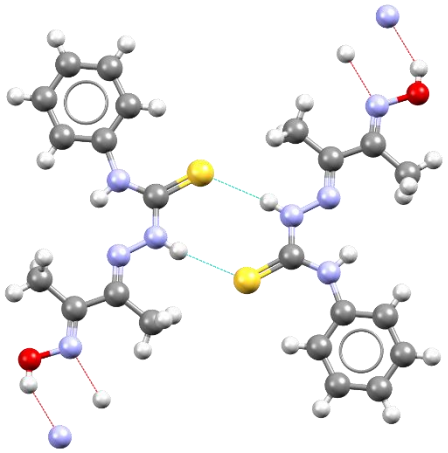
Motifs and Crystal Packing Feature Search in Mercury (MAT-006)

Developed using
2023.3 CSD Release



CCDC
advancing structural science

Table of Contents

Introduction

This workshop shows you how you can analyse hydrogen-bonding interactions and packing trends in crystal structures using the CSD-Materials Search functionality in Mercury. Mercury is the CSD software for visualizing crystal structures and acts as an interface to a multitude of functionality which allows you to further explore, amongst other things, crystal packing, molecular geometry and intermolecular interactions. In this workshop, we will focus on searching the CSD for common motifs and packing features using the Motifs and Crystal Packing Feature search funtionality from the CSD-Materials suite.We will explore both pre-defined, common arrangements as well as creating our own custom ones.

Learning Outcomes

After completing this workshop, you will be able to:

- Use the Motifs feature to search for hydrogen bond motifs for a specified set of functional groups.
- Analyse unusual H-bond interactions using the Crystal Packing Feature.

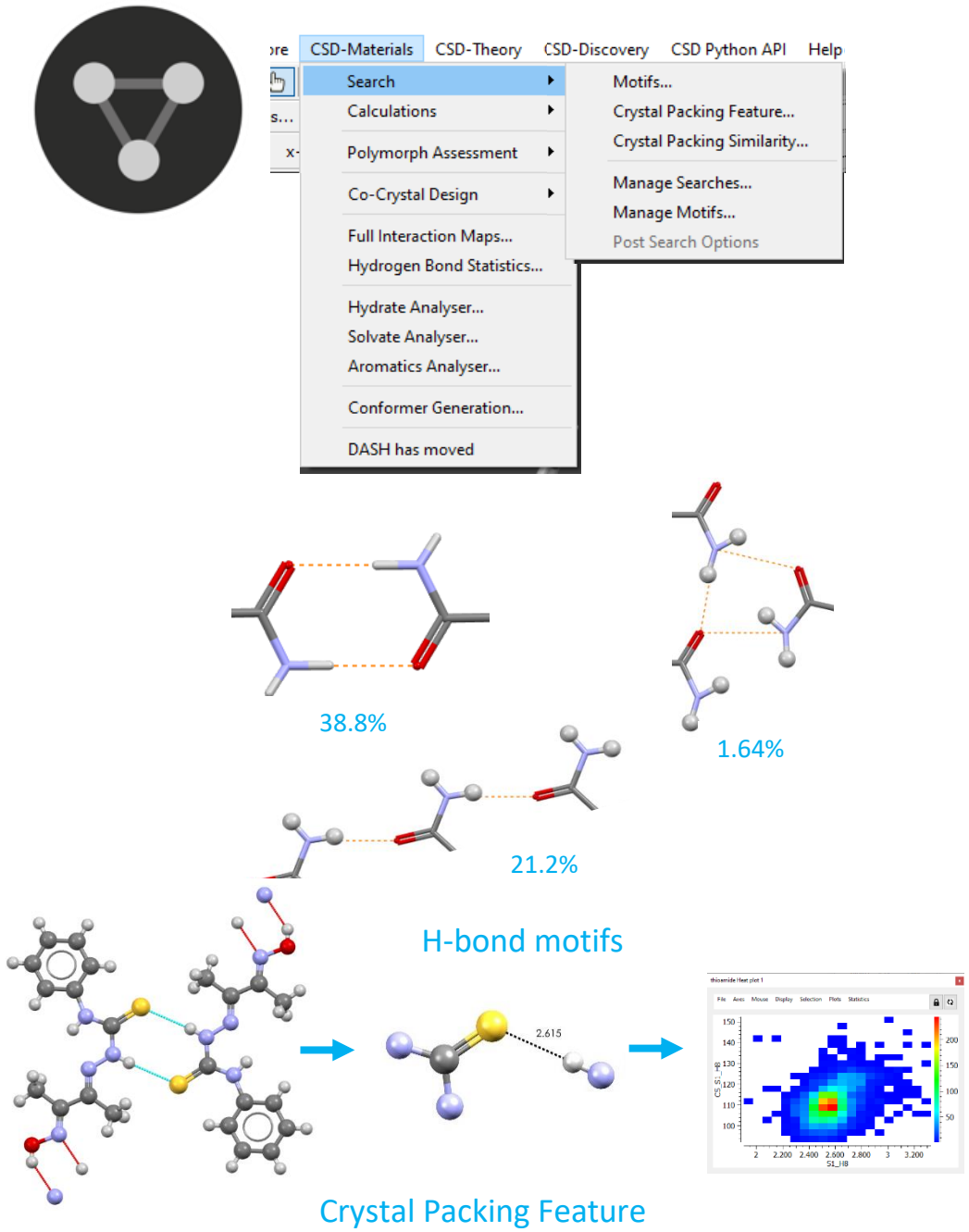
This workshop will take approximately 30 minutes to be completed. The words in *Blue Italic* in the text are reported in the Glossary at the end of this handout.

Pre-required Skills

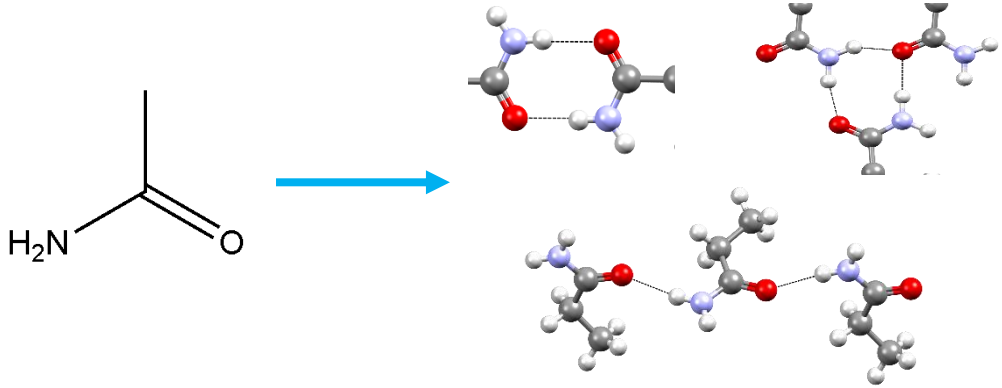
For this tutorial, we recommend being familiar with basic display and manipulation of structures from a 3D coordinates file in Mercury. A summary review is given at the end of this workshop.

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 one 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

First, we will search for primary amides with the motif **CONH2** ***R2,2(8).***

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

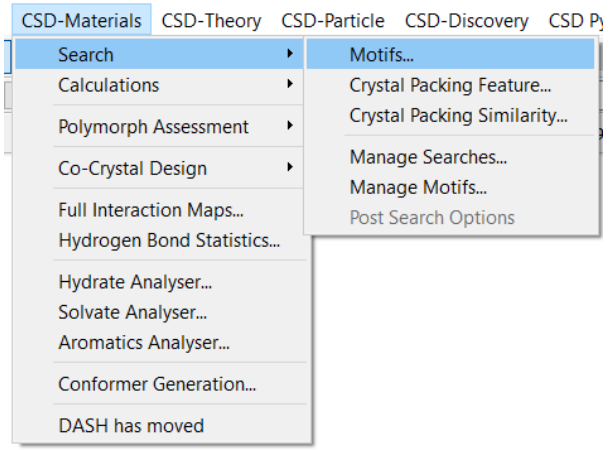
2. From the top-level menu select *CSD-Materials > Search > Motifs…* to launch the **Motif Search Wizard** window.

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

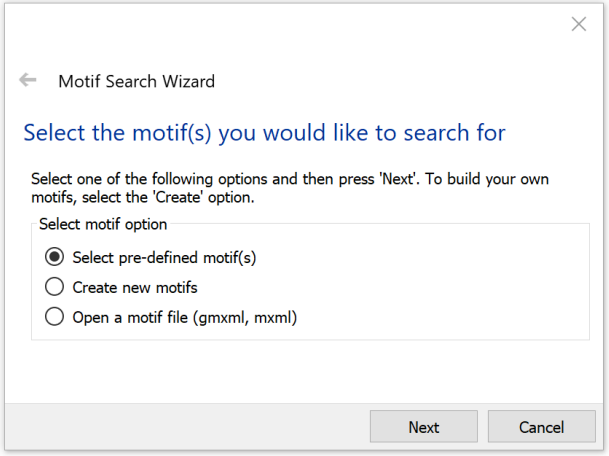
4. Click on the arrow next to ***homomeric*** to bring up the list of motifs.

1

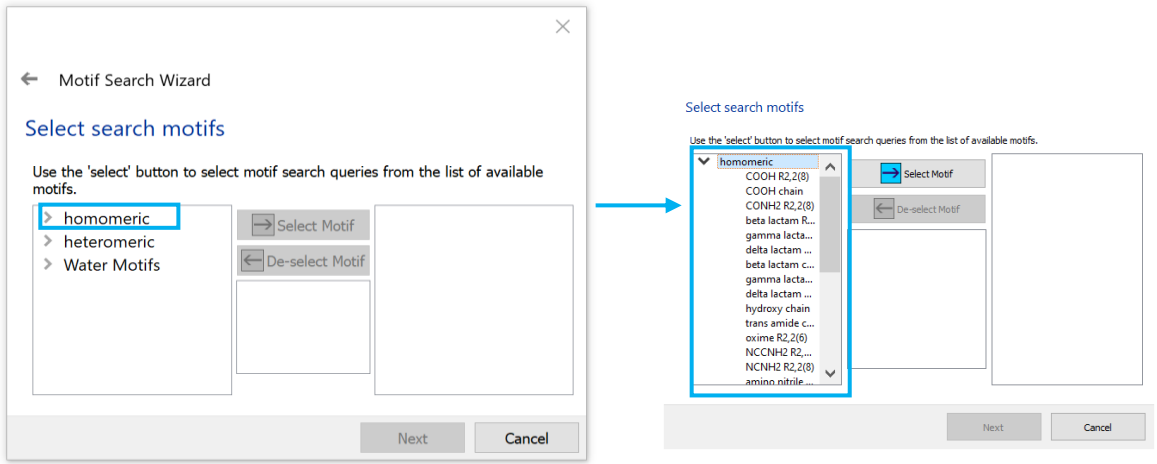
2



3



4



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

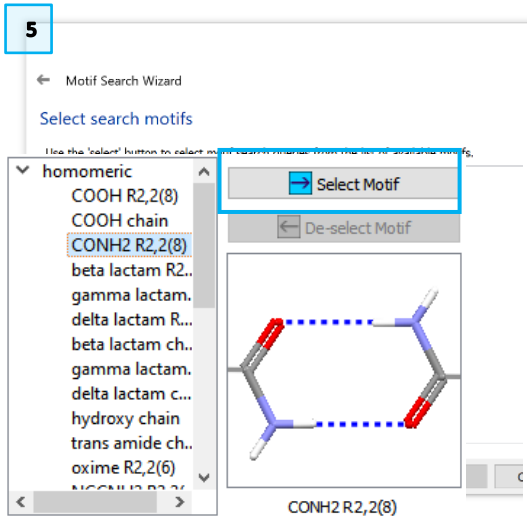
6. The CONH2 R2,2(8) motif should be listed on the right-hand side column. Select **Next** to continue.

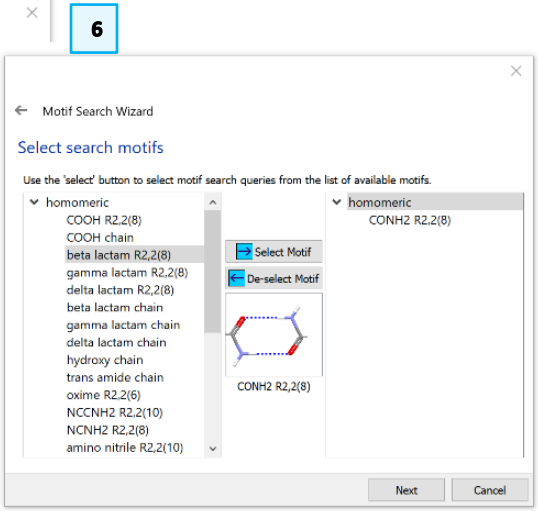
7. Select the *Databases* available (in this instance CSD 5.45 but your version may be different) and then click on the black arrow of the blue box button.

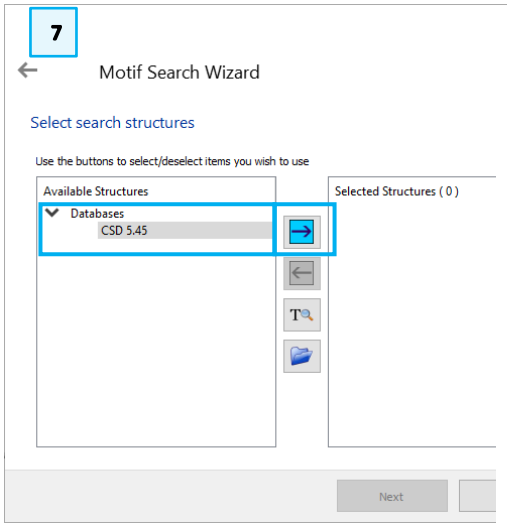
8. The selected Databases should appear on the right-hand side list. Select **Next** to continue.

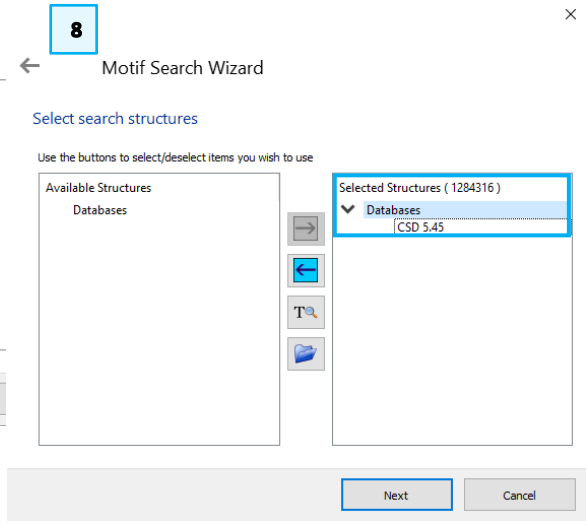
9. You can enter a search name if you wish, for this example we will leave the default name as “motif_search_1”. To start the search, click the **Start Search** button.

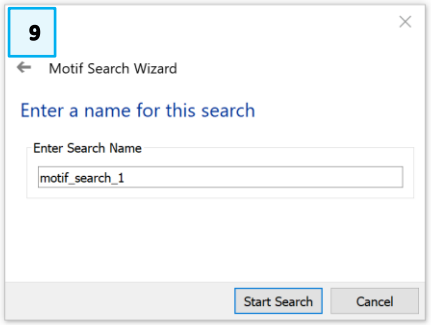
10. The Searches have now started, and you can notice a progression bar in the bottom right corner. Wait for the bar progression to reach 100% for the search to be completed.

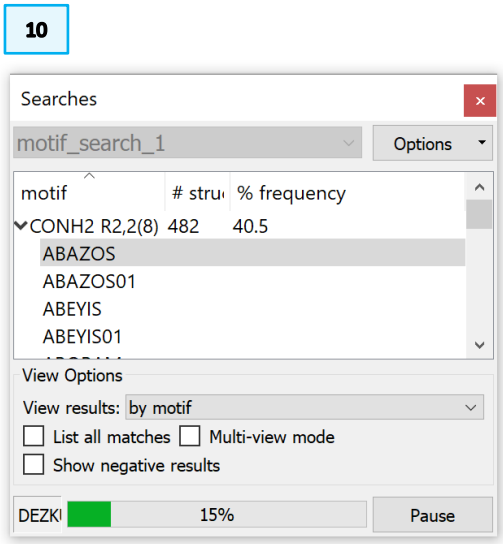












11. Once the search is completed all refcodes identified are listed in the Searches window on the right side. For the CONH2 R2,2(8) motif, at least 3417 structures were identified (with CSD 5.45) which represent an overall frequency of 38.4%.

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.

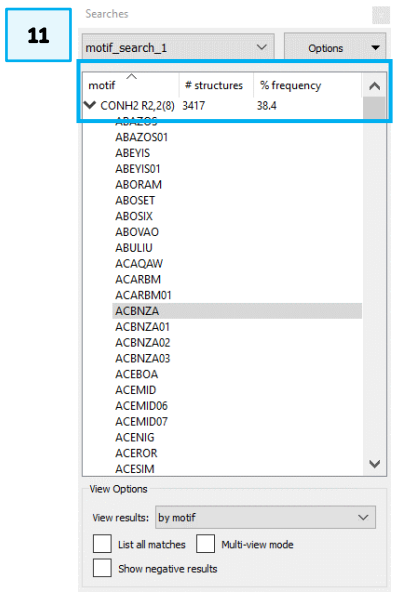
We will now conduct another search, for a different primary amide motif, **CONH2 *R2,3(8)***

13. Repeat **Steps 2 to 4** above and select the **CONH2 R2,3(8)** motif to perform a search for this type of interaction. Click on the **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 7-8** above and select **Next** to continue.

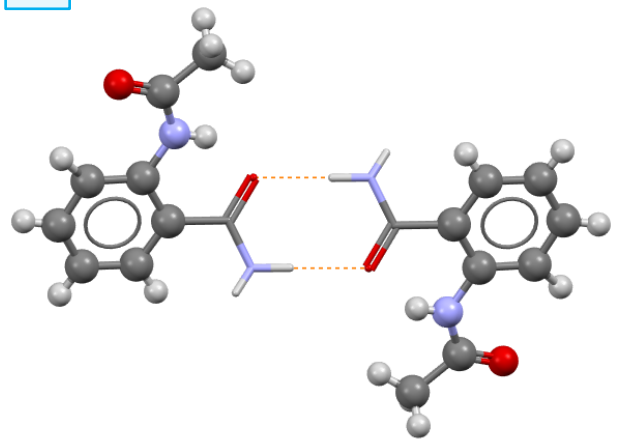
14. You can enter a search name if you wish, for this example we will leave the default name as motif_search_2. To start the search, click the **Start Search** button.

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% for the search to be completed.

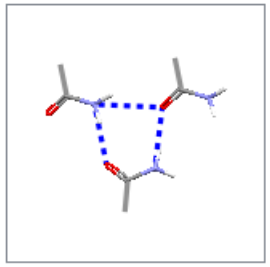
11



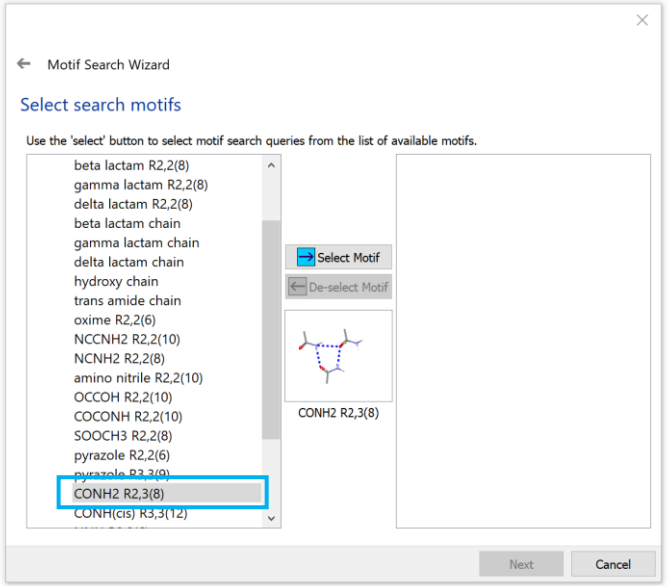
12



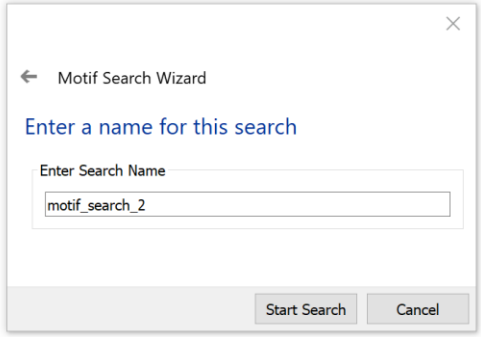
13



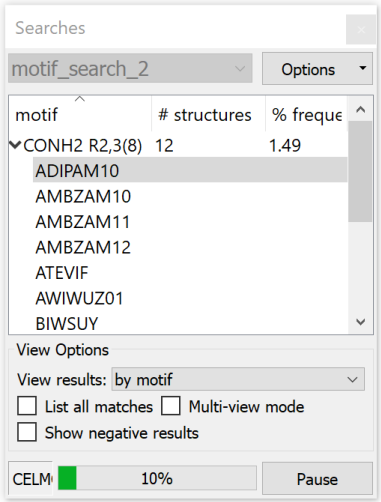
CONH2 R2,3(8)



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 146 structures were identified (with CSD 5.45) which represent an overall frequency of 1.64%.

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

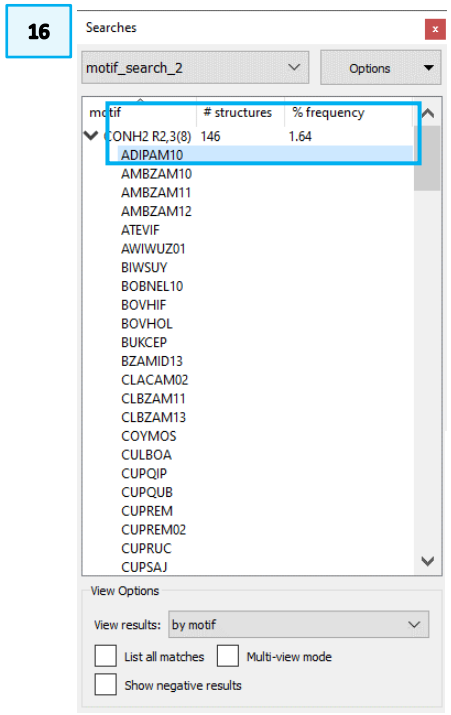
In the next search we will identify the frequency in the CSD of trans amide *chain*.

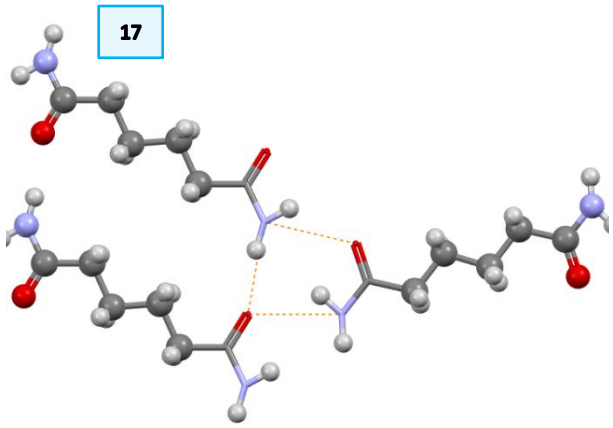
18. Repeat **Step 2** above and then select **Create new motifs**.

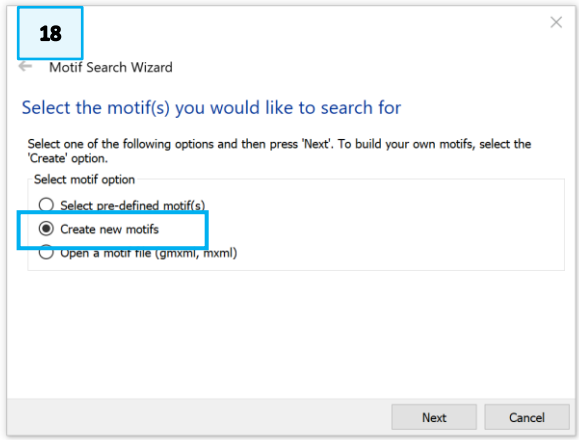
19. To generate the trans amide motif, select the **Add...** button.

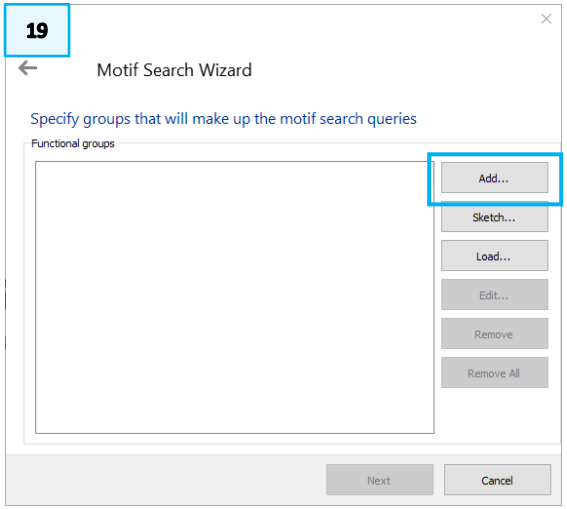
20. In the **Select substructure** window click on the *Substructure* menu and select the *carbamoyl_2* substructure the click **OK** button to continue.

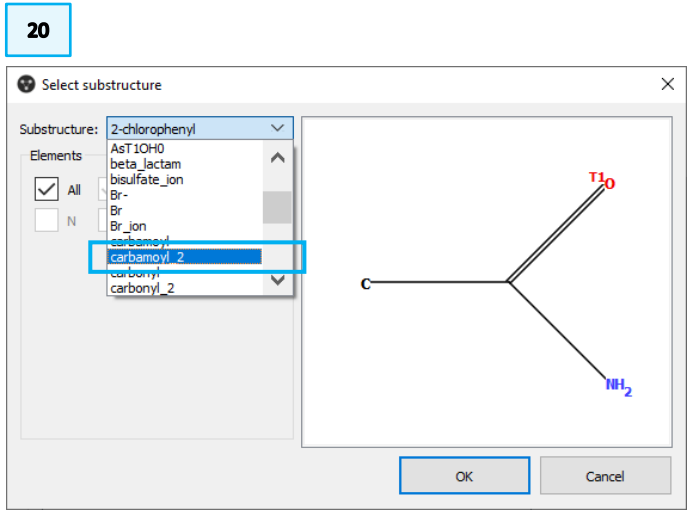
21. The carbamoyl substructure should be displayed in the **Motif Search Wizard** window, click **Next** to continue.

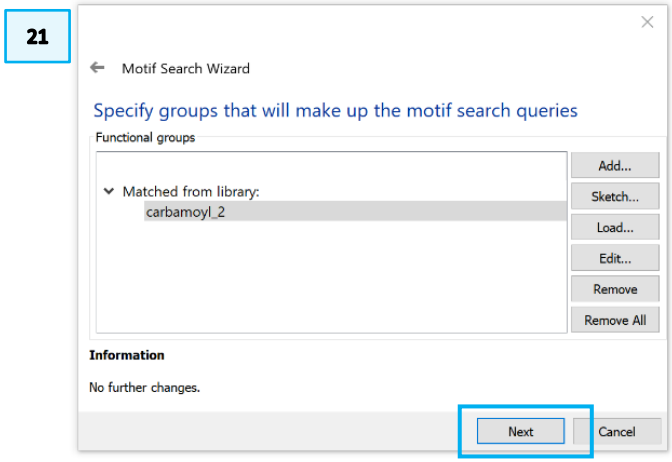












22. Select **Add...** to select the contacts present in the motif.

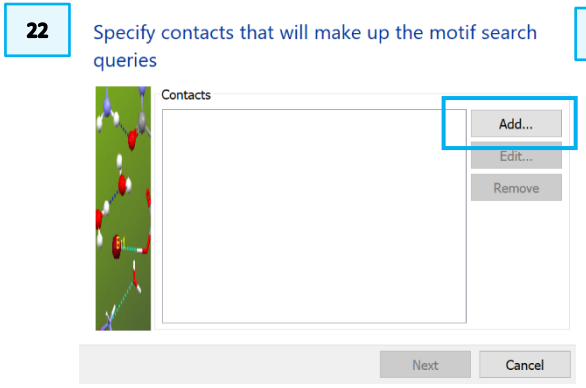
23. In the *From* column select *carbamoyl_2* in the Group drop down menu and N atom will be displayed in the Atom drop down menu. In the *To* column select *carbamoyl_2* in the Group drop down menu and O atom. The interaction will be illustrated in red in the display window. Click **OK** to continue.

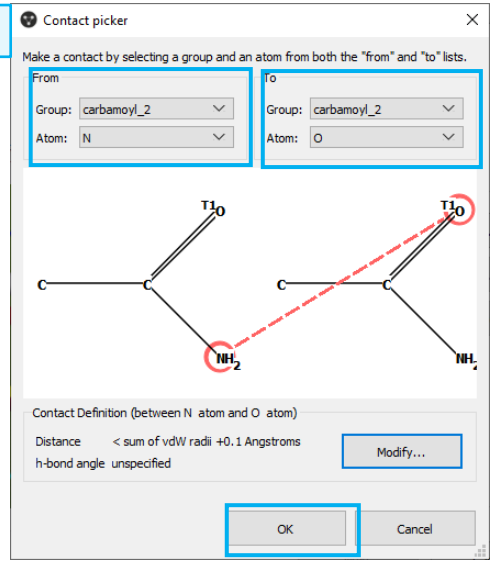
24. The specified contact will be illustrated in the wizard window. Click **OK** in the **Contact picker**, then select **Next** in the **Motif Search Wizard** to continue.

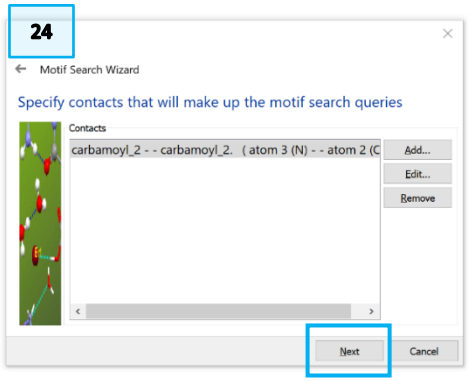
25. Select *Generate infinite chains* that repeat every 1, 2 and 3 unit the select **Next** to continue.

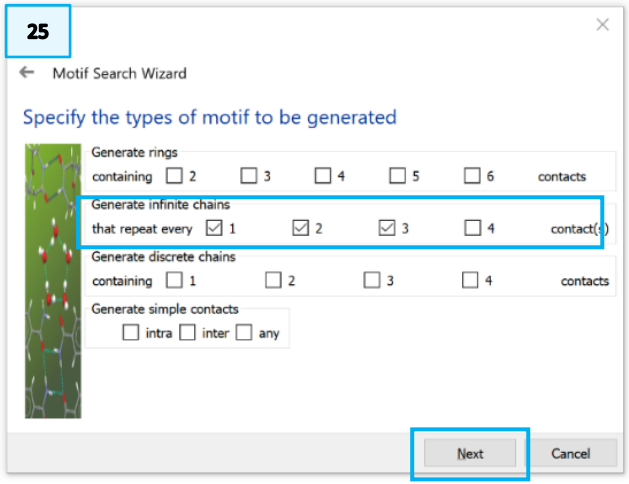
26. Note that 6 motif search queries were generated containing the chains to search for. Select **Next** to continue.

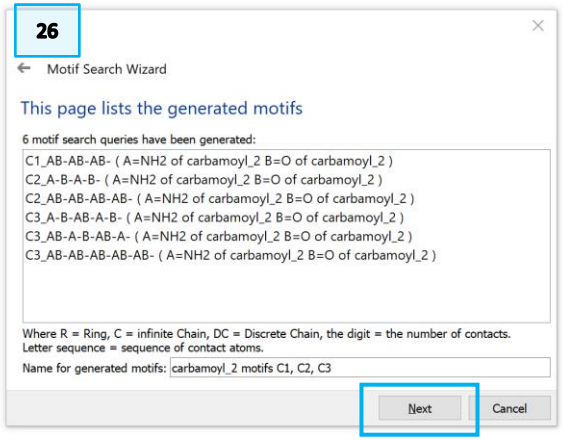
27. Ensure that the CSD is selected as shown in **Steps 7-8** above. You can enter a search name if you wish, for this example we will leave the default name as motif_search_3. To start the search, click the **Start Search** button.

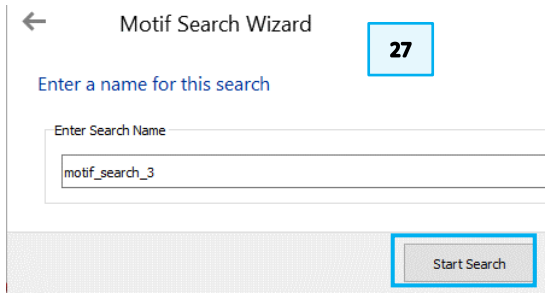












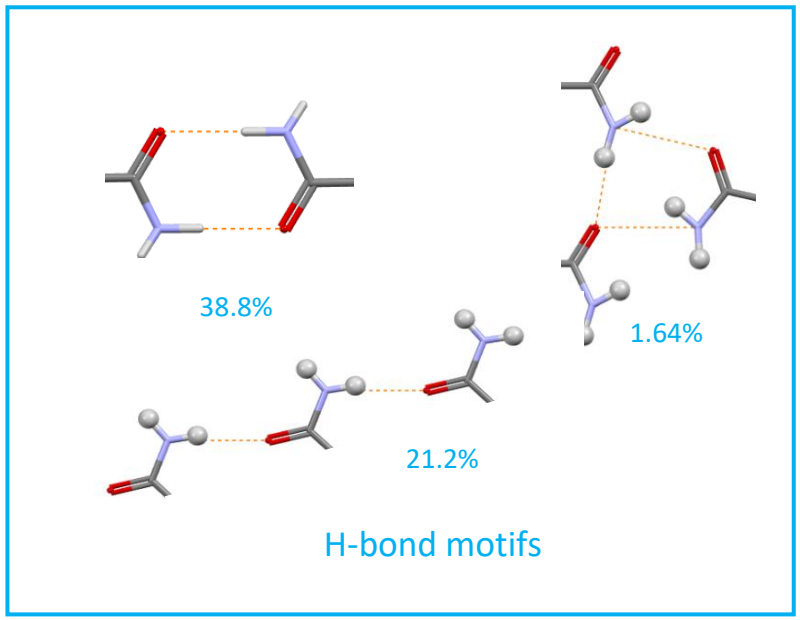
28. The Searches have now started, and you can notice a progression bar in the bottom right corner. Wait for the bar progression to reach 100% for the search to be completed.

29. Once the search is completed all refcodes identified are listed in the window wizard. For the trans amide chain motif at least 1175 structures were identified (with CSD 5.45) which represent an overall frequency of 21.2%.

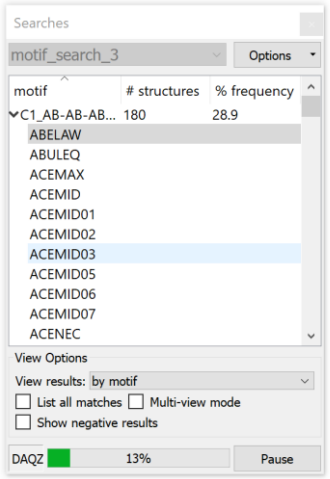
30. Scroll down through the refcode list to analyse the type of molecules that contain the trans amide chain motif. We are illustrating here the first refcode from the list ABELAW.

Conclusion

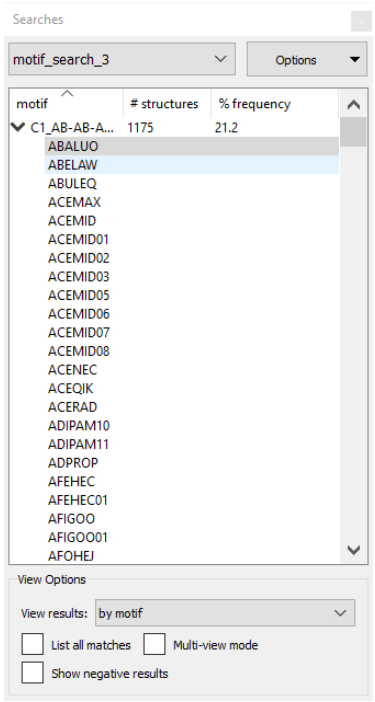
The Motif Search based on intermolecular interactions was used here to identify the abundance of different H-bond arrangements of primary amides present in the CSD. We identified the frequencies of occurrence in CSD for primary amides and noted that the CONH2 R2,2(8) motif is the most commonly observed (38.4%, *as of the 2023.3 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 most commonly observed (21.2%) and the CONH2 R2,3(8) is the least observed with only 127 structures found (1.64% frequency).



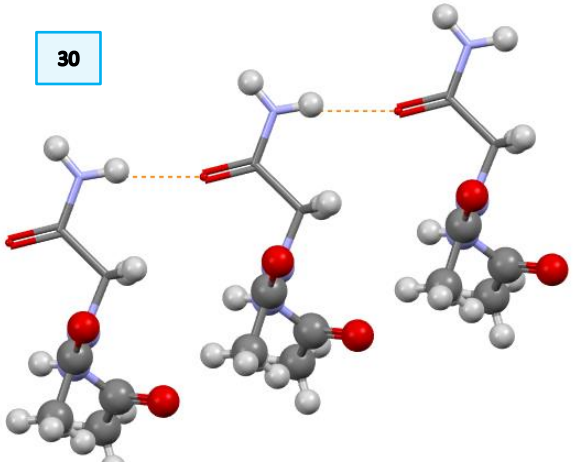
28



29



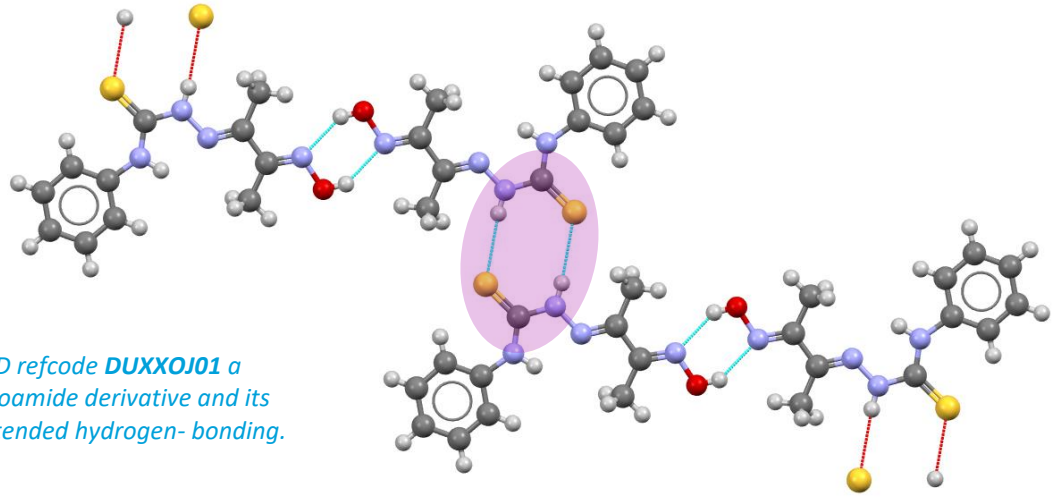
30



Example 2. Crystal Packing Feature: Investigating and 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.*

---

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

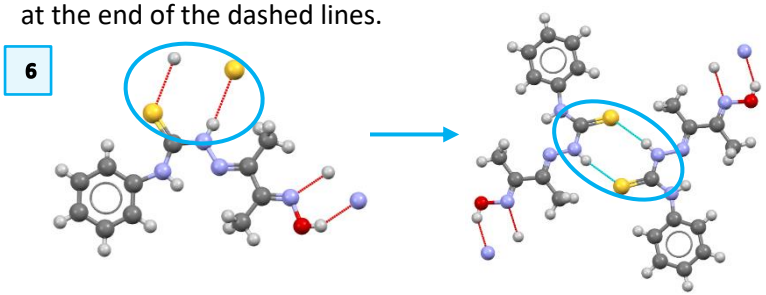
2. In the **Structure Navigator** toolbar type the “DUXXOJ01” refcode.

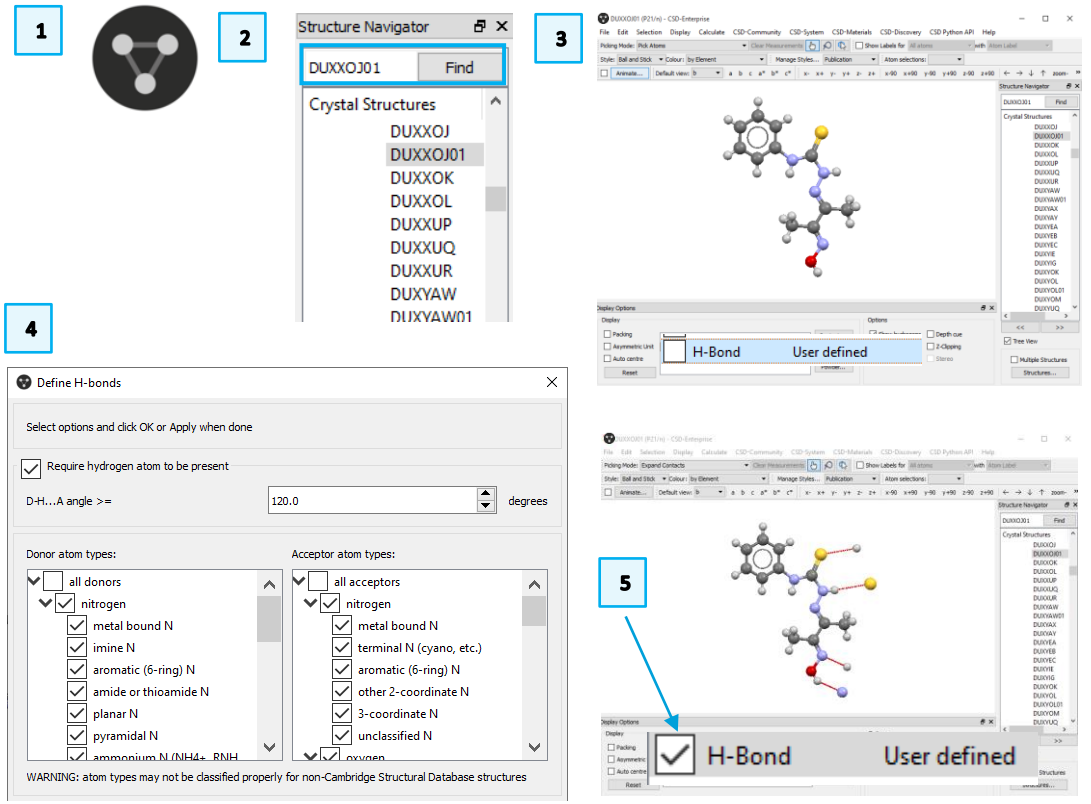
3. The structure will appear in the Mercury interface. Edit the H-bond definition by double-clicking on the *H-Bond* line in the **Display Options** toolbar.

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.

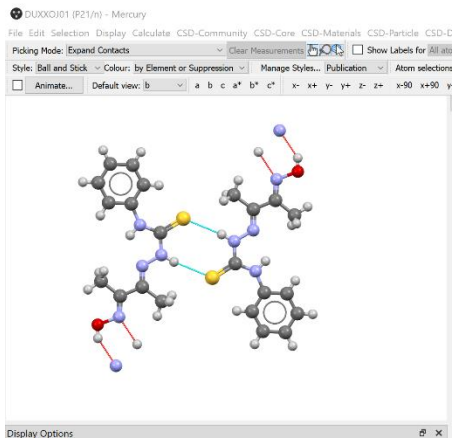
8. Click to select the atoms forming the interaction of interest in the structure, *i.e.* the thiourea – N-H donor group. Selected atoms will be highlighted in yellow. To de-select an atom, simply click it again.

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.

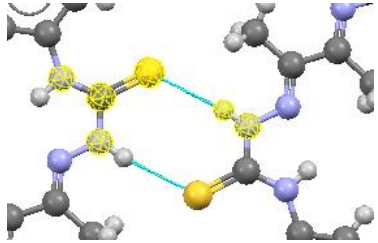
10. The Mercury display will now only show the atoms you selected; these may be hidden behind the dialogue box. In the top-level tool bar of the Mercury interface, tick the box to **Show Labels for All atoms** to label the atoms in the display. This will help in modifying atom properties in the following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

11. The wizard now allows you to define variable atom or bond types for the fragment interaction you are searching. Click in the visualizer 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.

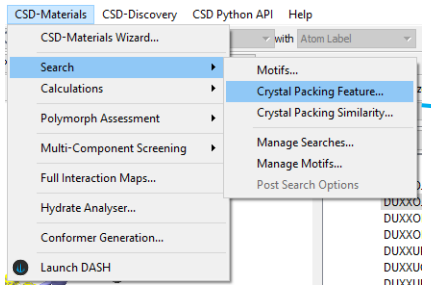
7

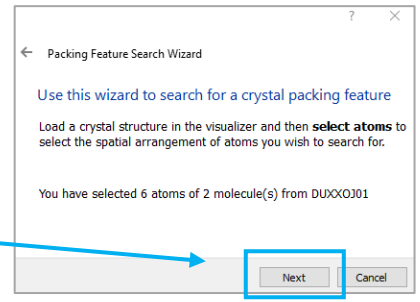


8

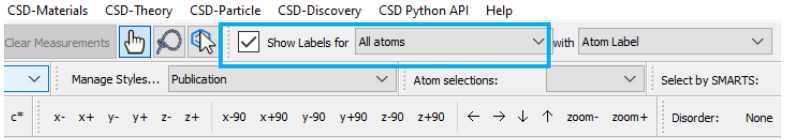


9

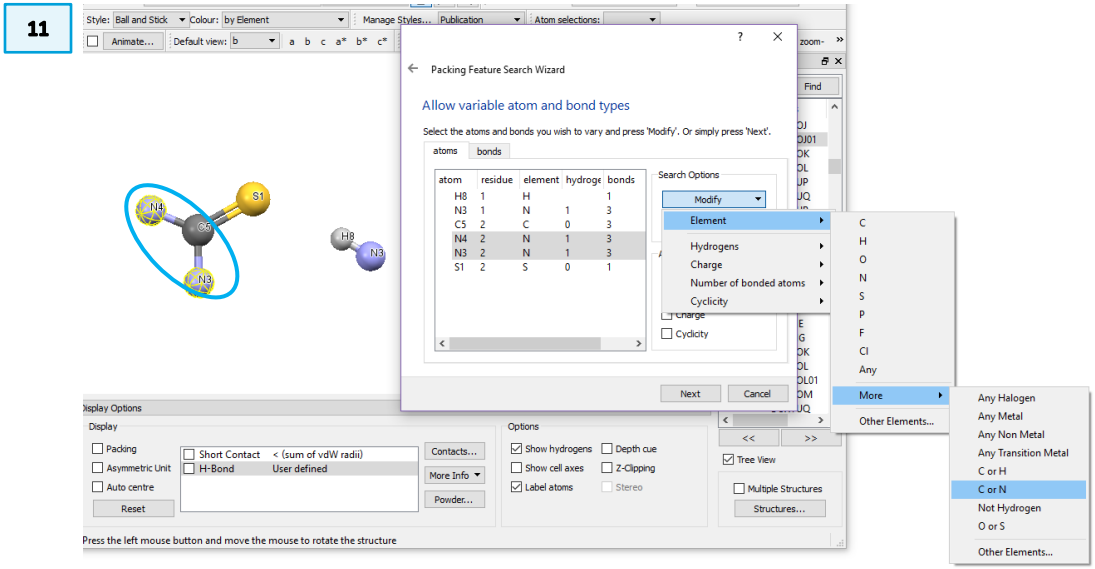


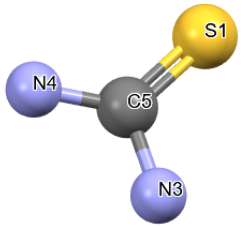


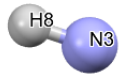
10



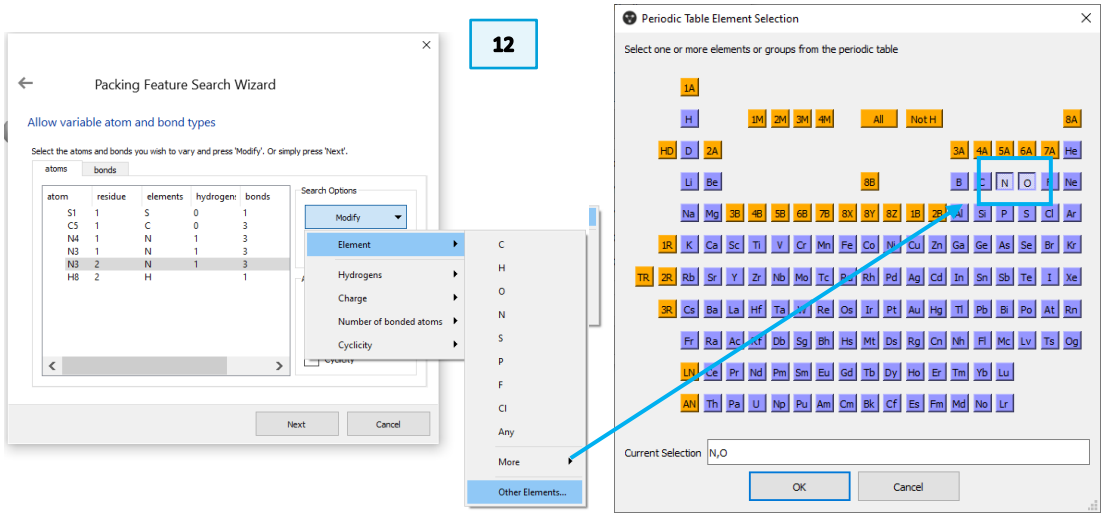
11







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**.



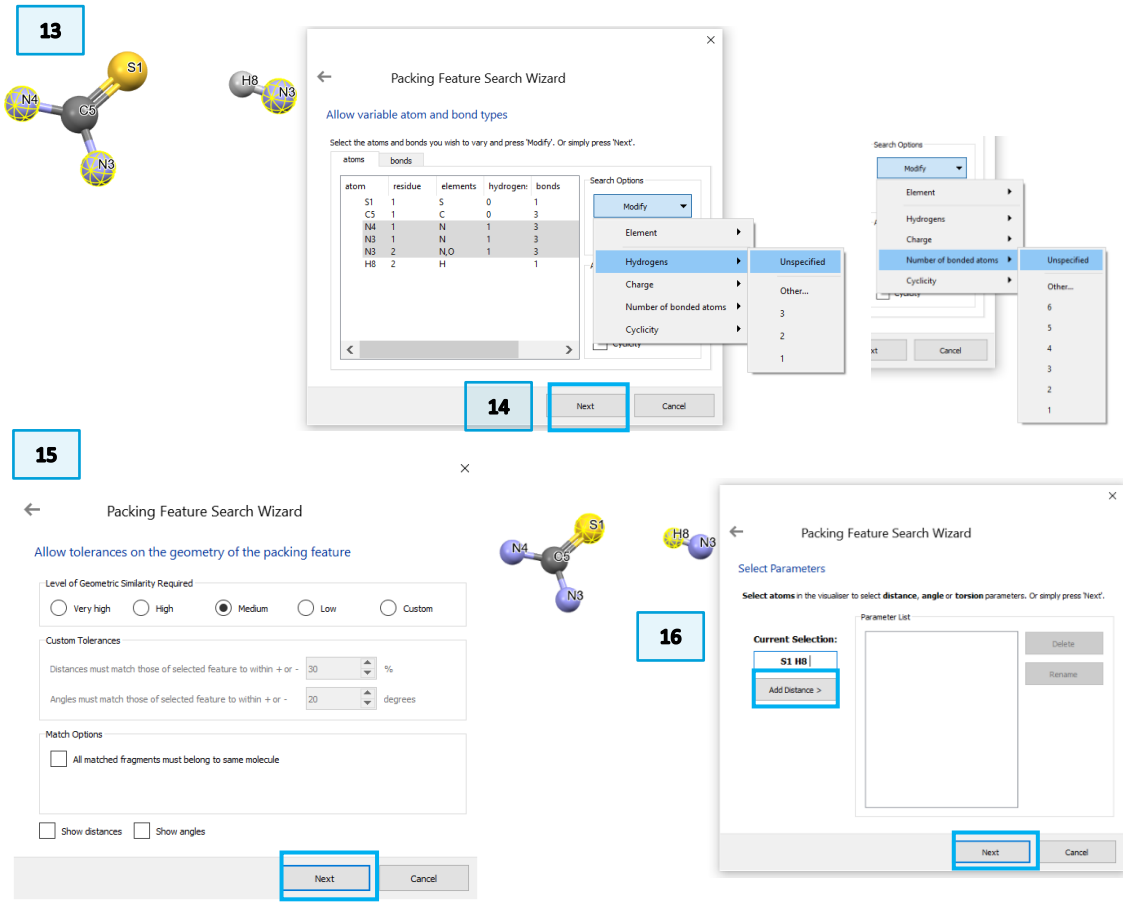
13. We also want to allow the number of hydrogen atoms and bonded atoms to vary on all three nitrogen atoms. To do this, click to select all three nitrogen atoms in the Mercury display. Choose *Modify > Hydrogens > Unspecified* and then *Modify > Number of bonded atoms > Unspecified* from the drop-down 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. You should have 3 parameters defined in the list. Click **Next** to continue.

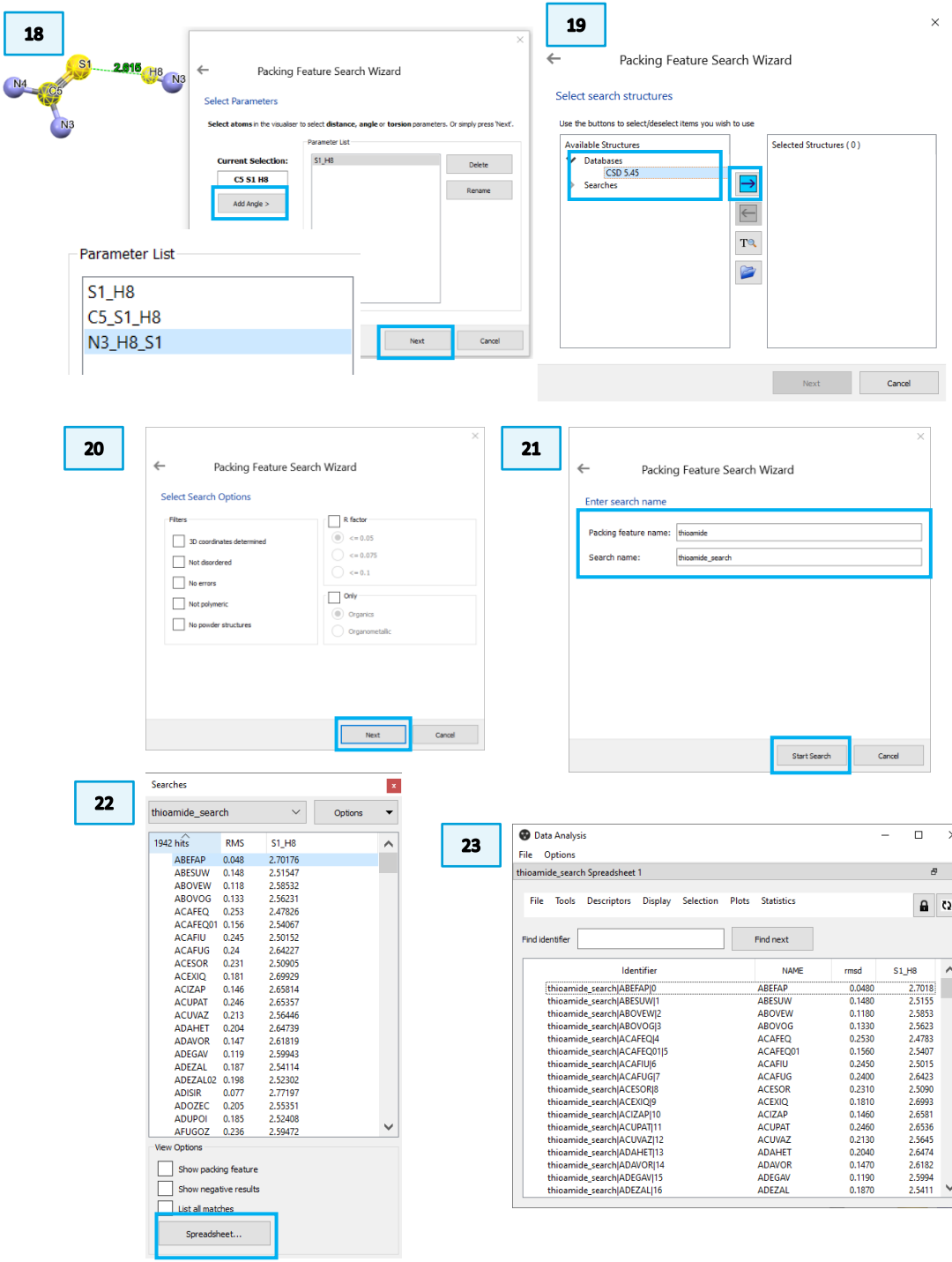
19. The next step allows you to choose which data to search. For this example, use all available data. To do this, click the headline **Databases** at the top of the list and then the right arrow, to move all data to the *Selected Structures* box. If you have your own in-house databases, you can add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

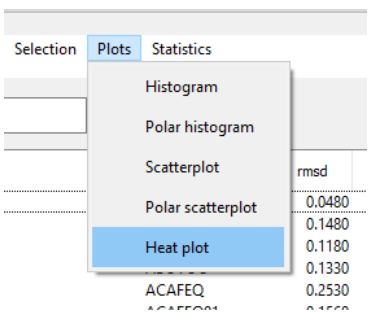
25. Choose the S···H distance for the X Axis and the C=S···H angle for the Y Axis. Click **OK** to close the dialog box.

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.

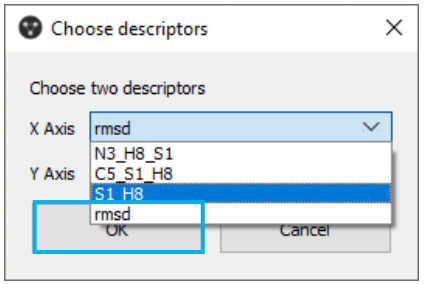
27. Return to the spreadsheet view by clicking the appropriate tab. Choose *Plots* > *Scatterplot*. Choose the S···H distance for the X axis and the N-H···S angle for the Y axis. Click **OK** to continue.

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.

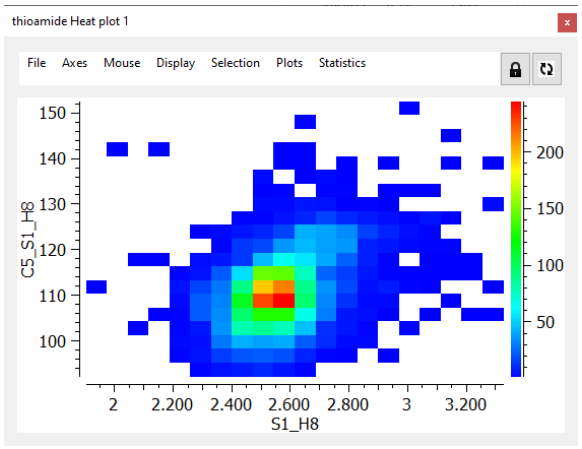
24



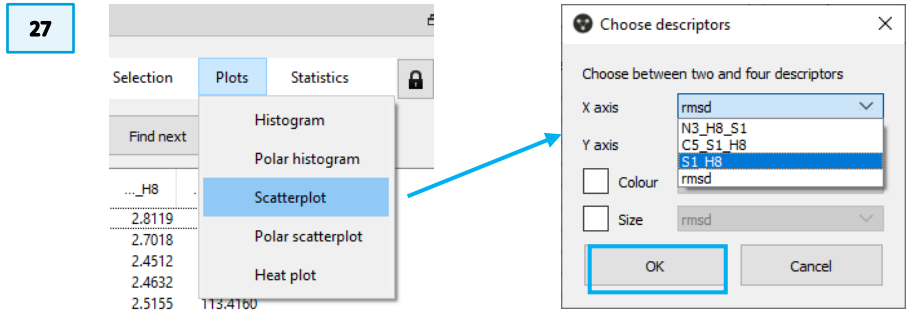
25



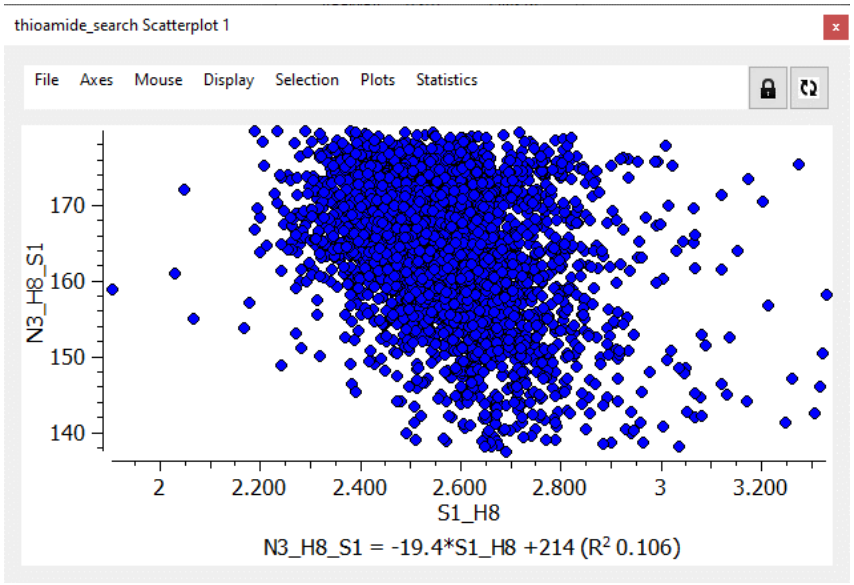
26



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.

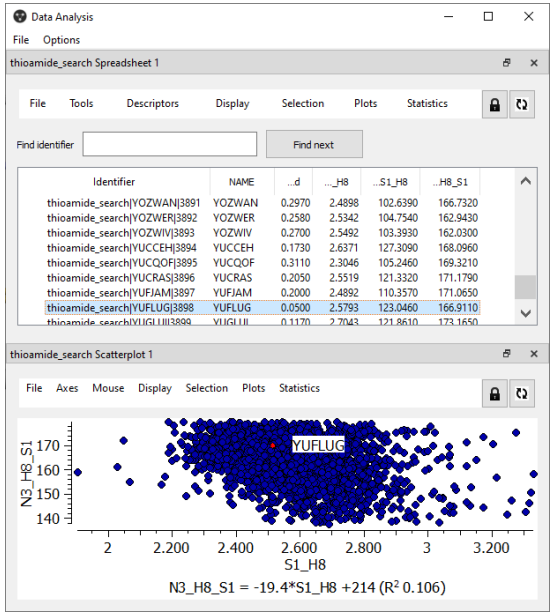
30. Visual inspection of the structures in the densest part of the distribution shows that most acceptors have one or two nitrogen atoms covalently bonded to the C=S carbon.

Conclusion

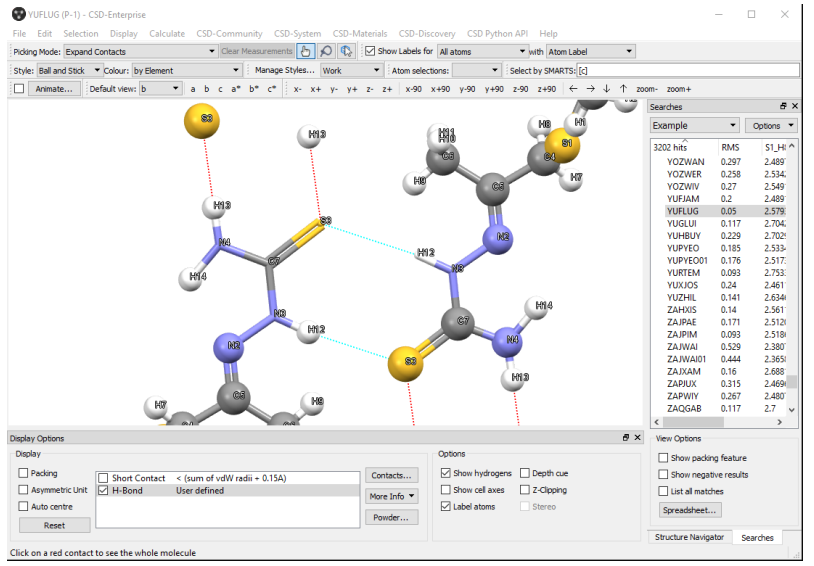
This example shows a high frequency of hits with a particular geometry. This clustering suggests that C=S groups can act as hydrogen-bond acceptors with clear directional preferences. Most observations of C=S as a hydrogen-bond acceptor are in systems where there are one or two nitrogen atoms covalently 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



Summary

This workshop illustrated searches that can be performed using CSD-Materials, focussing on uncovering structures with common hydrogen bonding motifs. You have learned how to identify the frequencies of occurrence of primary amides in different hydrogen bond networks using the Motif search feature and you have seen how to apply the Crystal Packing Feature search to uncover unusual hydrogen interactions such as S…H-N. You should now be able to:

- Set up and run a Motif search in Mercury using pe-defined crystal packing motifs to retrieve the frequency of occurrence of a motif.
- Use Mercury's Crystal Packing Feature search, starting from a reference crystal structure, to discover structures with common packing features.
- Analyse search results using Mercury's Data Analysis module.

For your reference, you can find the Mercury user manual at this link

Next steps

If you have enjoyed this workshop, you may like to explore the Crystal Packing Similarity search workshop (MAT-008), which follows on from this workshop. You can find more CSD-Materials self-guided workshops here.

Feedback

We hope this workshop improved your understanding of *Motif and Crystal Packing Feature searching* 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-006*. It will only take 5 minutes and your feedback is anonymous. Thank you!

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.

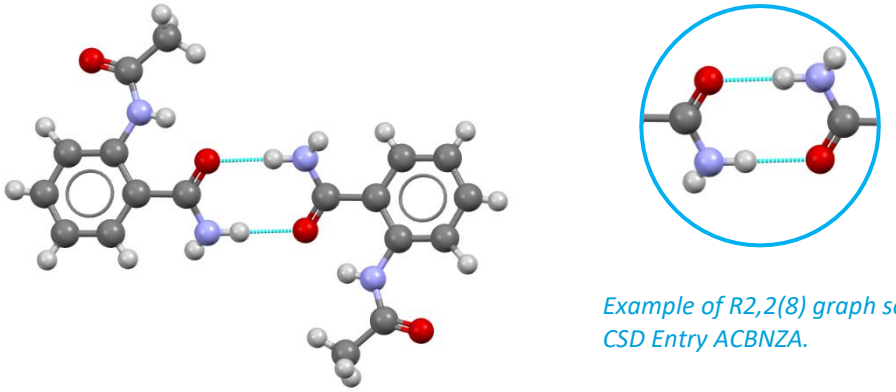
A graph-set descriptor is written as $\mathbf{G^a_d(n)}$, in which **G** represents the type of pattern, **a** is the number of hydrogen bond acceptors involved in that pattern, **d** is the number of donors and **n** the number of atoms in the pattern. The pattern type, G, can be one of four different 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:

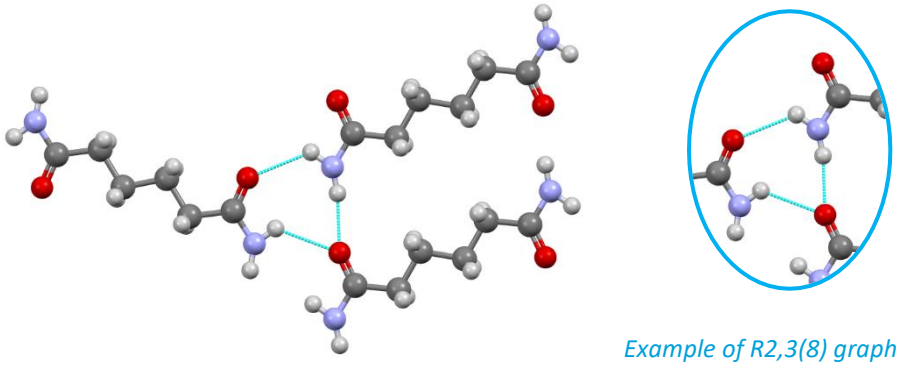
- M. C. Etter, Acc. Chem. Res., 23, 120, 1990
- J. Bernstein, R. E. Davis, L. Shimoni and N.-L. Chang, *Angew. Chem. Int. Ed.*, 34, 1555, 1995
- W. D. S. Motherwell, G. P. Shields and F. H. Allen, Acta. Cryst. B56, 466, 2000

Examples from this handout:

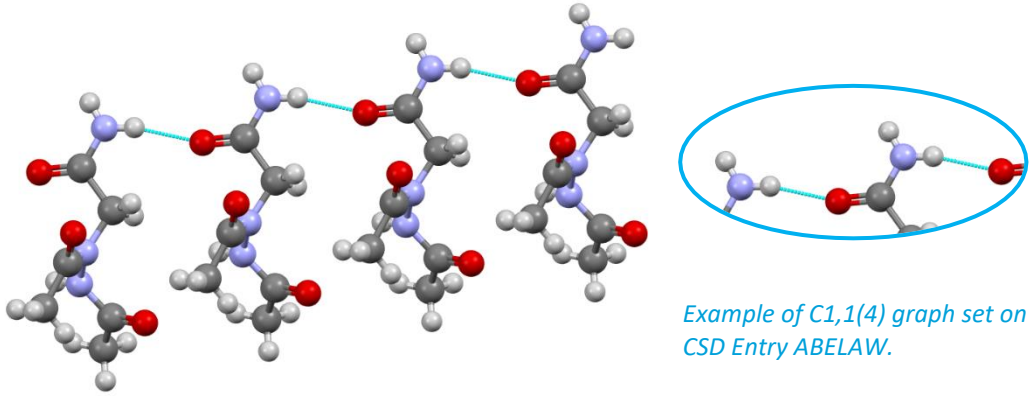
- **R2,2(8)** is a ring formed by 2 acceptors and 2 donors with 8 atoms involved in the ring.
- **R2,3(8)** is a ring formed by 2 acceptors and 3 donors with 8 atoms involved in the ring.
- **C1,1(4)** is a chain formed by 1 acceptor and 1 donor with 4 atoms involved in the ring.



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.

Heteromeric vs Homomeric

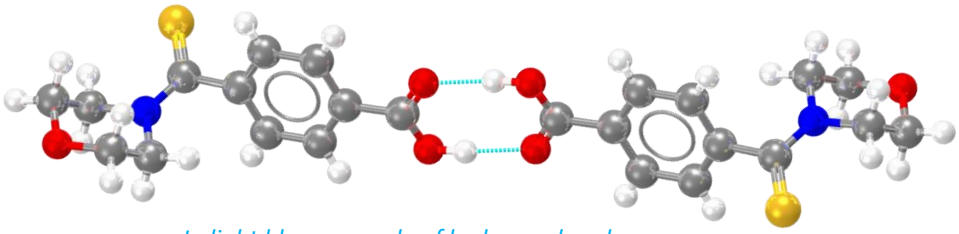
In the context of motif searching, homomeric refers to motifs that are created by the interaction and repetition of the same sub-unit or group. Heteromeric refers to when groups or sub-units are different.

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.

Motif

A motif is a characteristic pattern of interactions 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.



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

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