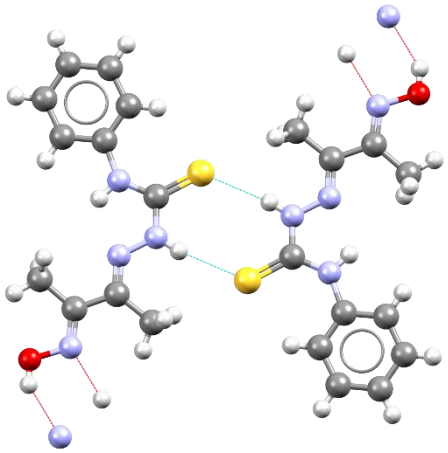
Motifs, Crystal Packing Feature, Crystal Packing Similarity

Developed using
2021.3 CSD Release



CCDC
advancing structural science

Table of Contents

Introduction

This workshop shows you how to analyse hydrogen-bonding interactions and packing trends in crystal structures using the features under the CSD-Materials > Search menu in Mercury. The motifs, packing feature and packing similarity searches will allow you to easily interpret packing trends by quick comparison with data available in the CSD.

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

Objectives

In this workshop you will learn how to:

- Use the Motifs feature and search for hydrogen bond motifs for a specified set of functional groups.
- Analyse unusual H-bond interaction using Crystal Packing feature.
- Use Crystal Packing Similarity tool to analyse structures that form solvates.

During the Virtual Workshop Try One part, work on one of the examples – you can pick the one on the components you are the most interested in. If you finish early, you may start a new example, or you can return to any of the examples at another time.

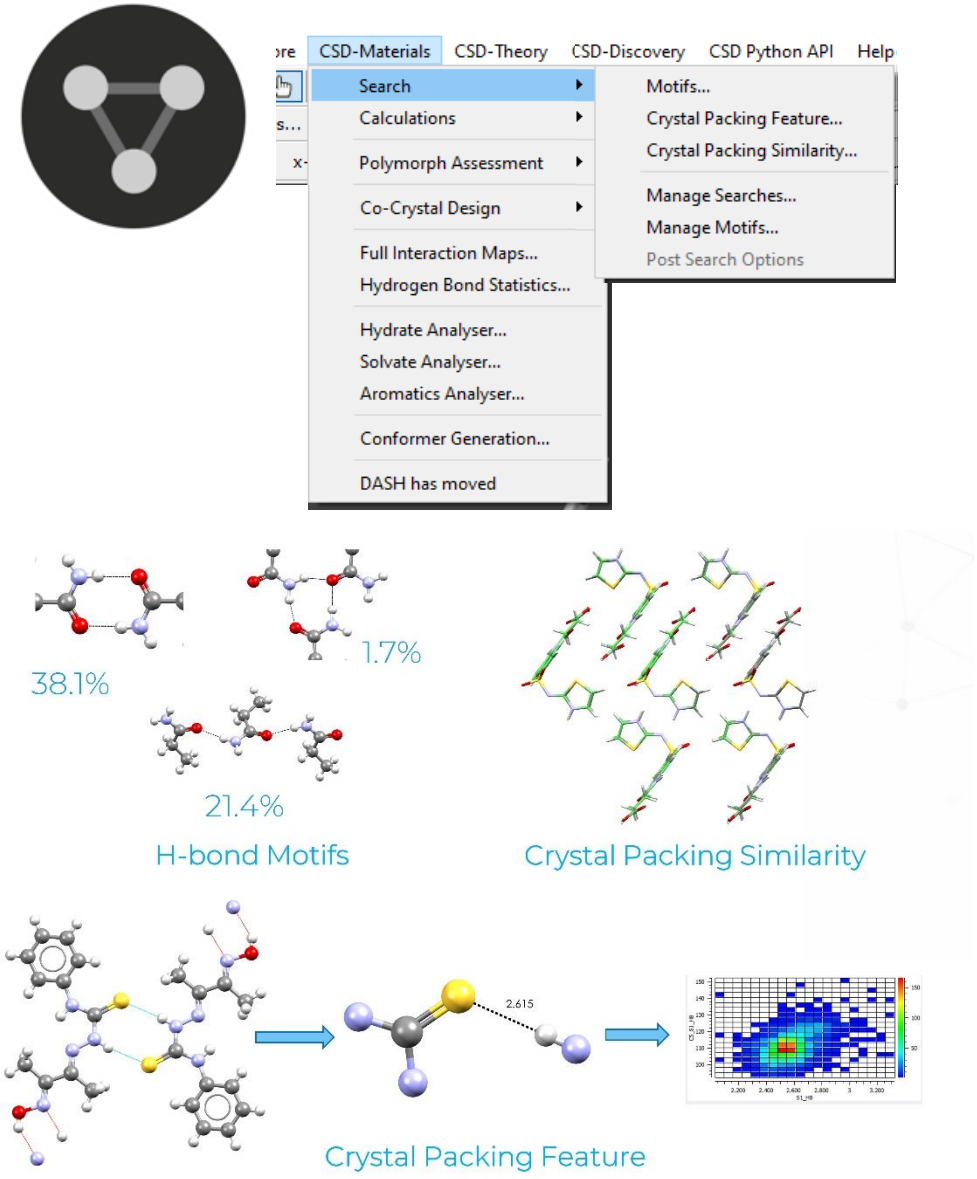
Note: The words in *Blue Italic* in the text are reported in the Glossary at the end of this handout.

Pre-required skills

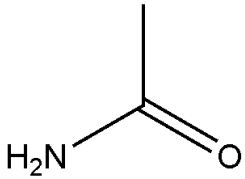
The following exercises assume you have a working knowledge of the basics of visualizing structures with Mercury, namely, how to display and manipulate structures from a 3D coordinates file.

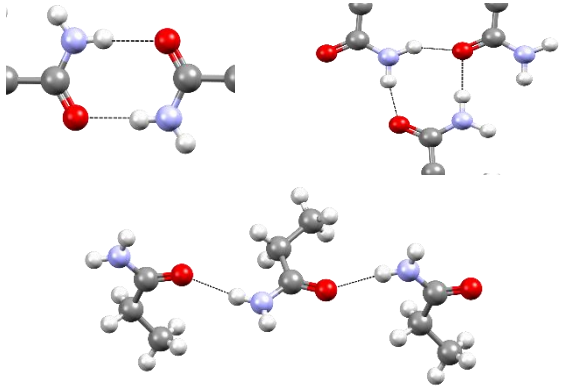
Materials

There are no additional materials required for this workshop.



Example 1. Motif searches for Primary Amides





The Motif searches functionality in CSD-materials determines the likelihood of *hydrogen bond motifs* for a specified set of functional groups. Using this functionality, you can assess different *motifs* by their relative frequency of occurrence in the CSD. The tool allows to search for auto-generated and bespoke motifs. Also, motif searches can be used to analyse the results of crystal structure prediction runs by identifying the range of predicted motifs
In this example we will use the Motifs functionality to search for Primary Amides and identify their frequencies occurrence in CSD.

Searching a pre-defined motif

CONH2 *R2,2(8)*

1. Open Mercury by double-clicking the Mercury icon on the desktop.

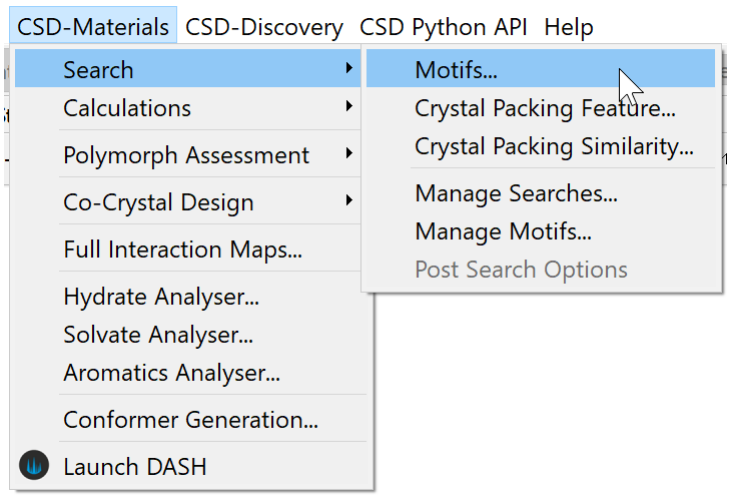
2. From the top-level menu select **CSD-Materials > Search > Motifs…** to launch the *Motif Search Wizard* dialog box.

3. In the *Motif Search Wizard* dialog box select the **Select pre-defined motif(s)** option to choose a pre-defined motif then click Next.

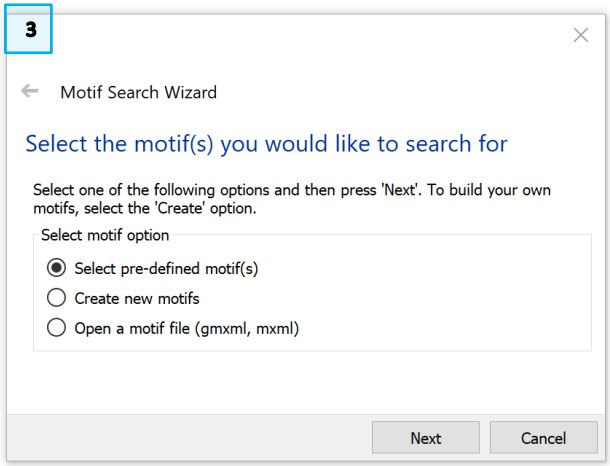
4. Click on ***homomeric*** to select the pre-defined motif.

1

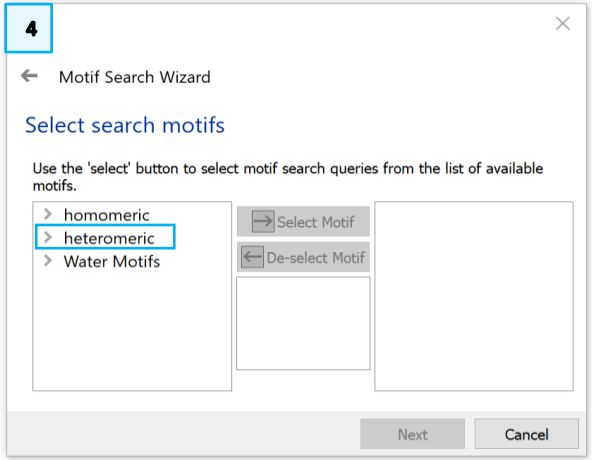
2



3



4



5. Under the **homomeric** list select the CONH2 R2,2(8), then click on Select Motif button.

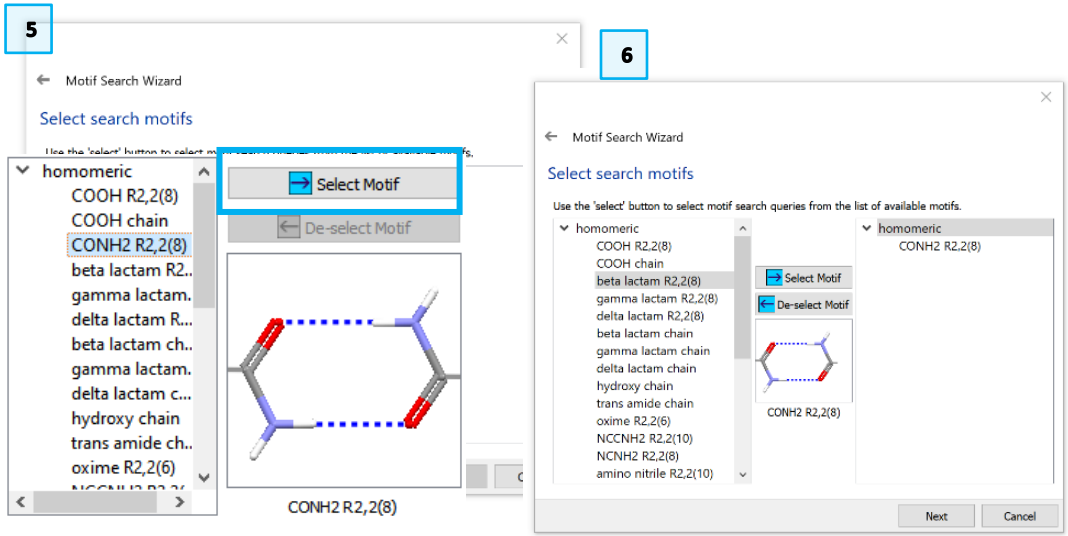
6. The CONH2 R2,2(8) motif should be listed on the right-hand side column. Select **Next** to continue.

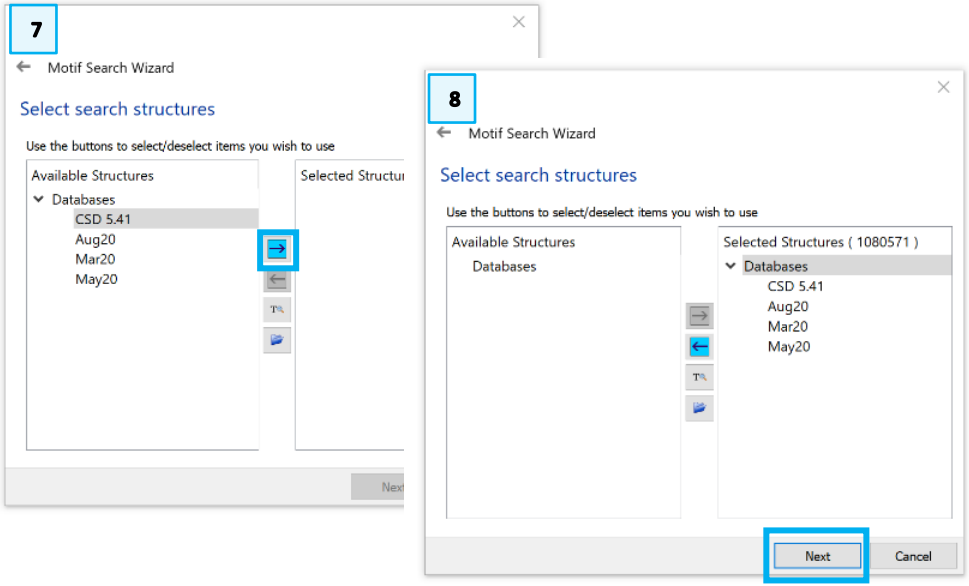
7. Select the *Databases* available and then click on the black arrow of the blue box button.

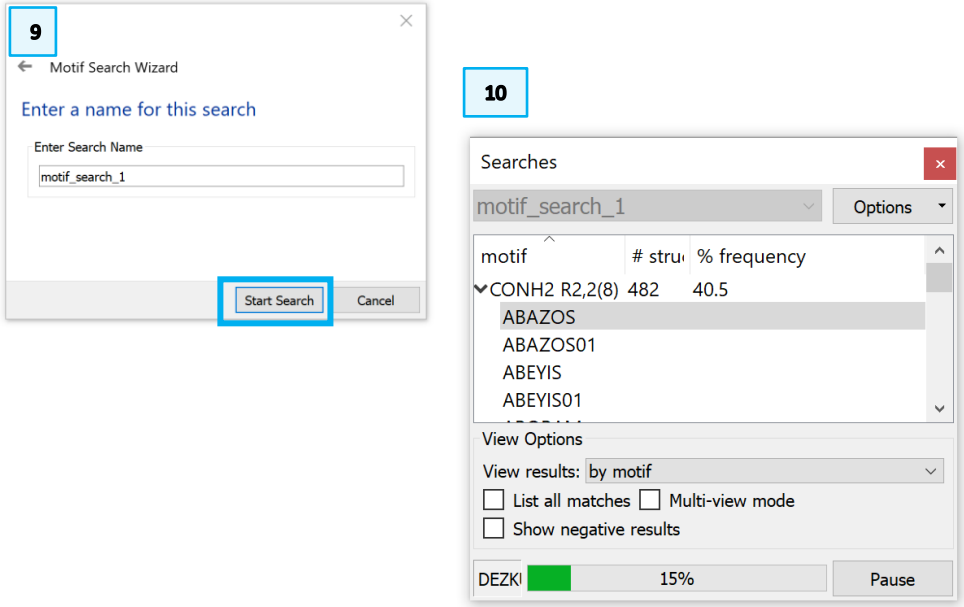
8. The selected Databases should appear on the right-hand side list. Select **Next** to continue.

9. You can enter a search name if you wish, for this example we will leave the default name as motif_search_1. To start the search, click the **Start Search** button.

10. The Searches have now started, and you can notice a progression bar in the bottom right corner. Wait for the bar progression to reach 100% in order to complete the search.







11. Once the search is completed all refcodes identified are listed in the window wizard. For the CONH2 R2,2(8) motif, at least 2905 structures were identified which represent an overall frequency of 38.1%.

12. Scroll down through the refcode list to analyse the type of molecules that contain the CONH2 R2,2(8) motif. We are illustrating here the presence of CONH2 R2,2(8) motif in o-Acetamidobenzamide, refcode ACBNZA.

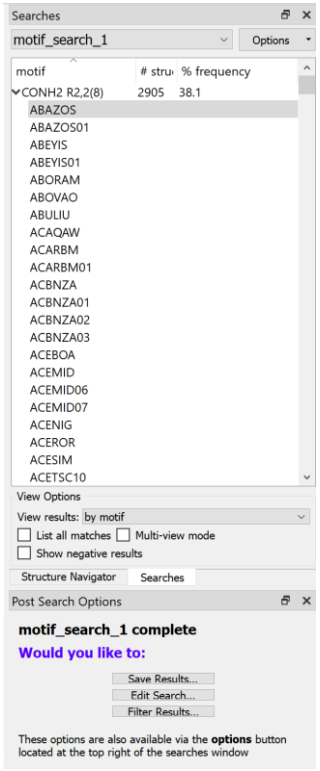
CONH2 *R2,3(8)* motif

13. Repeat **Steps 2 to 4** above and select the CONH2 R2,3(8) motif to perform a search for this type of interactions. Click on Select Motif button and the motif should be listed on the right-hand side column. Select **Next** to continue. In the next step ensure CSD is selected as in **Steps 8-9** above and select **Next** to continue.

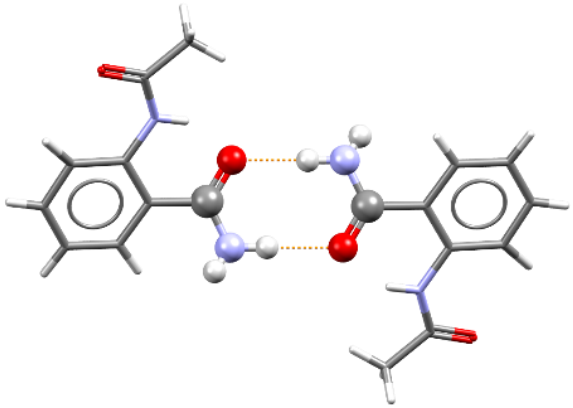
14. You can enter a search name if you wish, for this example we will leave the default name as motif_search_2. To start the search, click the **Start Search** button.

15. The Search has now started, and you can notice a progression bar in the bottom right corner. Wait for the bar progression to reach 100% in order to complete the search.

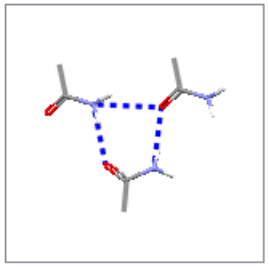
11

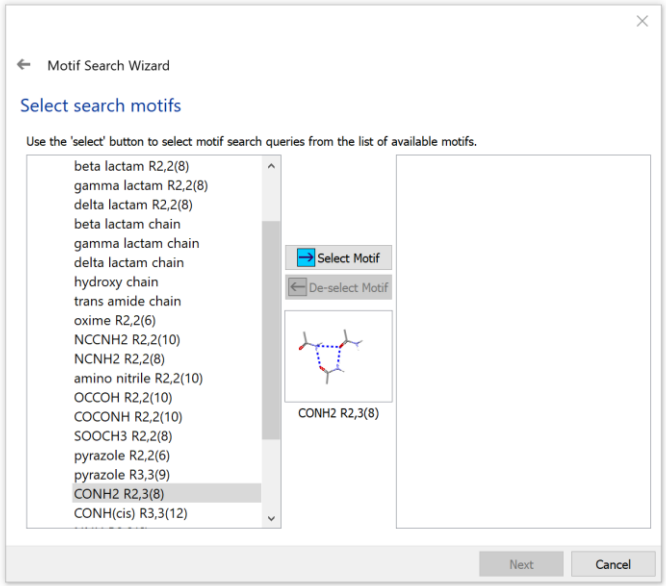


12

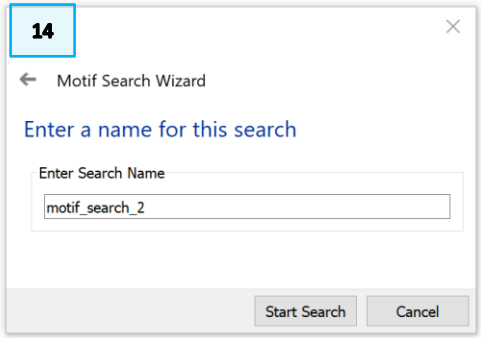


13

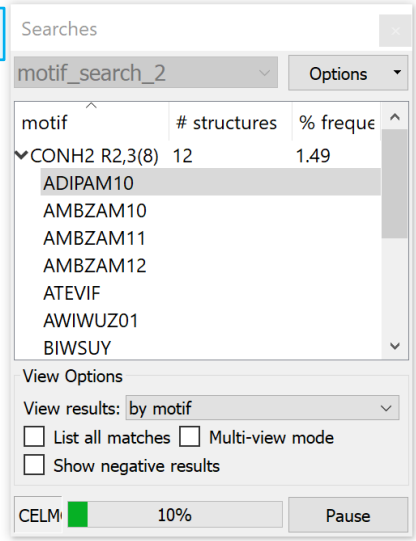




14



15



16. Once the search is completed, all refcodes identified are listed in the window wizard. For the CONH2 R2,3(8) motif at least 127 structures were identified which represent an overall frequency of 1.66%.

17. Scroll down through the refcode list to analyse the type of molecules that contain the CONH2 R2,2(8) motif. We are illustrating here the first refcode from the list ADIPAM10 which is the crystal structure for Adipamide.

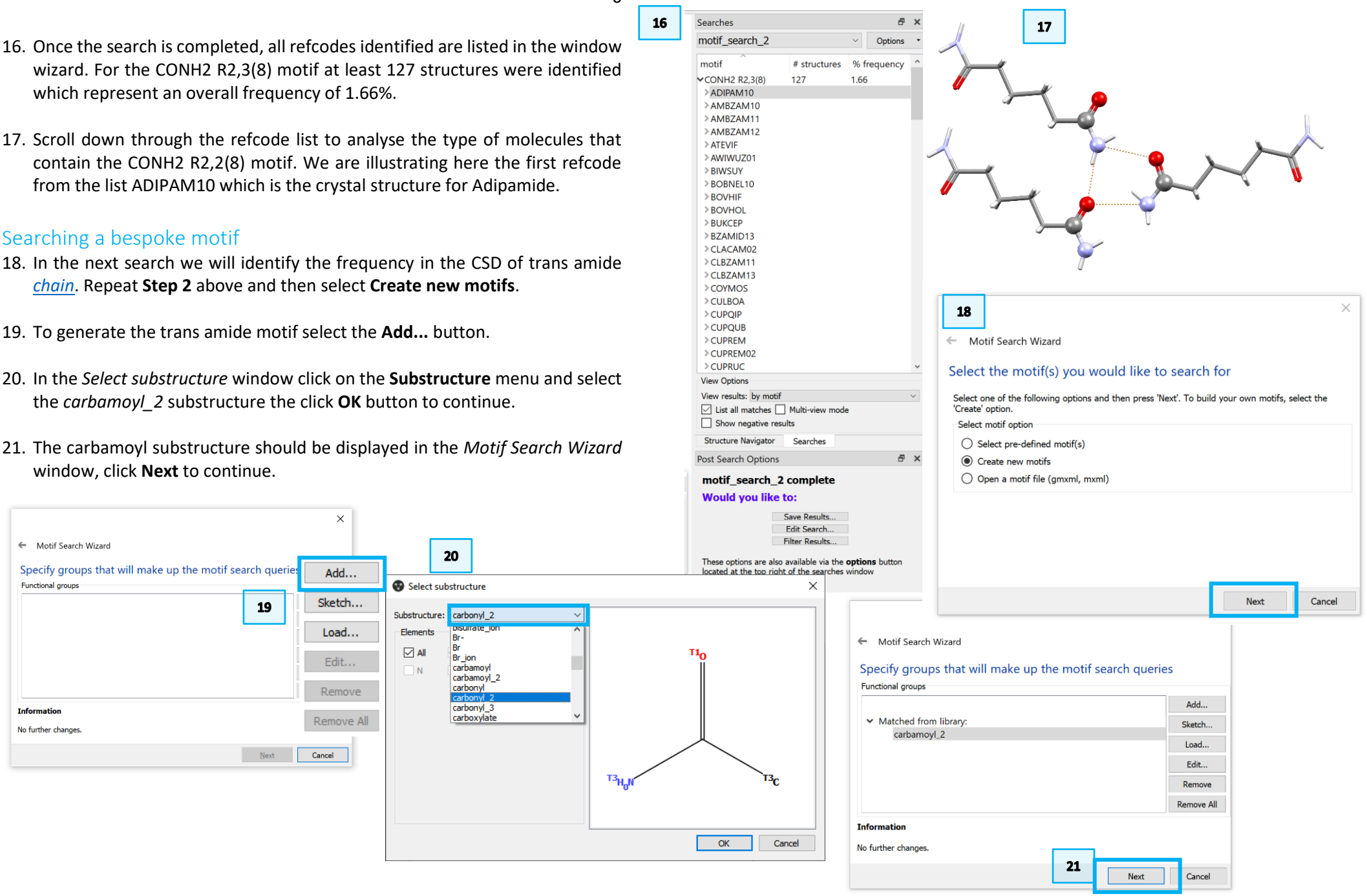
Searching a bespoke motif

18. In the next search we will identify the frequency in the CSD of trans amide *chain*. Repeat **Step 2** above and then select **Create new motifs**.

19. To generate the trans amide motif select the **Add...** button.

20. In the *Select substructure* window click on the **Substructure** menu and select the *carbamoyl_2* substructure the click **OK** button to continue.

21. The carbamoyl substructure should be displayed in the *Motif Search Wizard* window, click **Next** to continue.



22. Select **Add...** to select the contacts present in the motif.

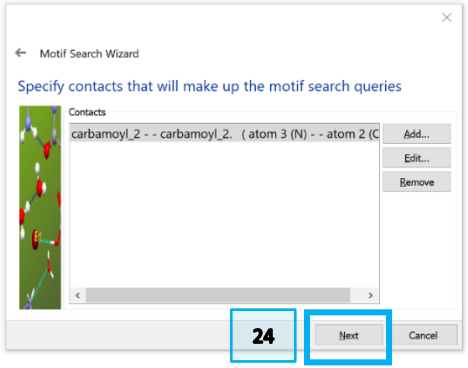
23. In the **From** column select *carbamoyl_2* in the Group drop down menu and N atom will be displayed in the Atom drop down menu. In the **To** column select *carbamoyl_2* in the Group drop down menu and O atom. The interaction will be illustrated in red in the display window. Click **OK** to continue.

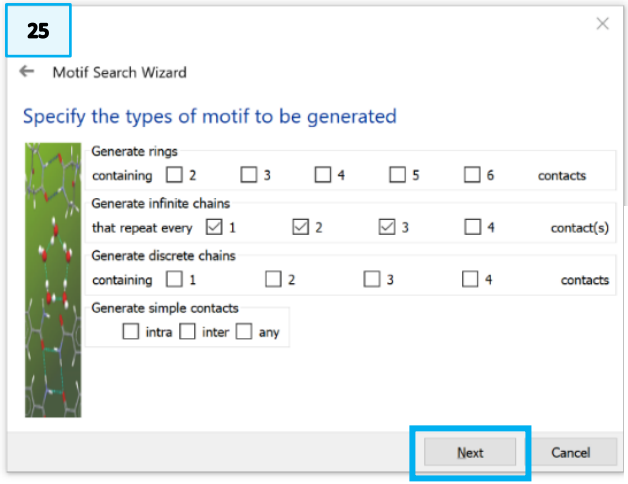
24. The specified contact will be illustrated in the wizard window. Select **Next** to continue.

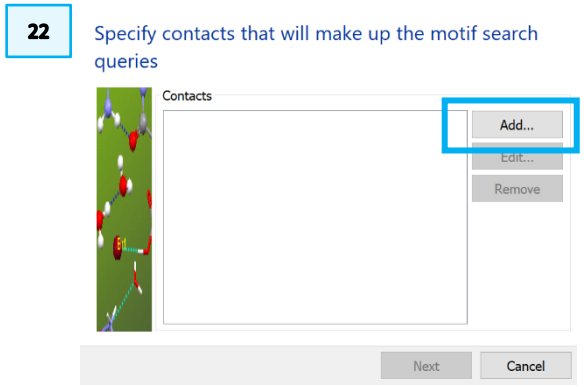
25. Select infinite chains that repeat every 1, 2 and 3 unit the select **Next** to continue.

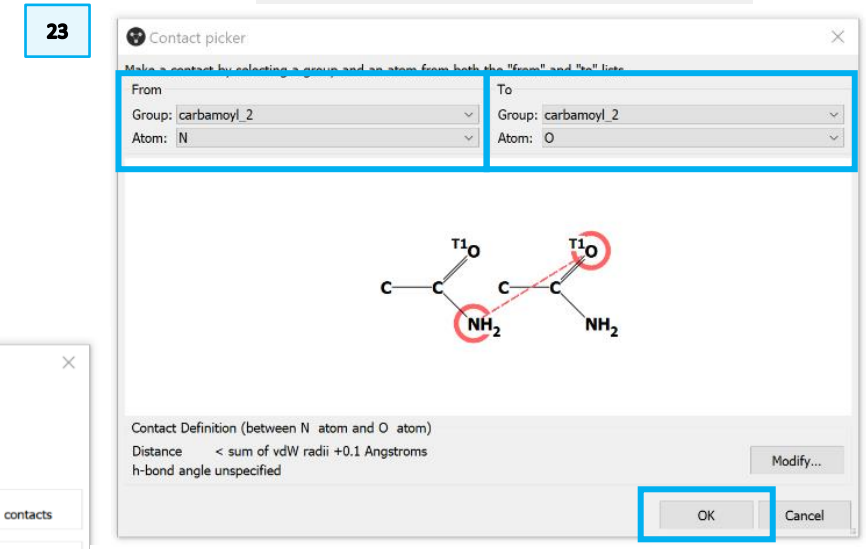
26. Note that 6 motif search queries were generated containing the chains to search for. Select **Next** to continue.

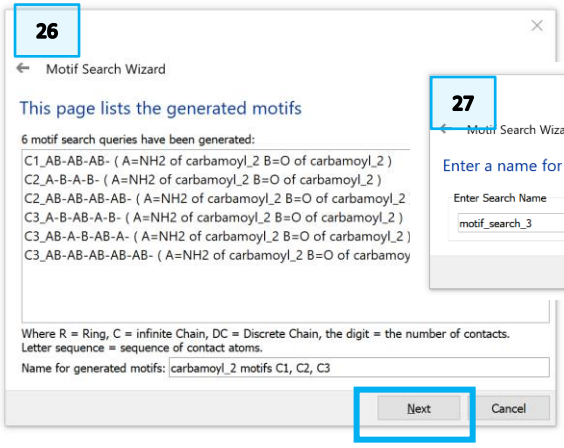
27. Ensure that the CSD is selected as shown in steps 7-8 above. You can enter a search name if you wish, for this example we will leave the default name as motif_search_3. To start the search, click the **Start Search** button.

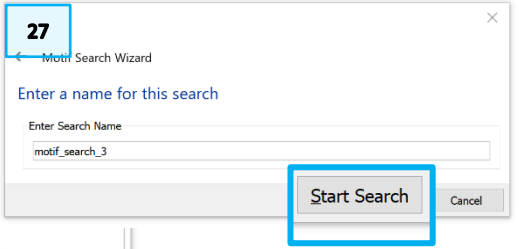












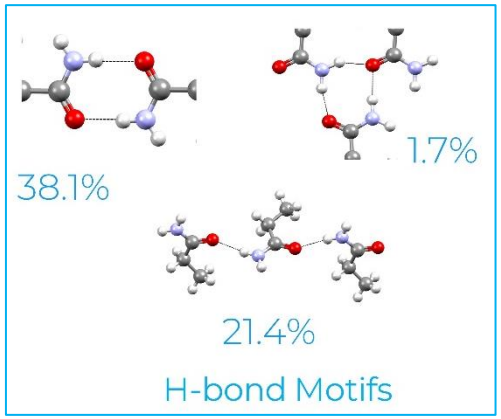
28. The Searches have now started, and you can notice a progression bar in the bottom right corner. Wait for the bar progression to reach 100% in order to complete the search.

29. Once the search is completed all refcodes identified are listed in the window wizard. For the trans amide chain motif at least 1008 structures were identified which represent an overall frequency of 21.4%.

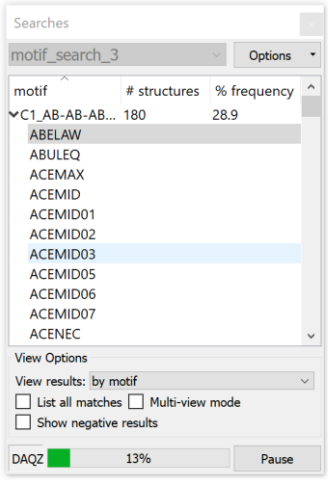
30. Scroll down through the refcode list to analyse the type of molecules that contain the trans amide chain motif. We are illustrating here the first refcode from the list ABELAW.

Conclusion

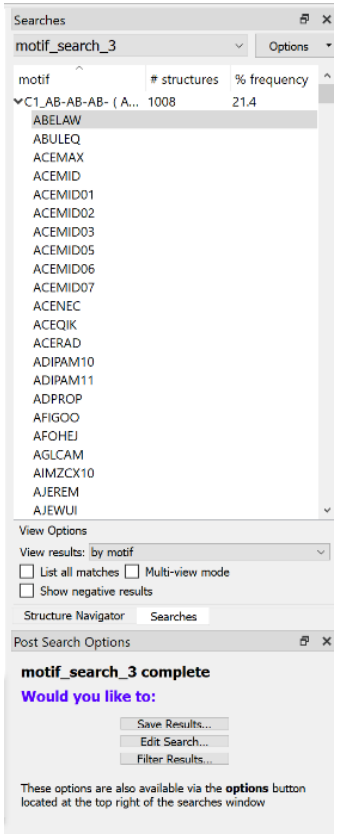
The Motif Search based on intermolecular interaction was used here to identify the abundance of different H-bond arrangements present in the CSD. We identified the frequencies of occurrence in CSD for primary amides and noted that the CONH2 R2,2(8) motif is the most commonly observed (38.1%, *as of the 2020.0 CSD release*). In this motif the amide utilises its donors and acceptors to form dimers. The trans amide forming a chain H-bonding motif is the second commonly observed (21.4%) and the CONH2 R2,3(8) is the least observed with only 127 structures found (1.7% frequency).



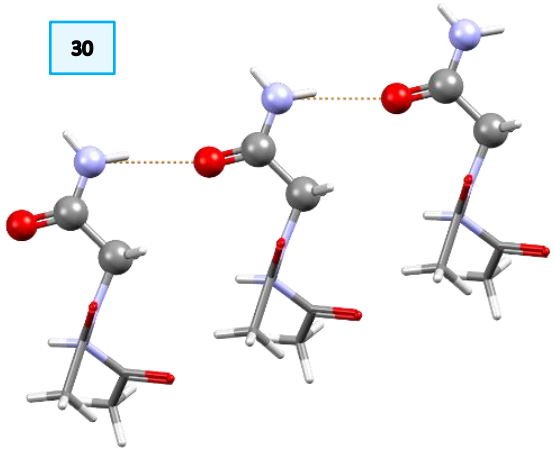
28



29

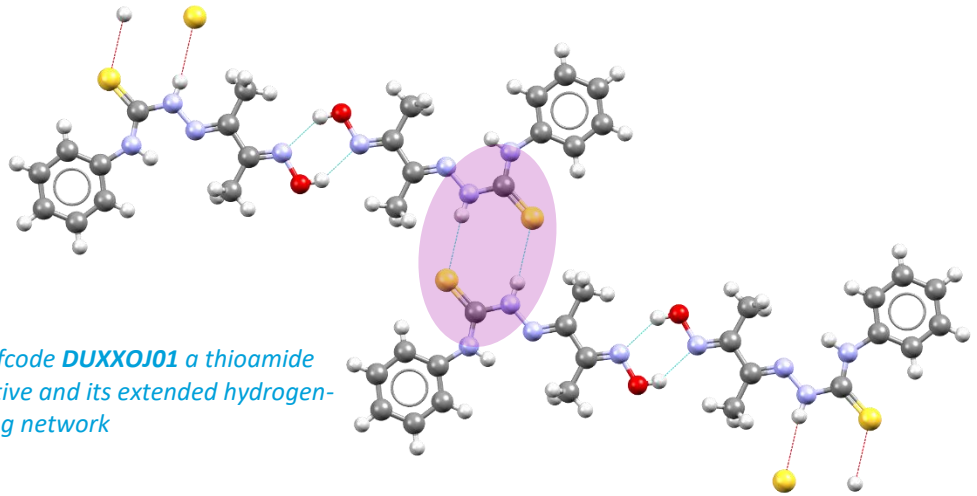


30



Example 2. Crystal Packing Feature - Investigating & analysing intermolecular interactions.

To use intermolecular interactions in crystal engineering, it is important to understand the relative geometric preferences of those interactions. The CSD is a valuable source of information on intermolecular interactions and their geometries.
This example uses a thioamide derivative to demonstrate how to determine if the C=S moiety is a genuine hydrogen-bond accepting group, and if so, what geometries the hydrogen bonds to the C=S group might adopt. The crystal packing feature search tool in CSD-Materials allows users to search the CSD for a specific molecular feature to determine the most preferred geometries.



*CSD refcode **DUXXOJ01** a thioamide derivative and its extended hydrogen-bonding network*

1. Open Mercury by double-clicking the Mercury icon on the desktop.

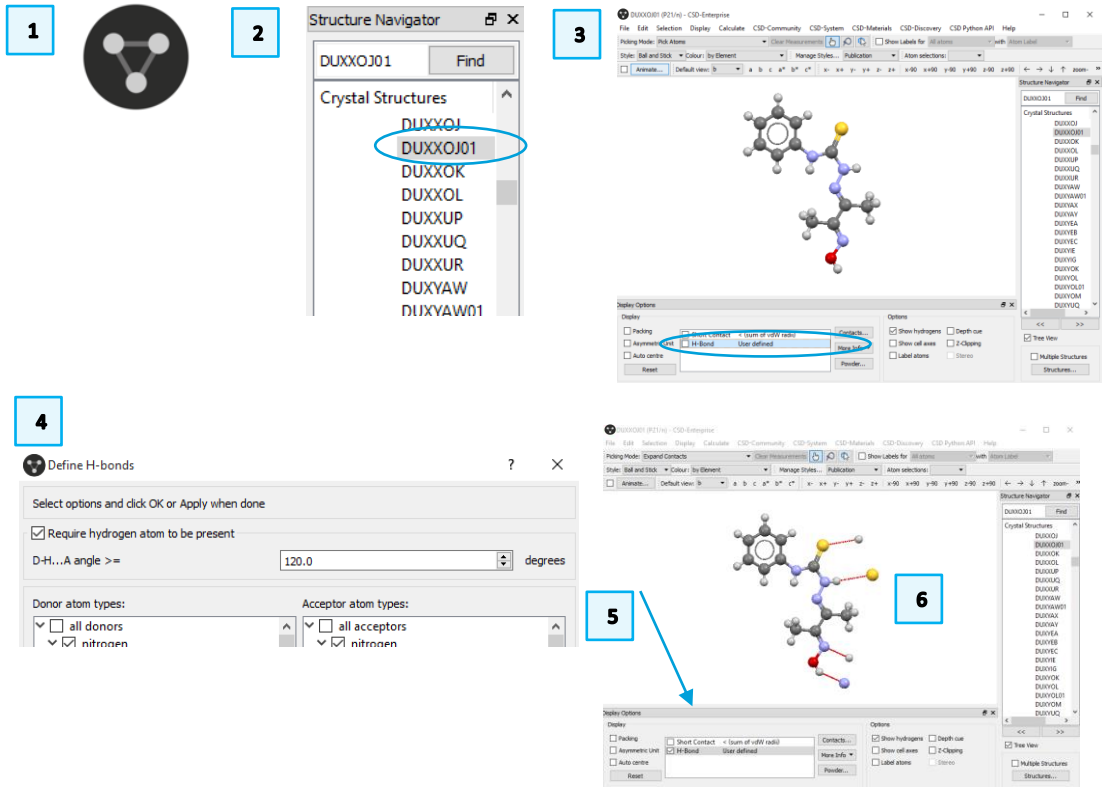
2. In the **Structure Navigator** toolbar type the DUXXOJ01 refcode.

3. The structure will appear in the Mercury interface. Edit the H-bond definition by double-clicking on the *H-Bond* line in the Display Options toolbar.

4. This will launch the *Define H-bonds* dialogue. Tick the box next to “Require hydrogen atom to be present”. Click **OK** to close the dialog box and return to the Mercury interface.

5. Tick the box next to “H-Bond” in the Display Options toolbar to turn on hydrogen-bond displays. This will show hydrogen bonding interactions as dotted red lines in the display

6. Expand the hydrogen bonds around the C=S group by clicking on the atoms at the end of the dashed lines.



7. You should now see two molecules linked by the hydrogen bonds between C=S groups. This is the interaction that we will be investigating.

8. Click to select the atoms forming the interaction of interest in the structure, *i.e.* the thiourea – N-H donor group. Selected atoms will be highlighted in yellow. To de-select an atom, simply click it again.

9. Open the Crystal Packing Feature Wizard, by clicking *CSD-Materials* from the top menu, then *Search*, then *Crystal Packing Feature…* The wizard will tell you there are 6 atoms selected from 2 molecules of DUXXOJ01. Click **Next** to continue.

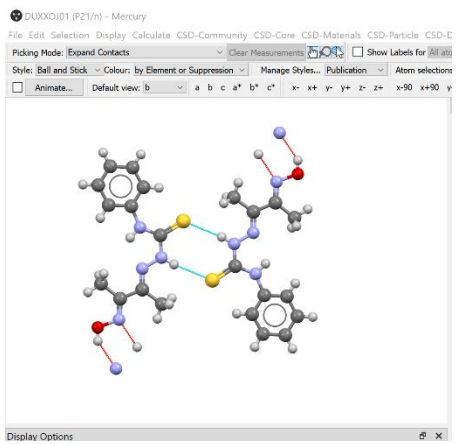
10. The Mercury display will now only show the atoms you selected; these may be hidden behind the dialogue box. In the top-level tool bar of the Mercury interface, tick the box to "Show Labels for All atoms" to label the atoms in the display. This will help in modifying atom properties in the following steps.

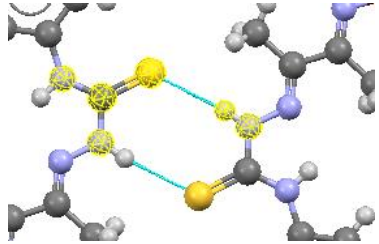
11. The wizard now allows you to define variable atom or bond types for the fragment interaction you are searching. Click to select the two nitrogen atoms bonded to the central carbon, N3 and N4 in the Mercury display. Then from the dropdown menu choose **Modify > Element > More > C or N**. This allows us to search for structures that have similar geometries but not necessarily the exact same elements.

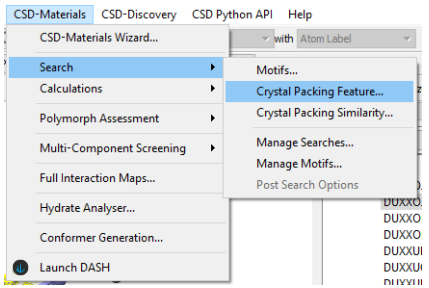
12. Click to de-select the two nitrogen atoms from the previous step. Click to select N3 in the donor N-H fragment and then choose **Modify > Element > Other Elements…** and click **N** and **O** from the periodic table. Then click **OK**.

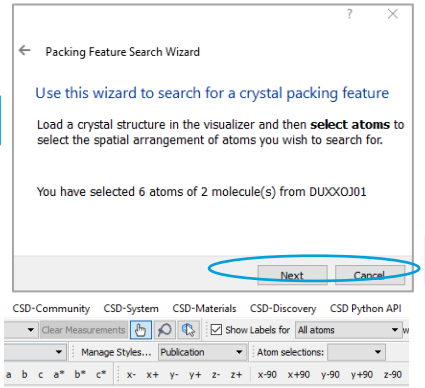
13. We also want to allow the number of hydrogen atoms and bonded atoms to vary on all three nitrogen atoms. To do this, click to select all three nitrogen atoms in the Mercury display. Choose **Modify > Hydrogens > Unspecified** and then **Modify > Number of bonded atoms > Unspecified** from the drop-down menu in the wizard.

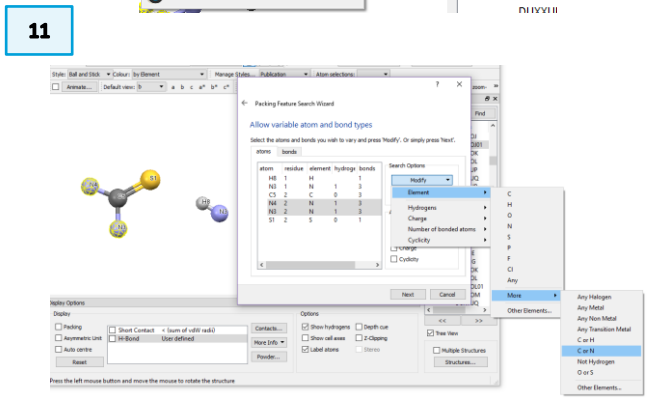
14. Click **Next** in the wizard to continue.

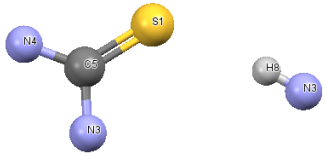


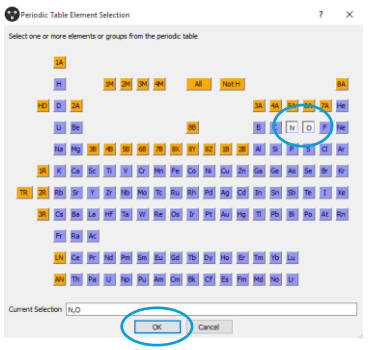


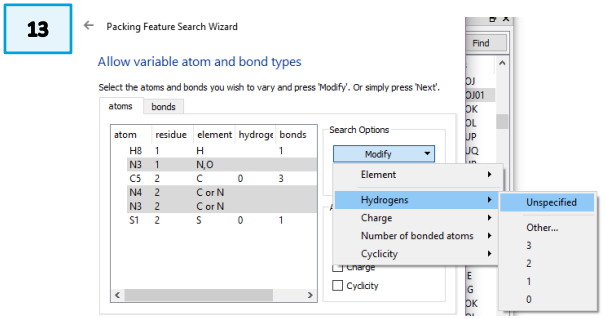












15. The next step of the wizard allows you to change the tolerances or customise search parameters. If your search fragment is very common, you may want to reduce the tolerance level. However, for this example, we will use the defaults. Click **Next** to continue.

16. Now we need to specify the bond lengths and angles we want to investigate for this fragment. We want to define the H···S distance and the N-H···S and H···S=C angles.

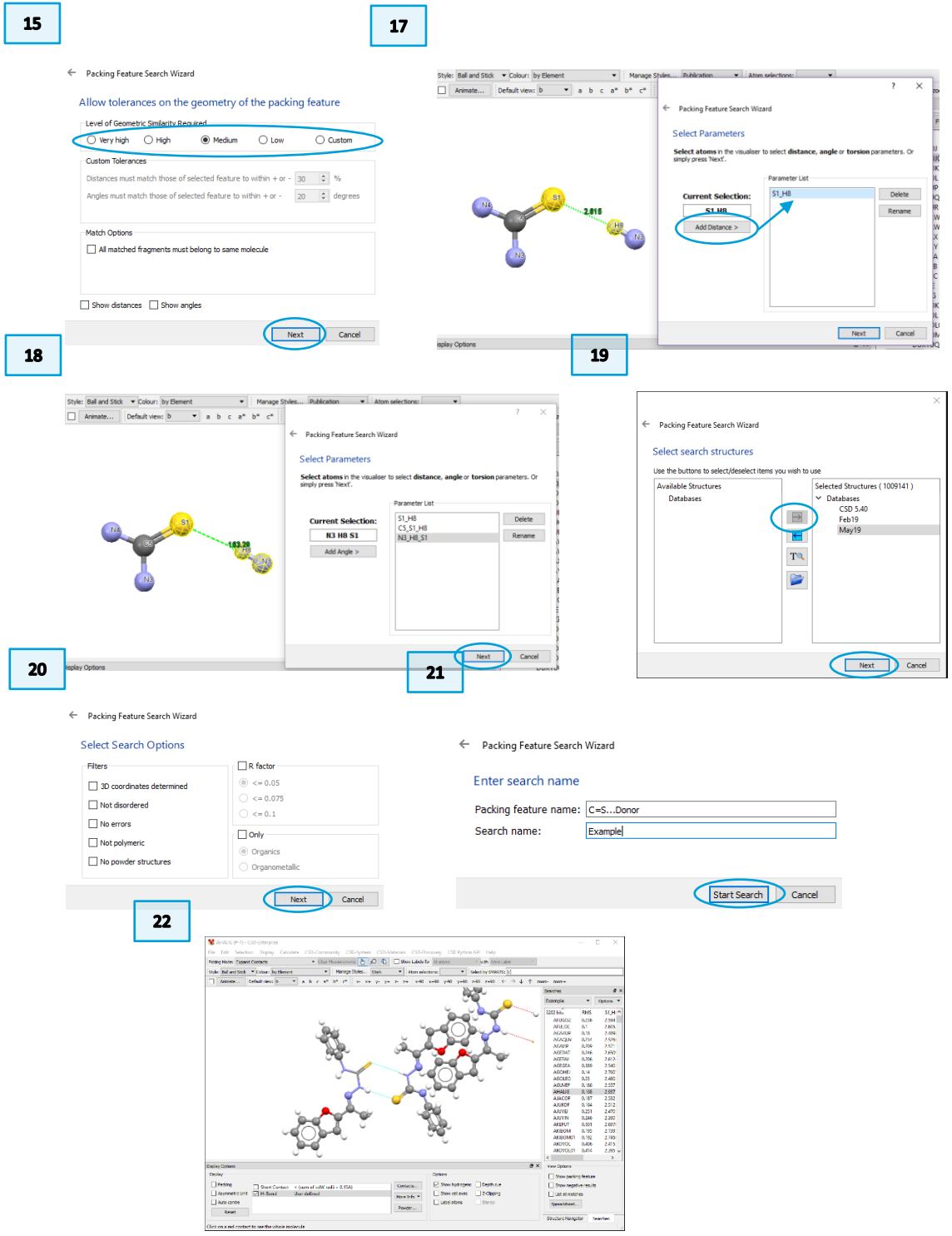
17. Make sure all atoms are de-selected by clicking in the background of the Mercury display. Click to select the S atom, and then the H atom then click the button **Add Distance >** to add this contact distance to the query.

18. Click, in order, the C atom, the S atom and the H atom, and then **Add Angle>** to add this parameter to the search. Likewise, click, in order, the N atom, the H atom and the S atom, and then **Add Angle >** to add this parameter as well. Click **Next** to continue.

19. The next step allows you to choose which data to search. For this example, use all available data. To do this, click the headline **Databases** at the top of the list and then the right arrow, to move all data to the *Selected Structures* box. If you have your own in-house databases, you can add them to the search in this way. Click **Next** to continue.

20. The next window allows you to apply filters. These can be used to target specific structure types (e.g. only organic molecules) or to improve the structure quality of our hits (e.g. no disorder.) We will use the defaults (*i.e.* no filters) for this example. Click **Next** to continue.

21. Choose an appropriate *Packing feature name* and *Search name* and enter them in the boxes. Click **Start Search** to continue.



22. When the search is done, click **Spreadsheet…** at the bottom of the Search toolbar window to open the **Data Analysis** window for analysing the results.

23. All the parameters selected during the set-up procedure will be listed in the Data Analysis spreadsheet.

24. To plot the data, select the *Plots* menu, and the type of plot you wish to use. For this example, we will start with a heat plot. Choose **Plots > Heat plot**.

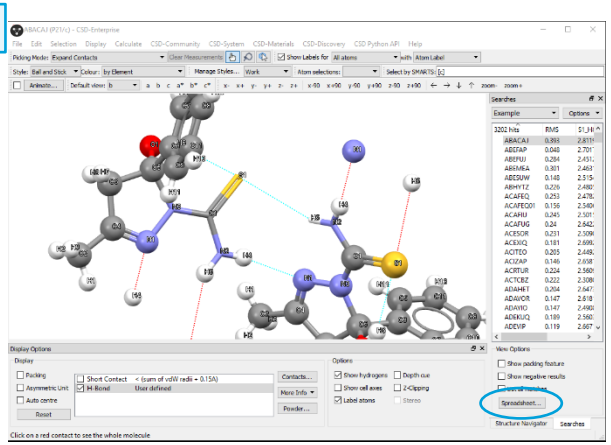
25. Choose the S···H distance for the X Axis and the C=S···H angle for the Y Axis. Click **OK** to close the dialog box.

26. The heat plot shows a strong geometric preference for a H···S contact distance of 2.5 to 2.6 Å and a H···S=C angle of 106° to 114°. This H···S contact distance is considerably less than the sum of the van der Waals radii of the atoms that are involved in the interactions (1.8 Å for S + 1.2 Å for H). This indicates a hydrogen bonding interaction.

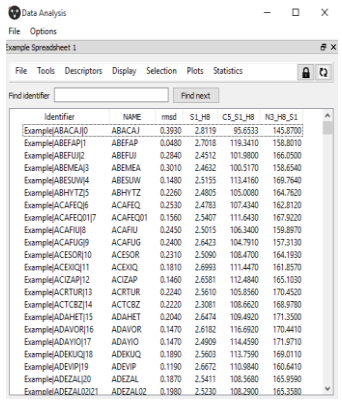
27. Return to the spreadsheet view by clicking the appropriate tab. Choose **Plots > Scatterplot**. Choose the S···H distance for the X axis and the N-H···S angle for the Y axis. Click **OK** to continue.

28. There is a dense cluster of structures with H···S contact distances in the region of 2.5 to 2.6 Å and more linear N/O-H···S bond angles in the region of 160° to 175° indicating that the C=S group is a common and good hydrogen-bond acceptor.

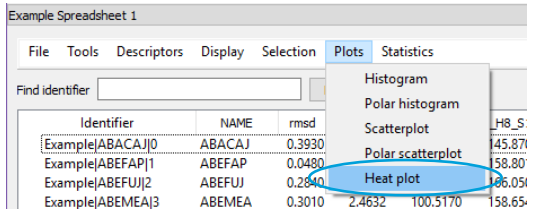
22



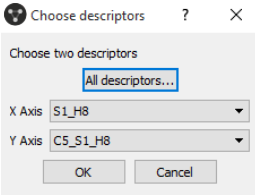
23



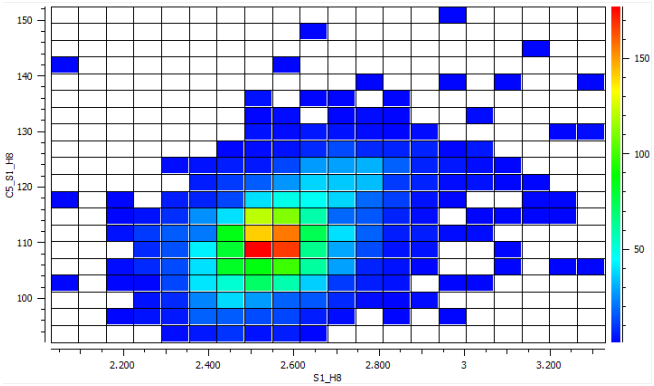
24



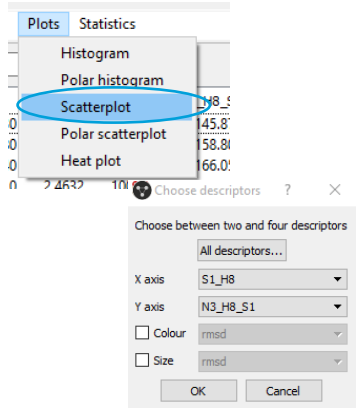
25



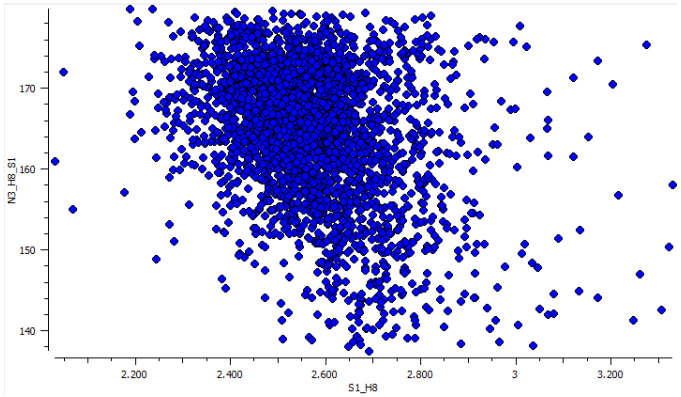
26



27



28



29. An individual structure in the cluster can be visualised in the Mercury display by clicking on a spot in the plot. The specific entry in the spreadsheet will also be highlighted.

30. Visual inspection of the structures in the densest part of the distribution shows that most acceptors have one or two nitrogen atoms covalently bonded to the C=S carbon.

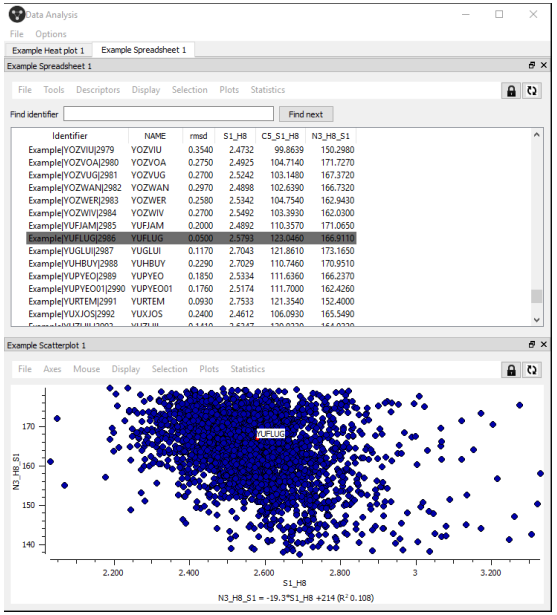
Conclusion

You should now be familiar with the *Crystal Packing Feature* search and analysing data through the *Data Analysis* tool.

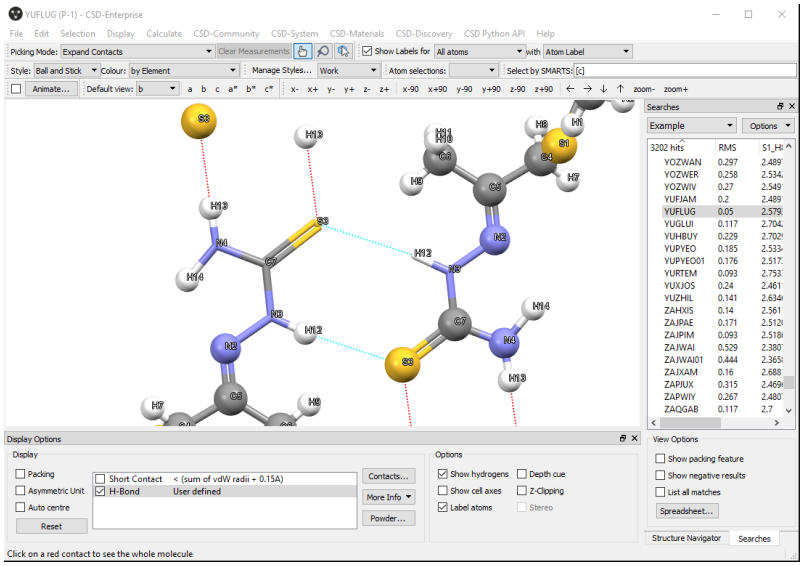
This example shows a high frequency of hits with a particular geometry. This clustering suggests that C=S groups can act as hydrogen-bond acceptors with clear directional preferences. Most observations of C=S as a hydrogen-bond acceptor are in systems where there are one or two nitrogen atoms covalently bonded to the carbon (thiourea or thioamide derivatives) which significantly enhances the accepting capability of the sulphur atom. The preferred interaction geometry for this moiety is an H···S distance of 2.5-2.6 Å and an H···S=C angle of approximately 110°. This is quite different to the equivalent geometries of C=O acceptors, which show distances of 1.8-1.9 Å and angles around 130°.

When running a packing feature search on a compound or moiety of interest, look for hits that cluster in specific distance and angle ranges. This suggests the most likely geometry a particular feature will adopt. This information can help guide development of co-crystals and other supramolecular assemblies.

29

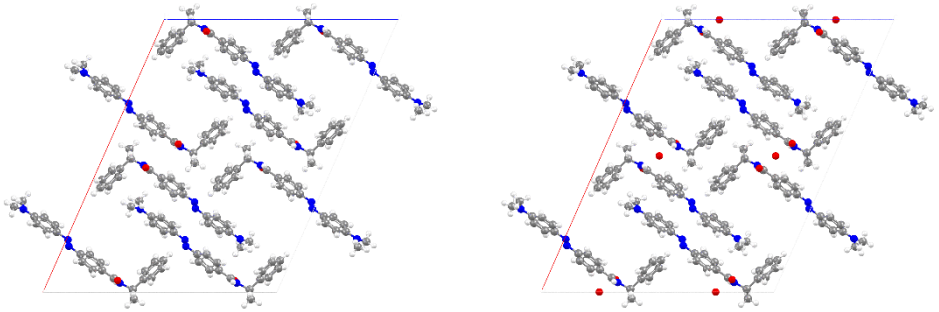


30



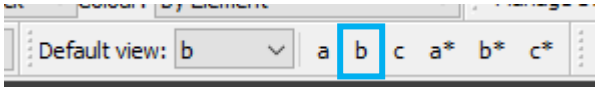
Example 3. Crystal Packing Similarity

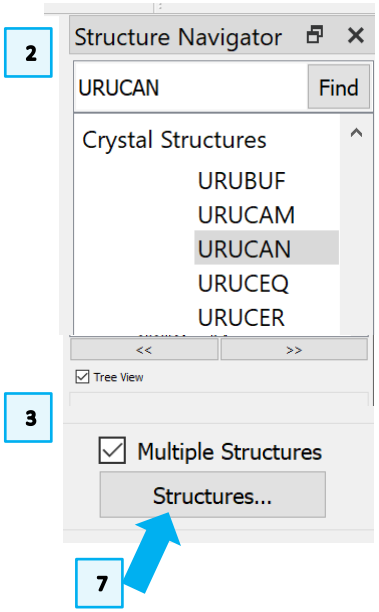
Crystal Packing Similarity functionality in CSD-Materials can be used to examine the packing similarities within a family of polymorphic structures, as well as anhydrous-hydrated and no solvated-solvated pairs. This example examines the crystal structure similarity between an anhydrous-hydrated pair for trans-rac-4-((4-(dimethylamino)phenyl)diazenyl)-N-(1-phenylethyl)benzamide. The similarity between the anhydrous-hydrated pair can give further insights about the hydration process and the ease of incorporating or removing water within the crystal structure of a material.

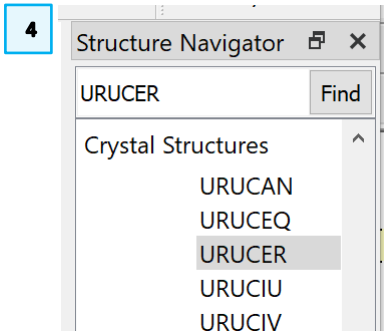


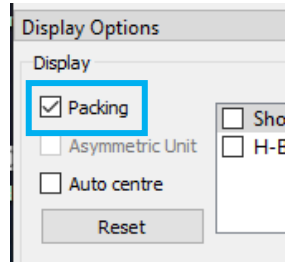
CSD Entry URUCAN (left) and URUCER (right), which are respectively anhydrous and hydrated forms of trans-rac-4-((4-(dimethylamino)phenyl)diazenyl)-N-(1-phenylethyl)benzamide

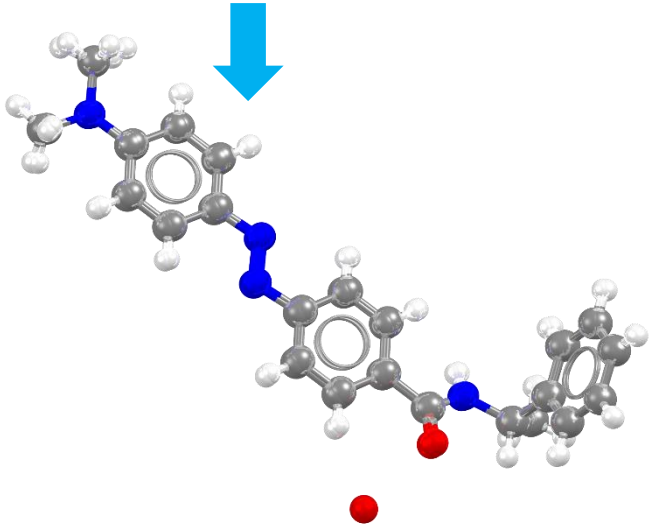
1. Open Mercury by double-clicking the Mercury icon on the desktop.
2. In the **Structure Navigator** toolbar type the refcode URUCAN.
3. Toggle on the *Multiple Structures* feature below the Structure Navigator.
4. Now, in the **Structure Navigator** type URUCER. You will now see in the 3D visualizer both CSD Entry URUCAN and CSD Entry URUCER. As this might however be a bit difficult to visualize, zoom in to appreciate the difference.
5. Toggle on the *Packing* in the **Display Options** dialog box below the 3D window. The packing of both structures is now shown.
6. To align the structures, select the view along the b axis from the top-level menu of Mercury.



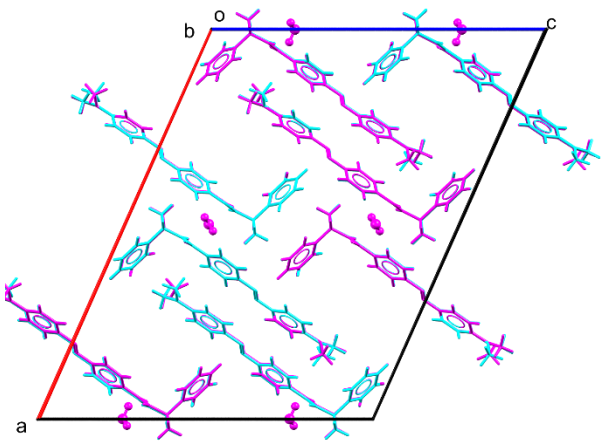








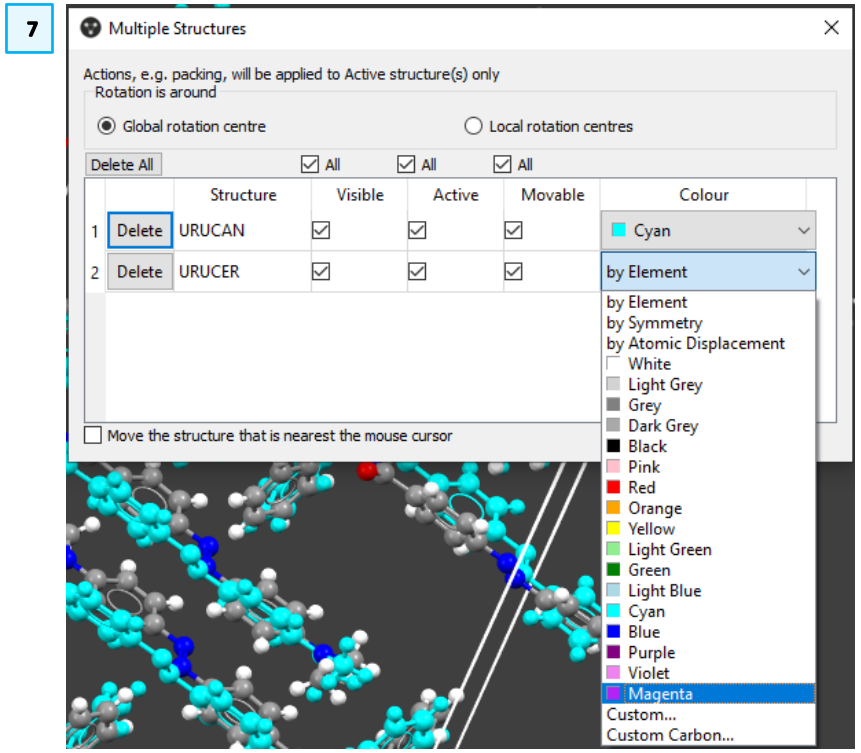
7. In the bottom right corner, just below the *Multiple Structures* tick box, click on **Structures…**. In the *Colour* column, select cyan for URACAN and magenta for URUCER, or any colour that will facilitate identifying the two structures for you. Observe: the packing is very similar between the anhydrous (URUCAN) and hydrated structures (URUCER).

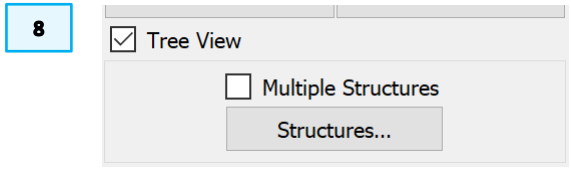


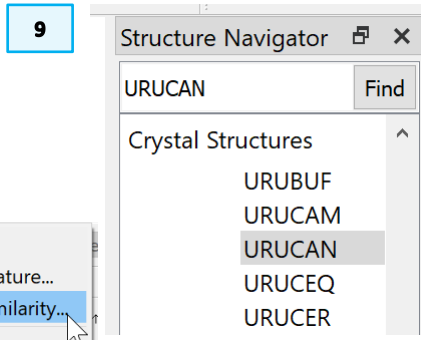
8. We will now move to analyse the packing similarity between these two structures. Toggle off *Multiple Structures*.

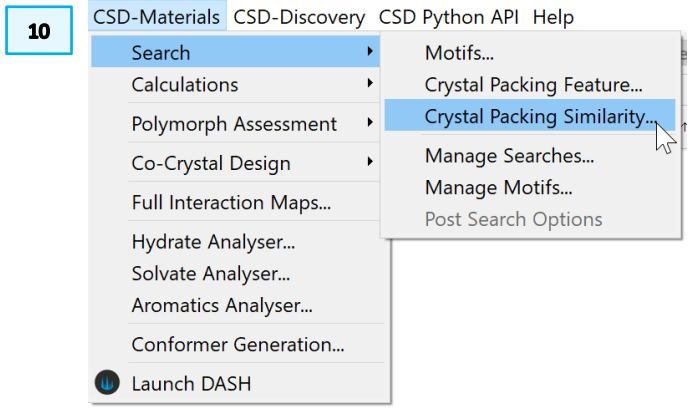
9. In the **Structure Navigator** type again the refcode URUCAN.

10. From the top-level menu in Mercury select **CSD-Materials > Search > Crystal Packing Similarity**. This will bring up the *Packing Similarity Wizard*.









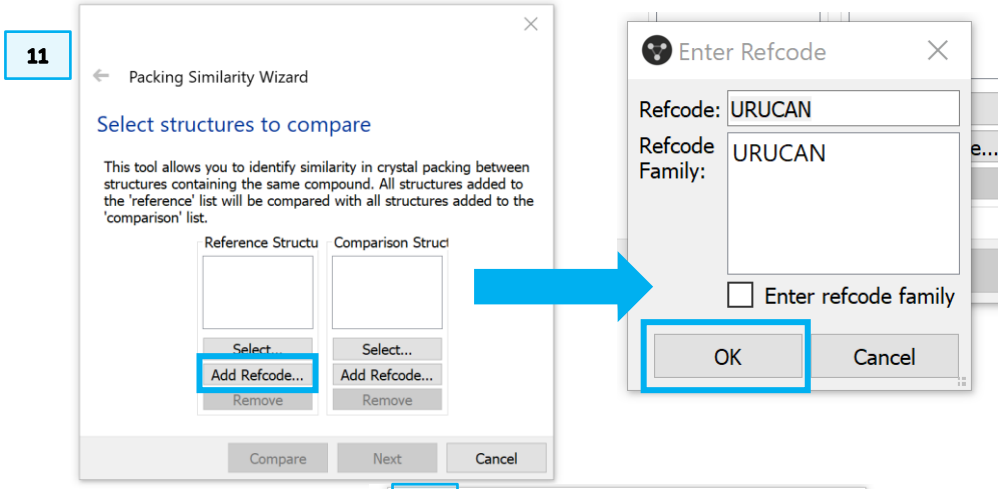
11. First, we need to add the structures to compare. In the *Packing Similarity Wizard*, under *Reference Structure* select **Add Refcode**. In the *Enter Refcode* pop-up type URUCAN, then click **OK**.

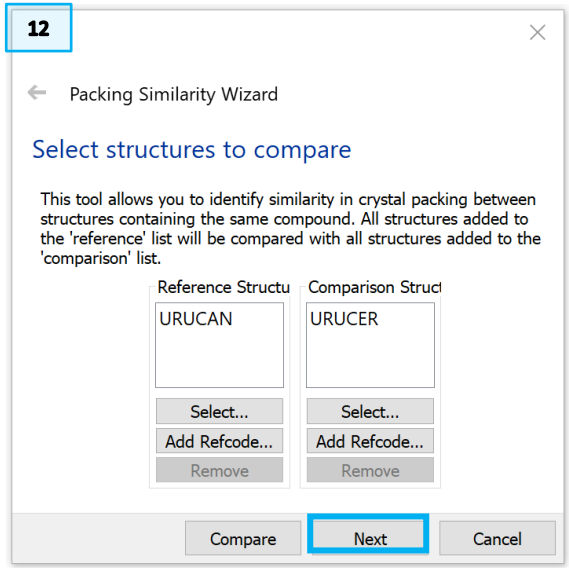
12. Repeat the same procedure as **Step 11** to add the comparison structure: under *Comparison Structure*, select **Add Refcode** and in the *Enter Refcode* pop-up type URUCER (you will need to overwrite the URUCAN refcode suggested), then click **OK.** Now click **Next** to continue with the settings.

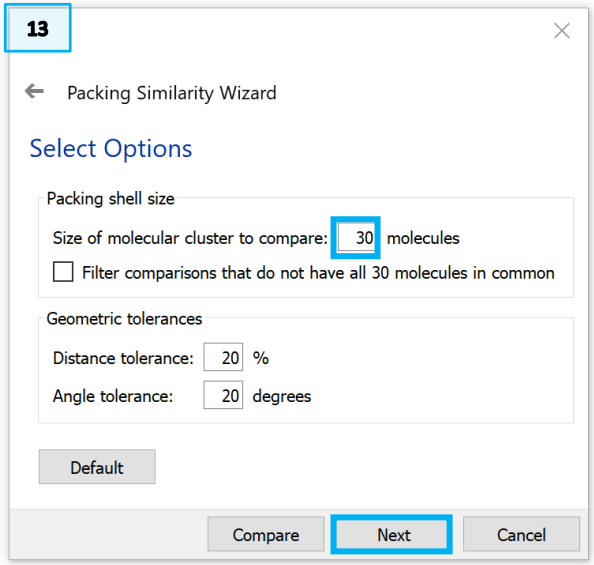
13. Under **Select Options** in the *Packing Similarity Wizard*, type 30 molecules as *Size of molecular cluster to compare*. Leave the other settings as default and click **Next**.

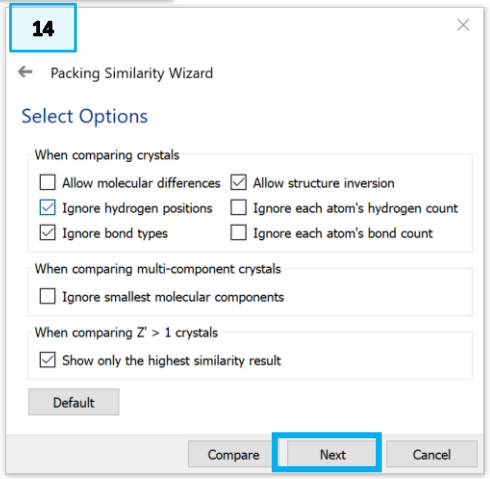
14. You have more options for your crystal packing similarity search. For this example, leave everything as default and click **Next**.

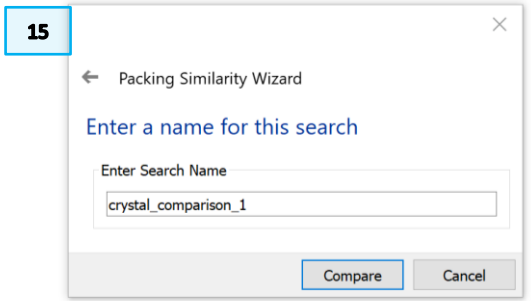
15. Note that the name of the search is given automatically as crystal_comparison_1, you can change this if you want. We will leave this as default for this exercise. Select Compare to begin the Crystal Packing Similarity Search.











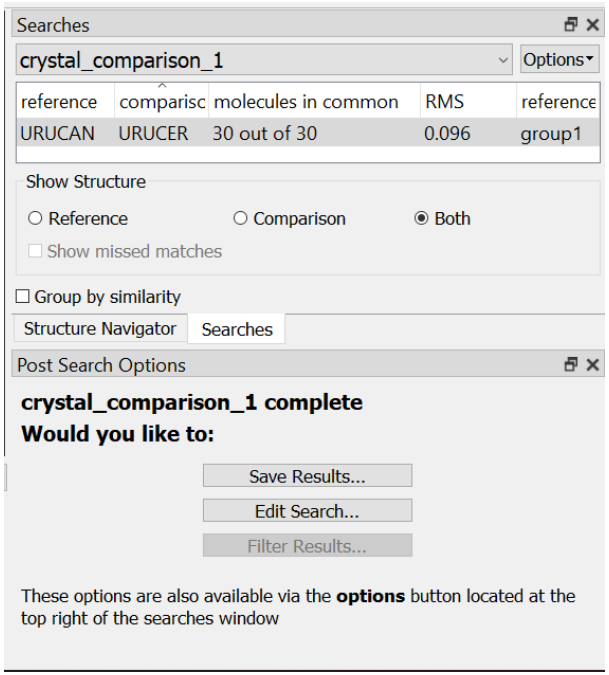
16. Once the search is completed, the outcome will appear in a table in the **Searches** toolbar on the right-hand side and the structures selected are displayed in the 3D visualizer. The table shows how many molecules the two structures have in common, in this case 30 out of 30, and the *RMS*, which in this example is 0.096, very low. The lower the RMS the more similar the structures are.

17. The 30 molecules compared are highlighted in the 3D display viewer of Mercury. You can toggle different options in the **Show Structure** box to visualise only the reference structure (*Reference* - carbon atoms in grey), only the comparison structure (*Comparison* – carbon atoms in green or red based on whether they match or not), or both overlapped. By default, only matching molecules are shown, but you can tick the *Show missed matches* box to display non-matching molecules. In this example it is greyed out as all molecules are a match.

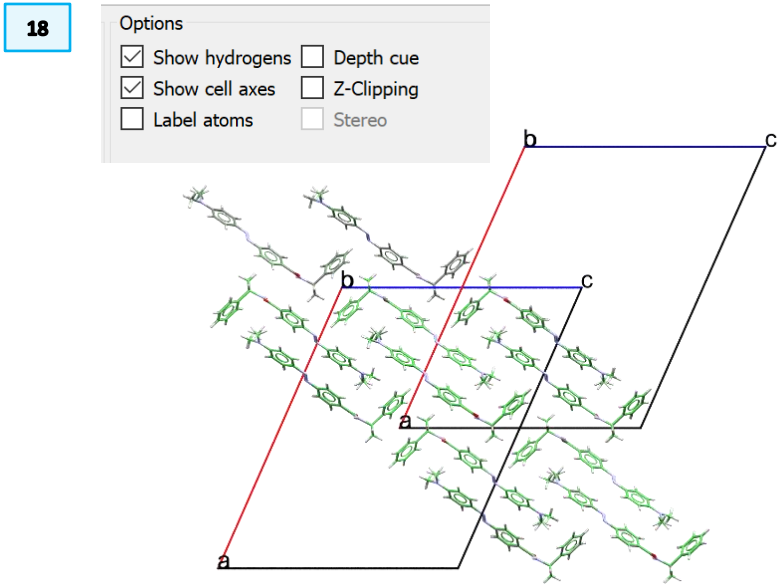
18. In the **Display Options** toolbar under the 3D visualizer, in the **Options** box toggle on the *Show cell axes* dialog box to see which molecules in the unit cell were compared.

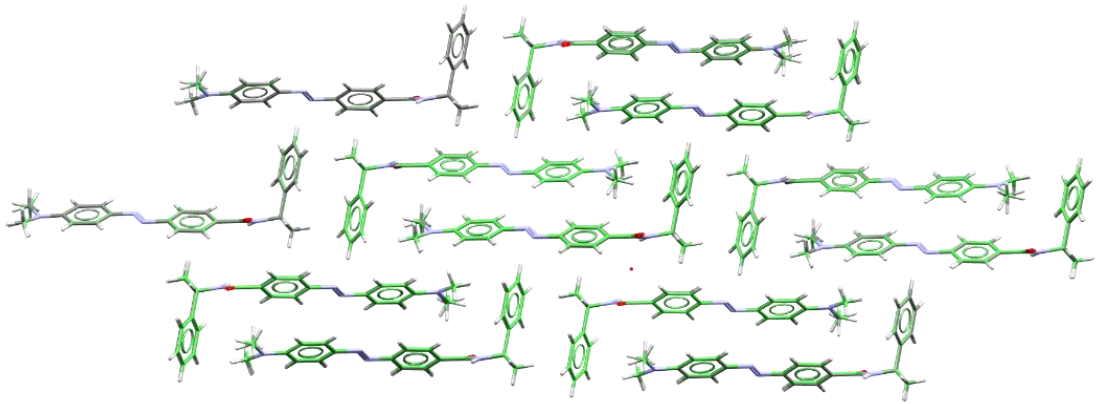
16

17



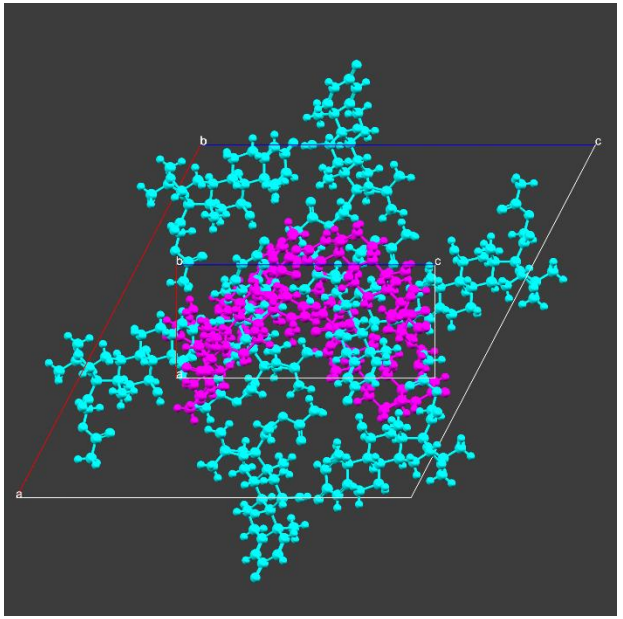
18





Conclusion

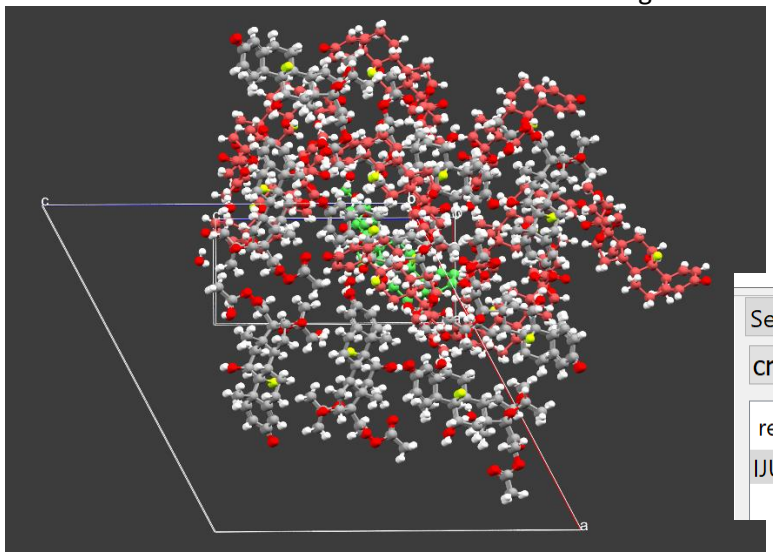
In this example we used the Crystal packing Similarity search tool to compare the anhydrous and hydrate forms of trans-rac-4-((4-(dimethylamino)phenyl)diazenyl)-N-(1-phenylethyl)benzamide. We found that for the cluster size analysed (30 molecules) the two forms have 30 out of 30 molecules in common and a low RSD. This data can inform further analysis of the structures and support further insights about the hydration process.



Bonus exercises

Perform a Crystal Packing Similarity study for the triamcinolone acetonide acetate pair, refcodes IJUTAJ02 (anhydrous) and QANVAF (monohydrate). What do you observe?

You will note that not all the anhydrated - hydrated pairs are similar. In this case only 1 out of 30 molecules are similar and the RMS is much higher 0.327.



Searches

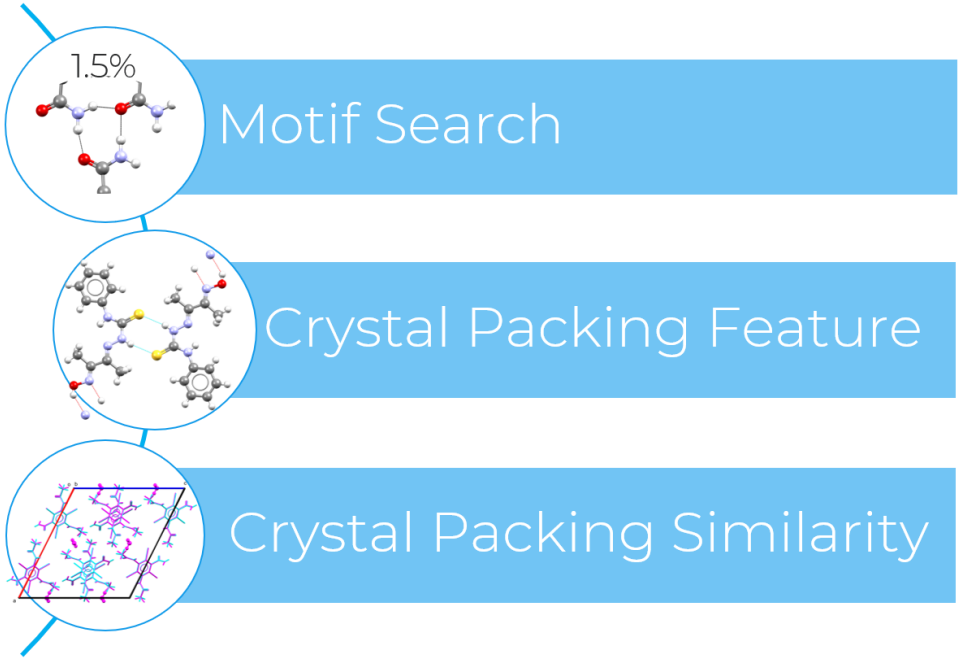
crystal_comparison_2 | Options

reference	compari	molecules in common	RMS	reference group
IJUTAJ02	QANVAF	1 out of 30	0.327	group1

Repeat the analysis for the VAWWEX - VAWWOH anhydrous-hydrated pair. What RMS do you notice? Is this pair very similar in packing?

Conclusions

- This workshop illustrated Searches that can be performed using CSD-Materials.
- You have learned how to identify the frequencies of occurrence of primary amides in different hydrogen bond network using Motif Search.
- You have learned how to apply Crystal Packing Feature to unusual Hydrogen Interactions such as S…H-N.
- You have learned how to apply Crystal Packing Similarity to anhydrous-hydrated pair.



Next steps

If you have completed one of the three exercises proposed and still have time, you can start any of the other two exercises or try these components on one of your own structures. Or ask the tutors for more examples! You could also check Tutorials 9 and 10 under Help > Tutorials top level menu in Mercury.

You may also explore more CSD-Materials components with our self-guided workshops:
https://www.ccdc.cam.ac.uk/Community/educationalresources/workshop-materials/csd-materials-workshops/

Reading suggestions

- If you want to see how Crystal Packing Similarity was applied to identify similar packing of an optoelectronic system with structures in CSD check out this paper: https://pubs.acs.org/doi/abs/10.1021/acs.cgd.1c00177
- To see how many similar anhydrous and hydrated pairs are in CSD check out this publication: https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/chem.201404693

Glossary

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

Examples from this handout:

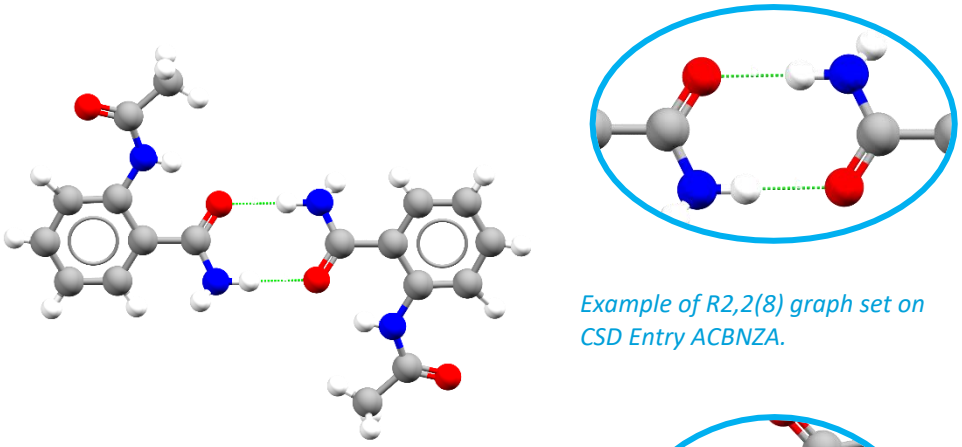
- **R2,2(8)** is a ring formed by 2 acceptors and 2 donors with 8 atoms involved in the ring.
- **R2,3(8)** is a ring formed by 2 acceptors and 3 donors with 8 atoms involved in the ring.
- **C1,1(4)** is a chain formed by 1 acceptor and 1 donor with 4 atoms involved in the ring.

Heteromeric vs Homomeric

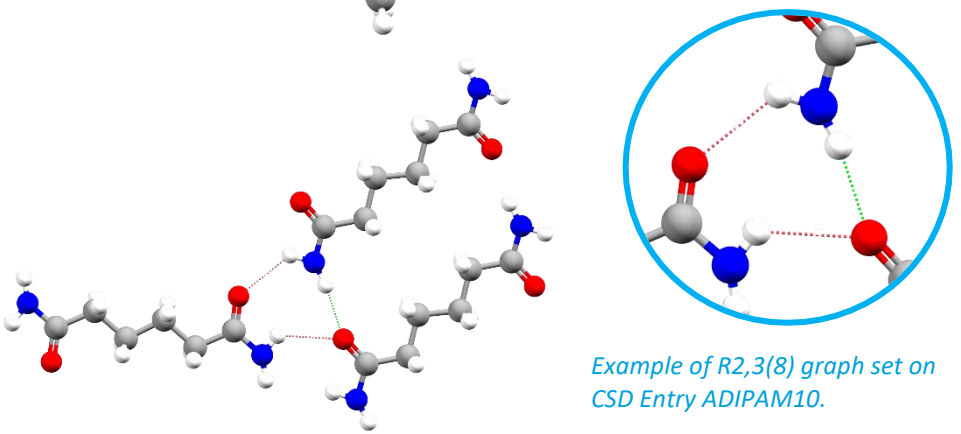
In the context of motif searching, homomeric refers to motifs that are created by the interaction and repetition of the same sub-unit or group. Heteromeric instead refers to when groups or sub-units are different.

Hydrogen Bonds

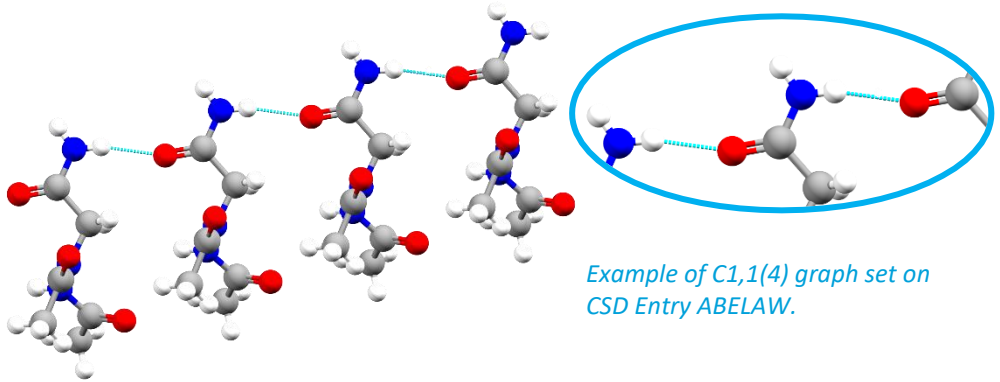
Hydrogen bonding occurs between donor-acceptor interactions precisely involving hydrogen atoms. The H-bonds interactions are classified as: strong (mostly covalent), moderate (mostly electrostatic) and weak (electrostatic). Their strength is observed to be between 12 and 30 kJ/mol.



Example of R2,2(8) graph set on CSD Entry ACBNZA.



Example of R2,3(8) graph set on CSD Entry ADIPAM10.



Example of C1,1(4) graph set on CSD Entry ABELAW.

Motif

A motif is a characteristic pattern of hydrogen bonds within a crystal structure. These interactions may be intermolecular, intramolecular or a combination of the two. Motifs are defined within Mercury as sets of interactions between functional groups with a particular pattern, e.g., rings, chains or discrete contacts.

Root Mean Square Deviation (RMSD)

The root mean square deviation (RMSD) is a commonly used measure of the difference between two sets of values (usually comparing observed data to estimated data). The RMSD is defined as the square root of the mean squared error. In Mercury this is used to measure the geometric difference between packing features or packing patterns in crystal structures.