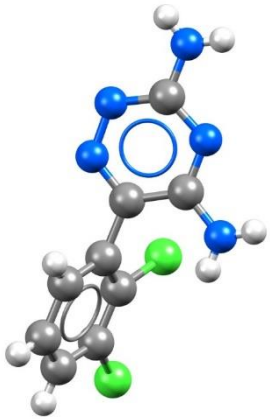
Exploring Hydrogen Bond Propensities

MAT-001

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 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.

In this workshop, we will explore the basics of running a Hydrogen Bond Propensities (HBP) analysis and assess the potential for a compound to exhibit polymorphism.

Learning Outcomes

After completing the workshop, you will:

- Be familiar with the Hydrogen Bond Propensities tool in Mercury.
- Know how to perform a HBP analysis and how to read and interpret the results.
- Understand the potential to use HPB in polymorphs analysis.

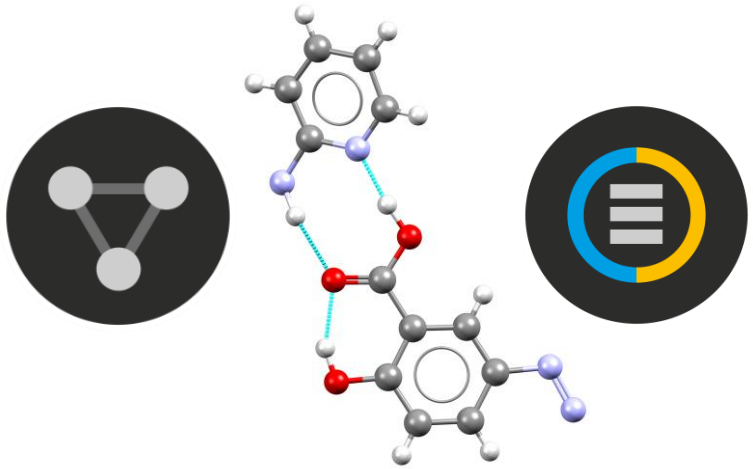
This workshop will take approximately **30 minutes** to be completed. The words in *Blue Italic* in the text are reported in the Glossary at the end of this handout.

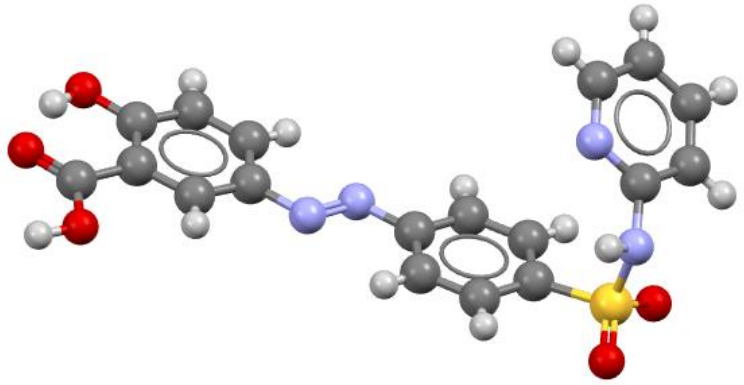
Pre-required Skills

Familiarity with the Mercury interface is important; you can access the Visualization in Mercury self-guided workshop here. A summary is given at the end of this document here.

Materials

There are no additional materials required for this workshop.

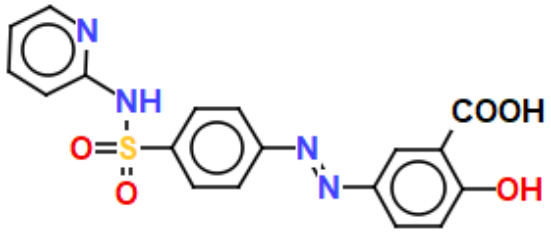




*Sulfasalazine (refcode **QIJZOY**).*

Example 1. A monomorphic system

Sulfasalazine is used to treat ulcerative colitis and Crohn's diseases. Only one polymorph has been reported 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**).*

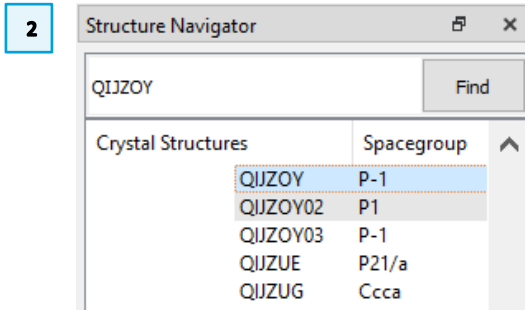
Examine H-bonding network.

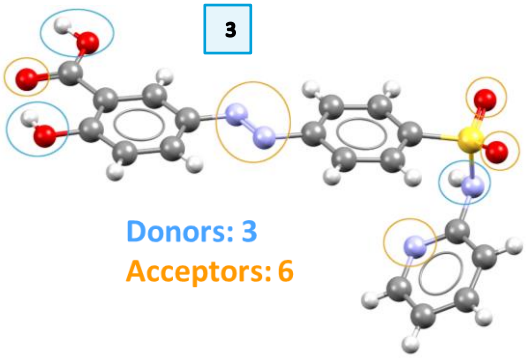
In this section we will examine the potential *hydrogen bond* *donors* and *acceptors* present in sulfasalazine.

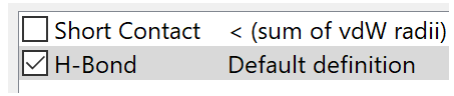
1. Start Mercury by double-clicking the icon on your Desktop or navigating from the Start Menu (*Start > CCDC > Mercury*)

2. In the **Structure Navigator** window, type the refcode "QIJZOY", to bring up the structure of sulfasalazine amide tautomer. The structure will be displayed in the 3D visualiser. There are 3 potential donors and 6 acceptors.

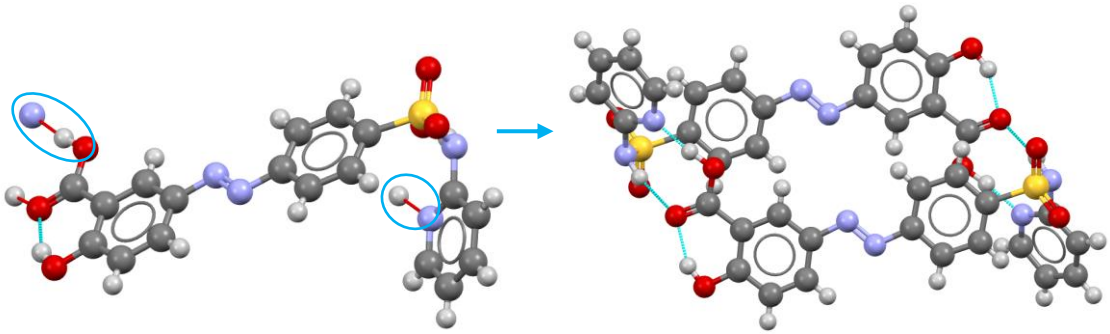
3. Toggle on the **H-Bond** check box in the **Display Options** toolbar to investigate how many of the potential donors and acceptors are utilised by sulfasalazine. Reveal the hydrogen bonded molecules by clicking on the red dotted lines or hanging atoms.





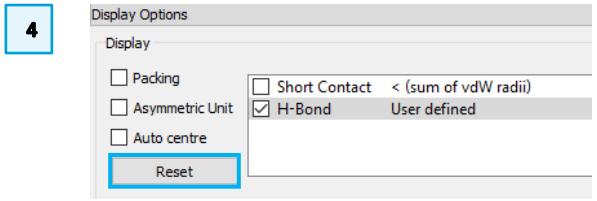


Two of the acceptors and two donors are used in intermolecular interactions, forming centrosymmetric dimers involving the carboxylic acid and pyridylamino functional groups. An intramolecular hydrogen bond is also formed between the hydroxyl group and the O atom of the carboxylic group.

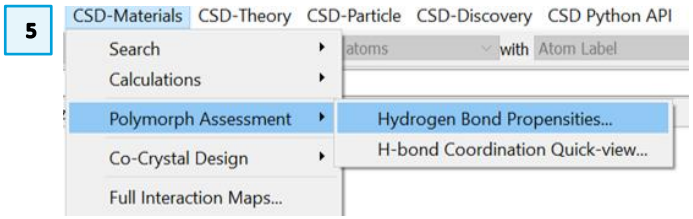


Calculate Hydrogen Bond Propensities

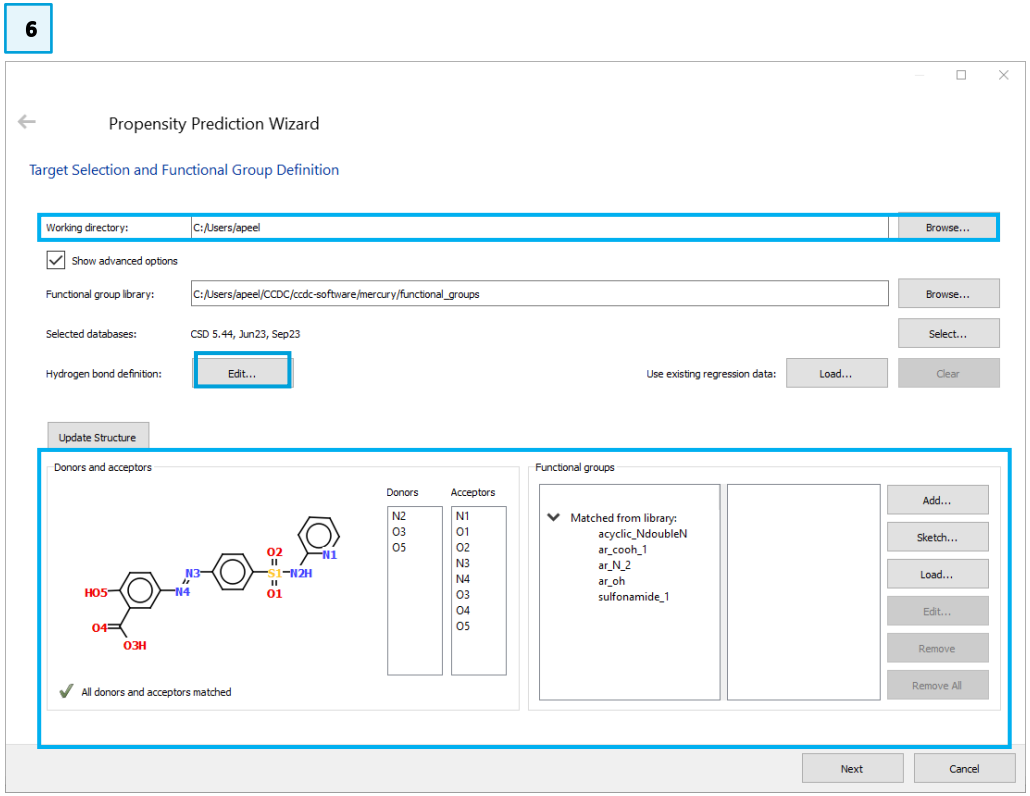
4. Press **Reset** in the Display Options toolbar.



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



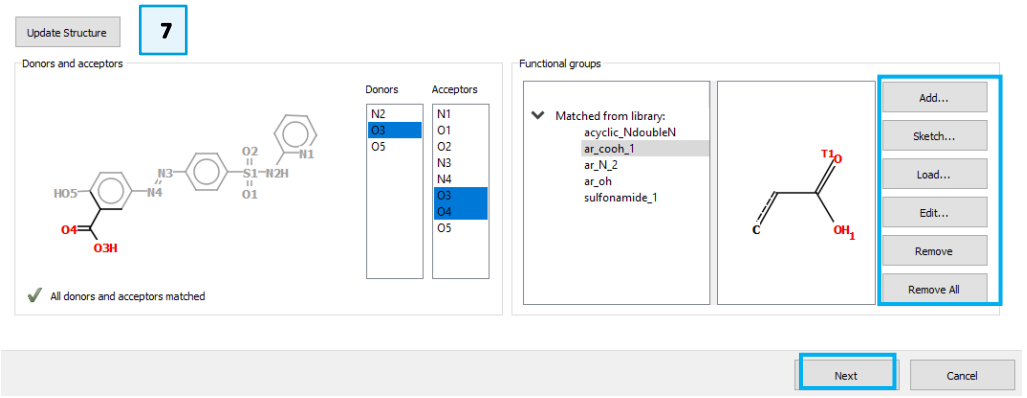
6. In the *Propensity Prediction Wizard* select a working directory by clicking on **Browse…** The potential hydrogen bond donor and acceptor atoms are automatically identified and linked to their functional groups. Three donors have been identified: N2 as sulfonamide_1, O3 as ar_cooh_1, and O5 as ar_oh. Eight acceptors have also been identified. Note that O3 and O5 are identified as both donor and acceptor as standard for a hydroxy group.



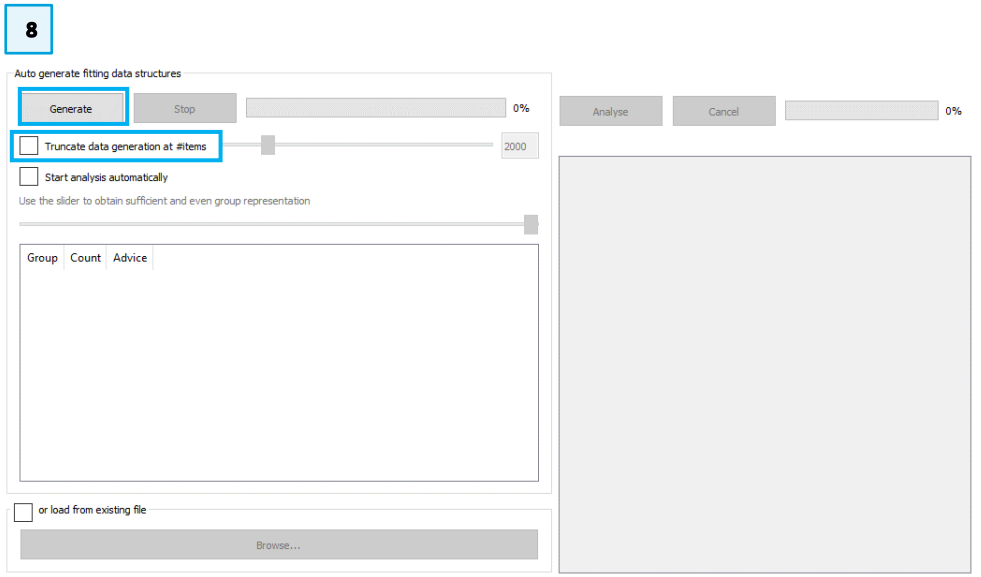
Tips and tricks

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…**

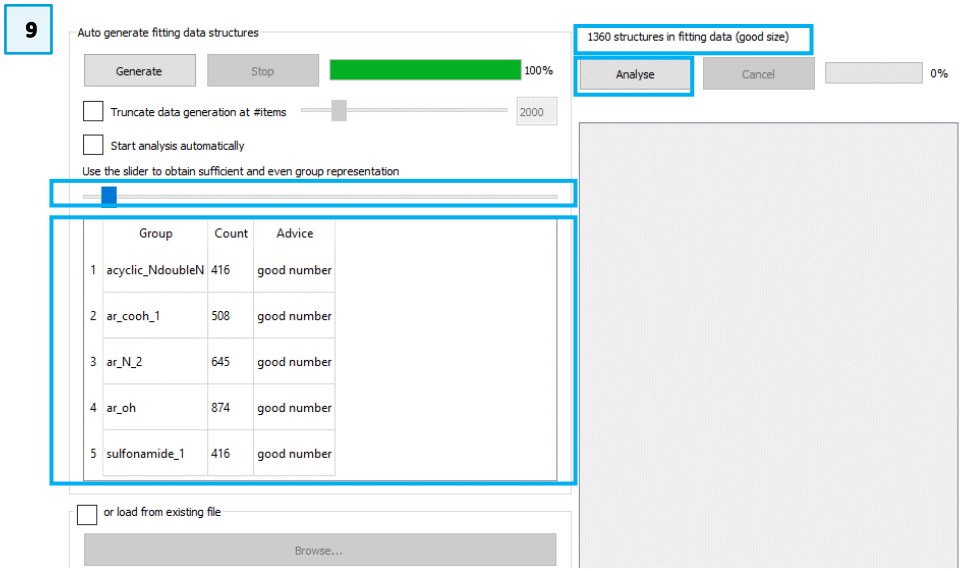
7. The *Donors* and *Acceptors* atoms can be highlighted in the 2D chemical diagram by selecting them from the list. You can also highlight a functional group from the *Match from library* list; the corresponding atoms will be automatically highlighted in the *Donors/Acceptors* lists. The functional 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**.



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.



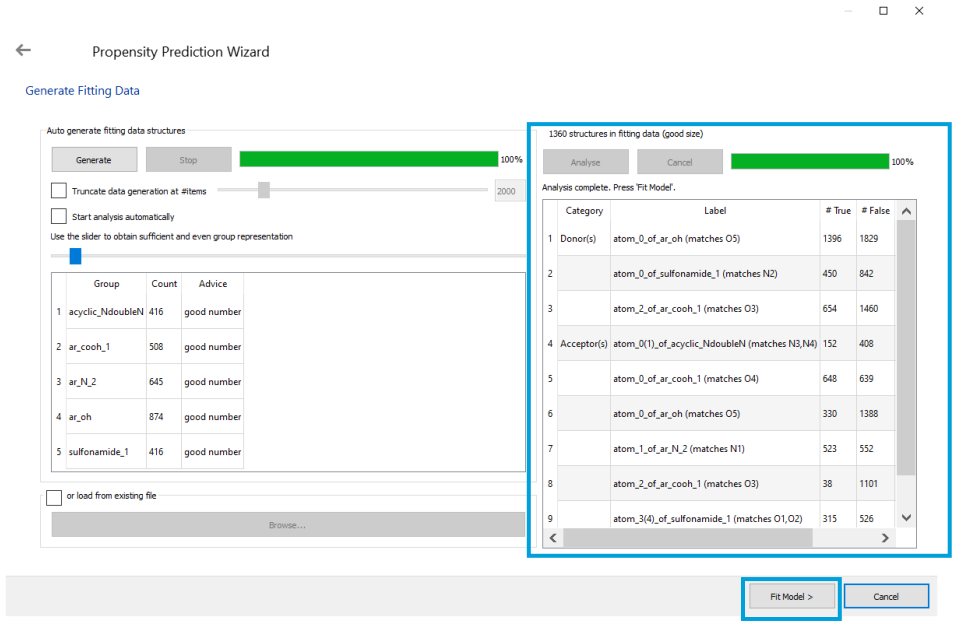
9. When the run is finished, it attempts to automatically select a sufficient number of hits (count) per functional group with fairly even representation across the groups. In general, around 400-500 structures per functional group should be enough. The group numbers can be adjusted by using the slider 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. Move the slider to select around 1400 structures in total and click **Analyse**.



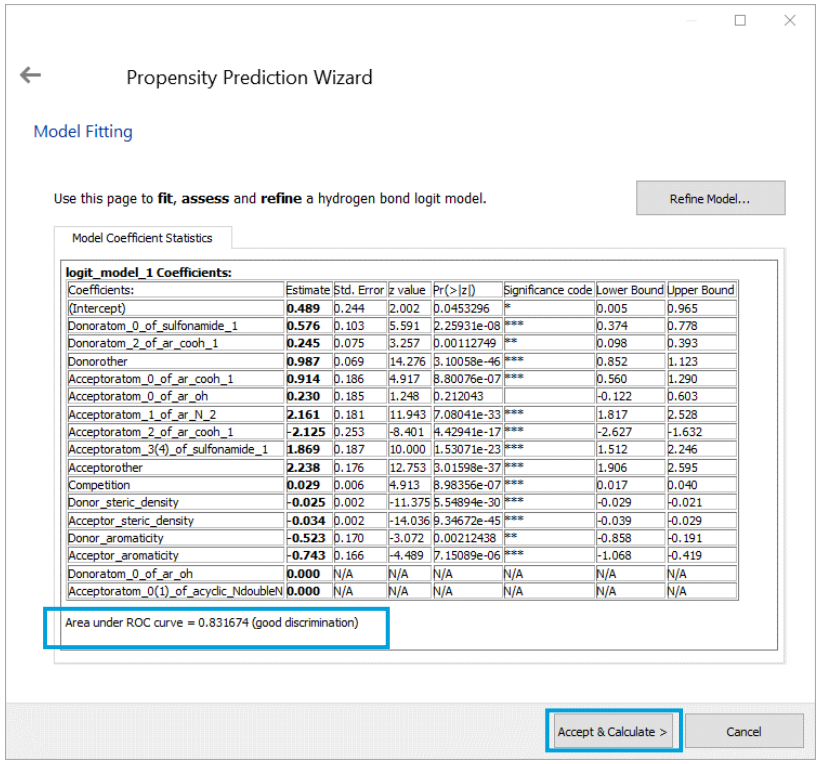
10. When the analysis is finished the number of the True and False outcomes will be listed. If there are very low numbers for True or False, you should check that they are ticked in the **Ignore?** checkboxes. There are no very low values in this example. Click the **Fit Model >** button to continue.

11. For this example, the Area under the *ROC* (receiver operating characteristic) curve (AUC) should be around 0.83. 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.

10



11

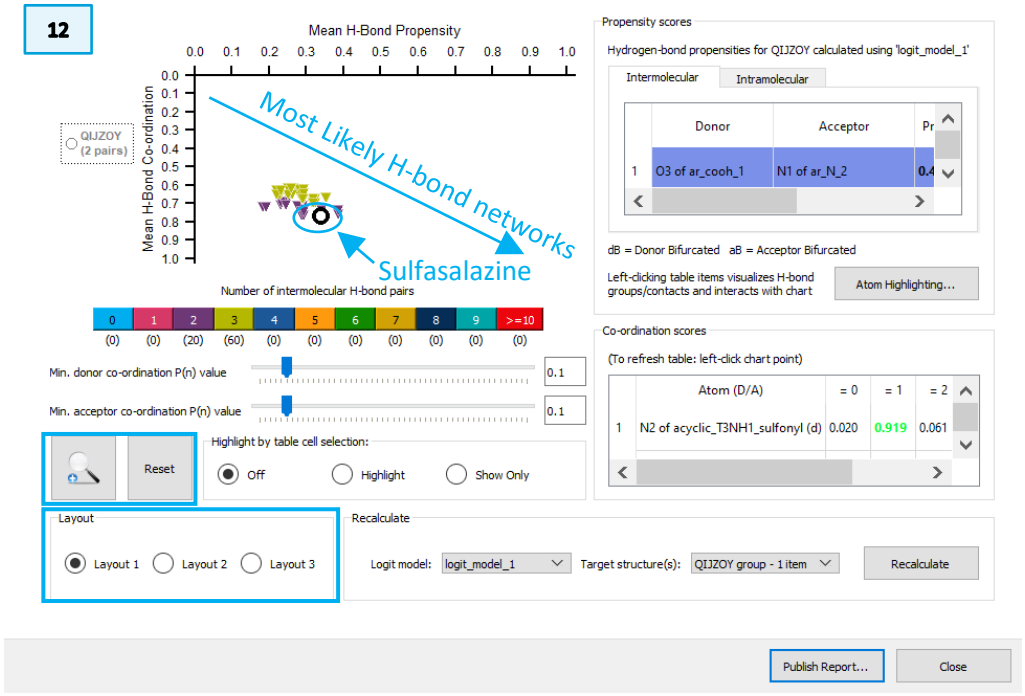


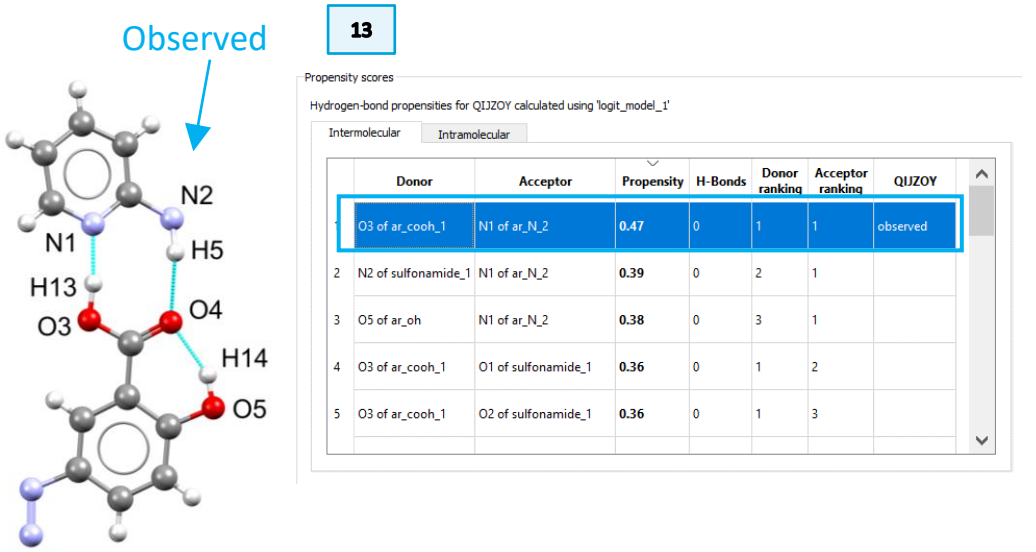
Summary of HBP results

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.
    - 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 network in the *Propensity score* table.
    - Hover over a point to display the mean propensity and mean co-ordination values.

13. *Propensity Score 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** 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.





14. *Co-ordination Scores Table:*
    - (d) stands for donor and (a) stands for acceptor.
    - = 0, = 1, = 2 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 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.

Conclusion

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

14 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

Summary

In this workshop, you have learnt how to use Hydrogen Bond Propensities in Mercury and have uncovered how this can enable you to predict different hydrogen bonding networks and the likelihood of these occurring. You should now:

- Be able to run a HBP analysis in Mercury.
- Interpret the results in terms of propensities and coordination scores.
- Understand how this can be used to explore hydrogen bond network stabilities.

For your reference, you can find the Mercury user guide here.

Next Steps

To explore Hydrogen Bond Propensities further and learn complementary techniques to assess hydrogen bonding networks, you can try the self-guided workshop "Investigating Solid Form Stability: Understanding Hydrogen Bond Propensities" (MAT-007) which is available in the CSD-Materials workshop area on our website.

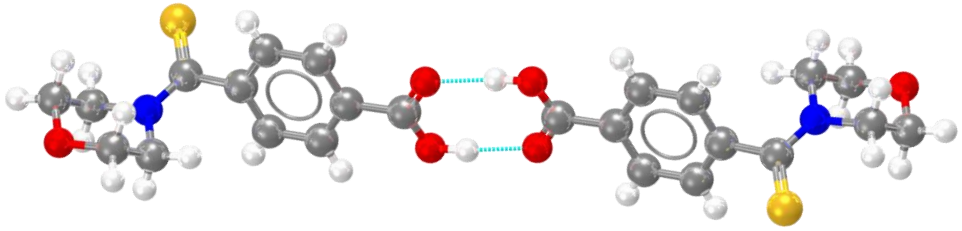
https://www.ccdc.cam.ac.uk/community/training-and-learning/workshop-materials/

Feedback

We hope this workshop improved your understanding of *Hydrogen Bond Propensities* and you found it useful for your work. As we aim to continuously improve our training materials, we would love to hear your feedback. Follow the link on the workshop homepage and insert the workshop code, which for this self-guided workshop is MAT-001. It will only take 5 minutes and your feedback is anonymous. Thank you!

Glossary

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.]

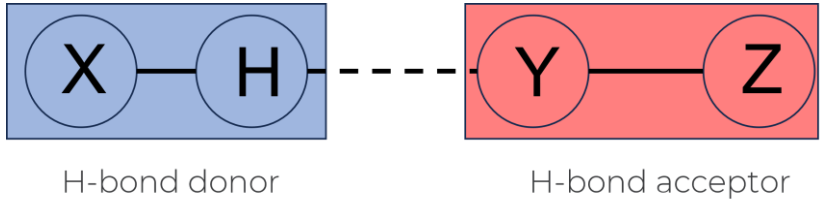
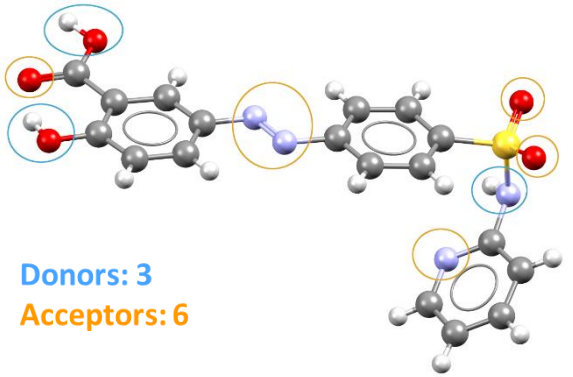


Illustration of a hydrogen bond interaction with between hydrogen bond donor X–H and hydrogen bond acceptor Y–Z.

Hydrogen Bond Propensity (HBP)

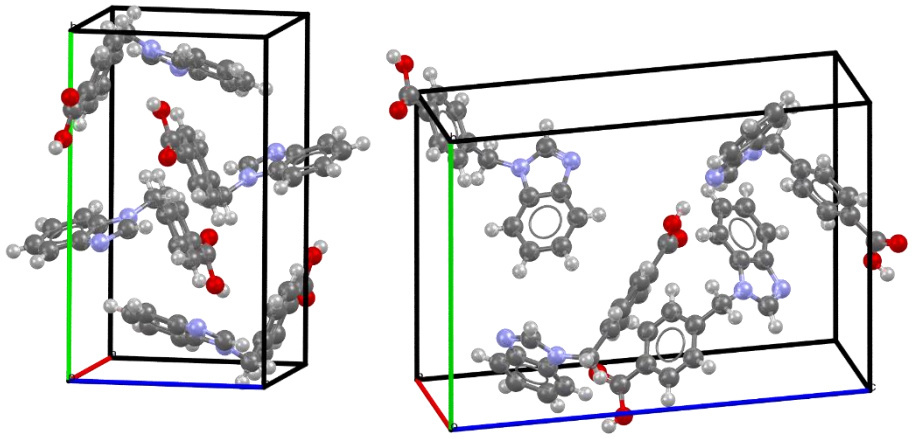
- The HBP tool in Mercury > CSD-Materials evaluates the relative likelihoods of possible H-bonding networks in any observed polymorphs of a target system.
- Probabilities for hydrogen bond pairings to form in the target system are calculated from a statistical model built from relevant structures in 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.



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

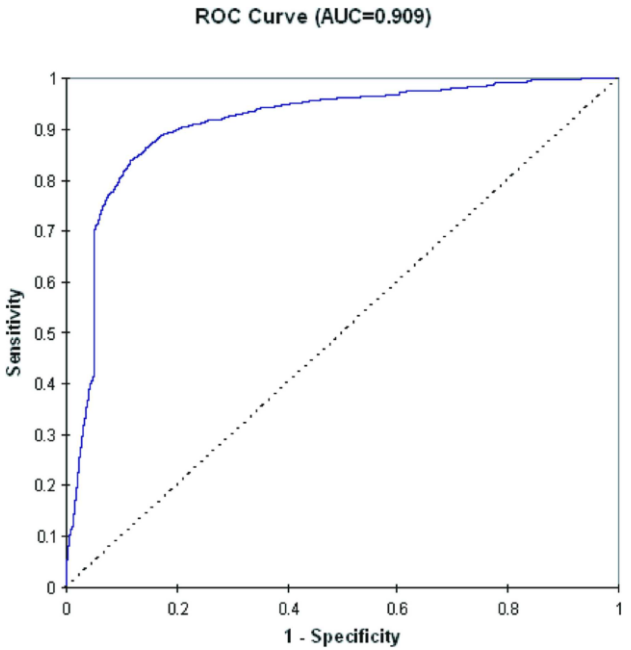
Polymorphism is the occurrence of two or more crystalline forms of the same substance. Where available, polymorph information can be displayed for Cambridge Structural Database (CSD) structures. Structures known to be polymorphic contain comments which include the word polymorph (when reported by the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

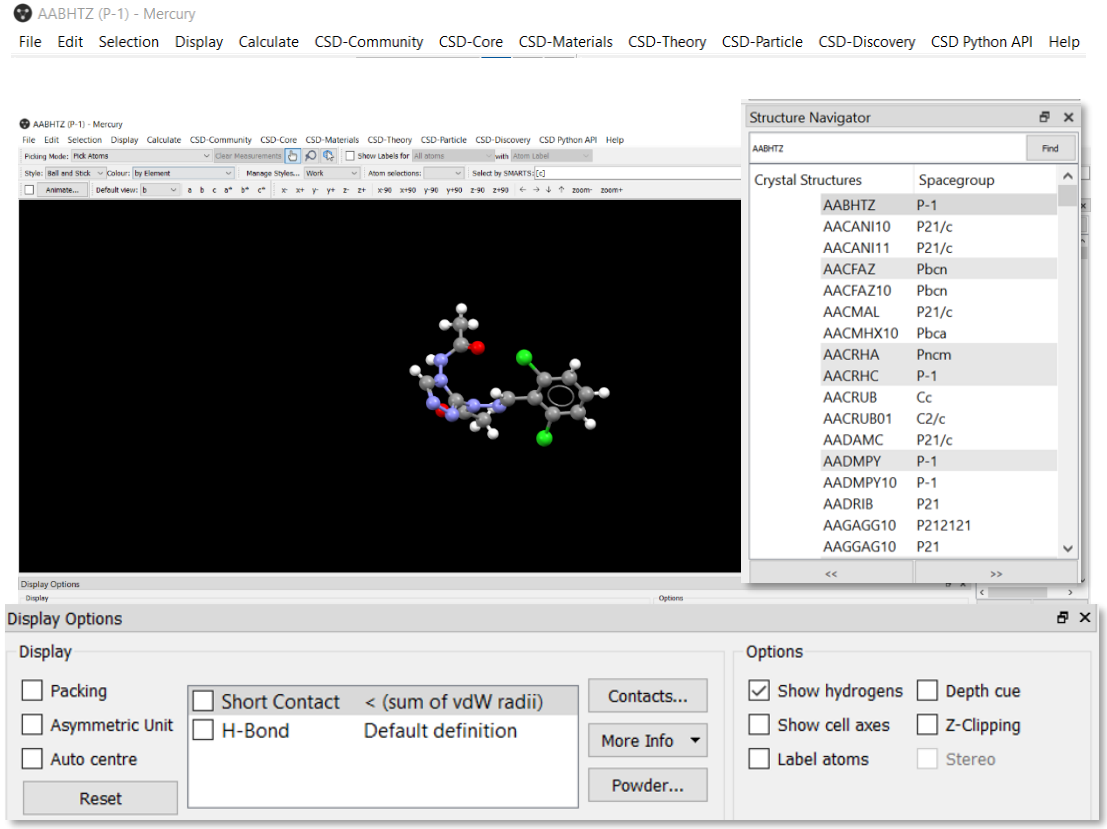
Basics of Mercury Visualization

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

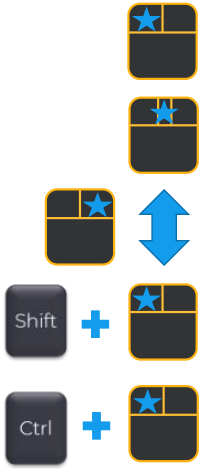
In the following we will see some of the basics of navigation and visualization in Mercury that you will find helpful to support your analysis.

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