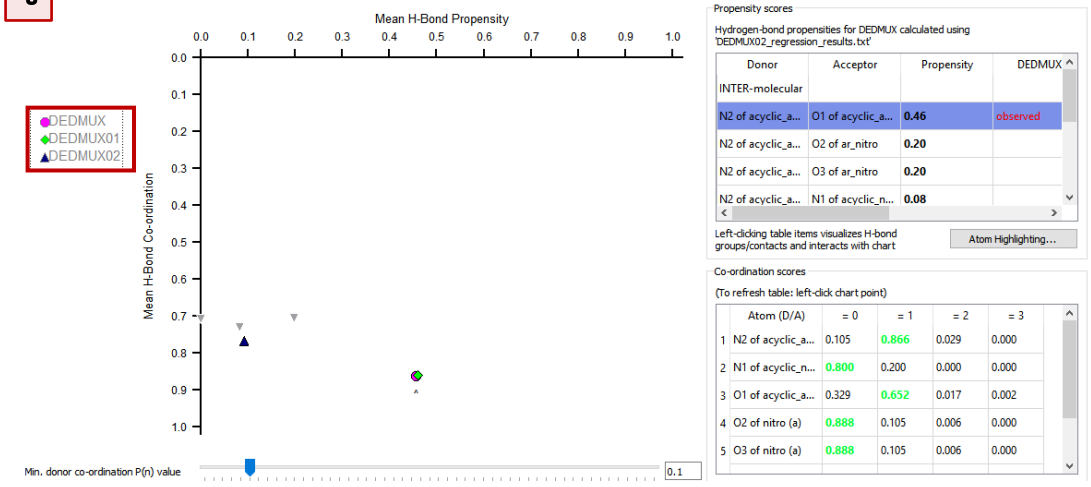
7



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 form I and form 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



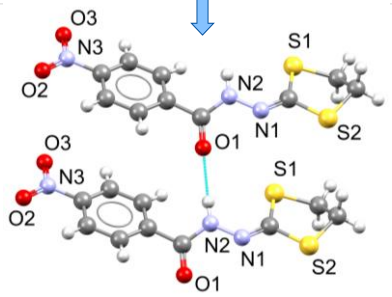
9

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