Investigating solid forms

Using motif searching, packing feature searching and packing similarity in Mercury

May 2022

CCDC
advancing structural science

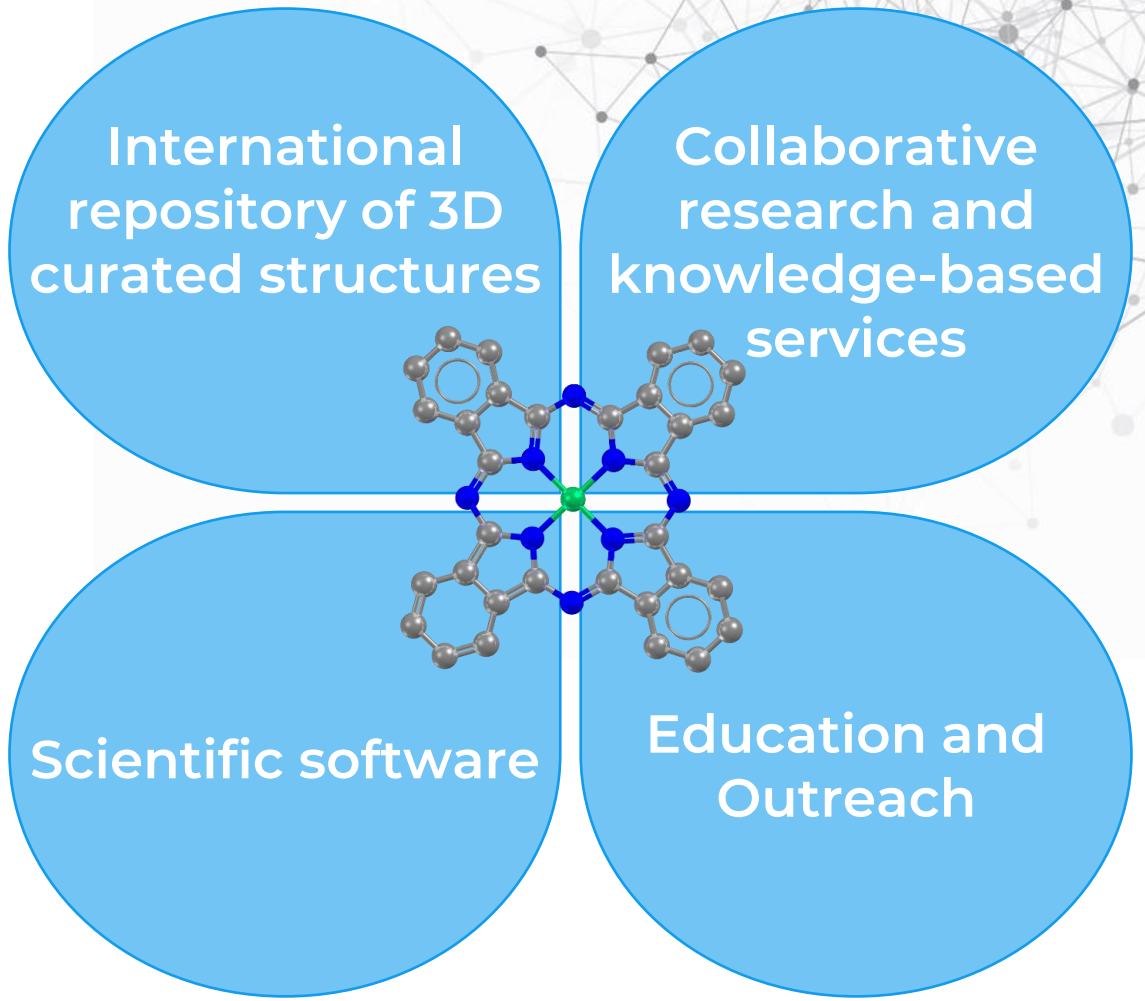
Learning outcomes

- Learn how informatics and data-driven approaches can be used to understand solid form design and risk assessment.
- Learn how to use Motif searching.
- Learn how to perform crystal packing feature searches.
- Learn how to perform crystal packing similarity searches.
- Familiarise yourself with what other tools are available in the CSD-Materials suite and what they can be used to do.

CCDC

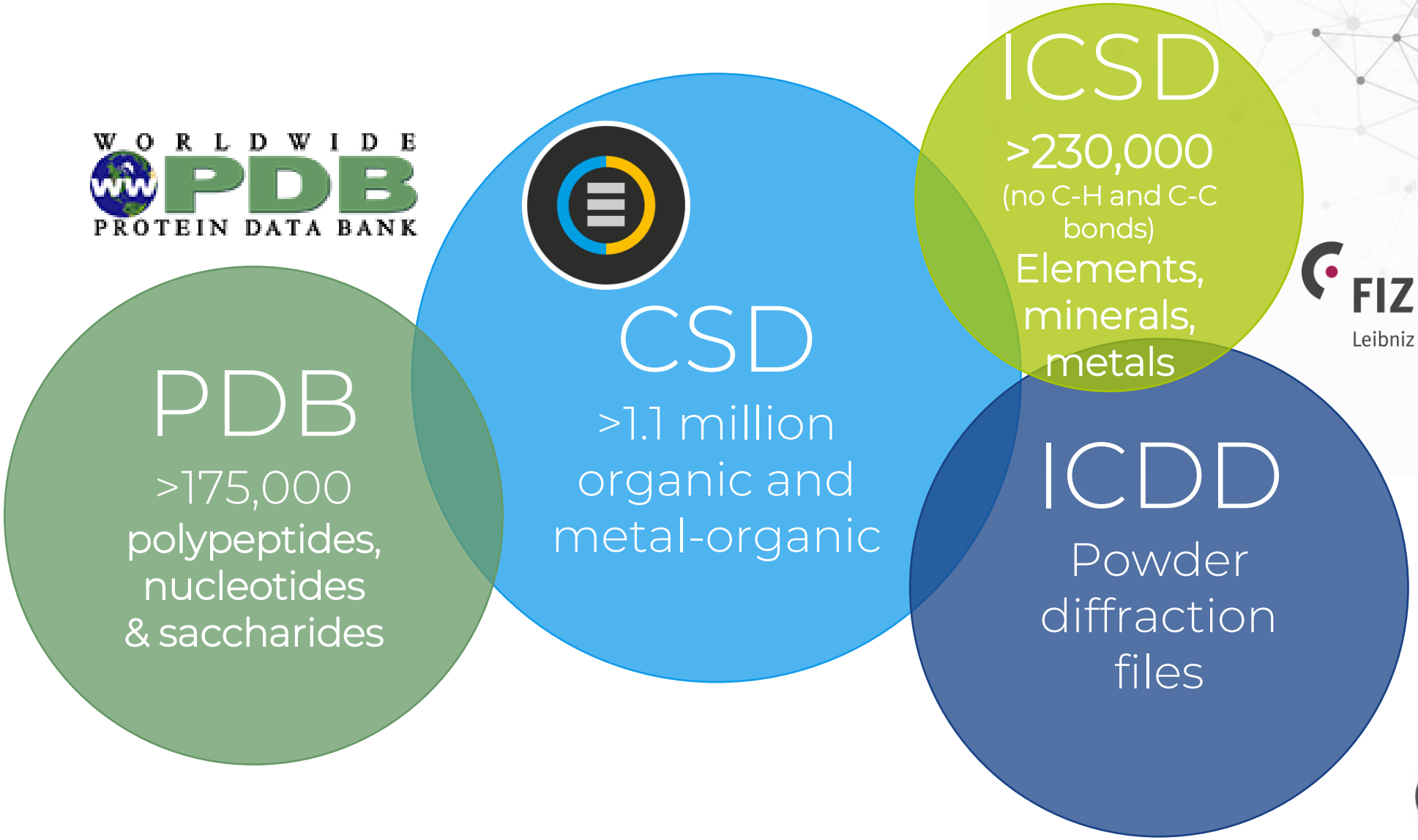
What we do at the CCDC

Charitable Objective: Advancement of chemistry and crystallography for the public benefit through providing high quality information services and software



CCDC

Structural databases



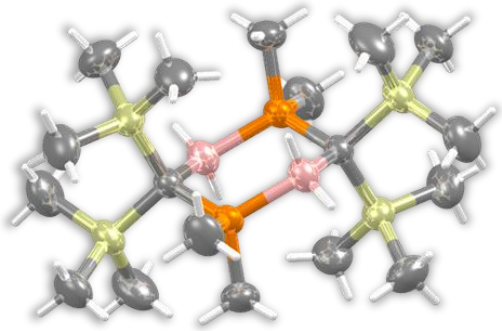
WORLDWIDE PDB
PROTEIN DATA BANK

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

ICDD®

CCDC

The Cambridge Structural Database



CSD Refcode ELOFUJ
The 1.1 millionth CSD structure
A *CSD Communication*
determined by Bill Clegg

>1,170,000

Structures published that year

Structures published previously

Number of Structures in the CSD

1,000,000
800,000
600,000
400,000
200,000
0

< 1973 1975 1977 1979 1981 1983 1985 1987 1989 1991 1993 1995 1997 1999 2001 2003 2005 2007 2009 2011 2013 2015 2017 2019 2021

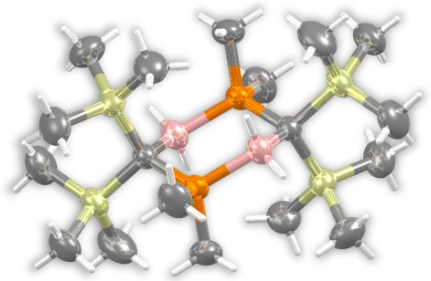
Year

- Every published structure
  - ASAP & early view
  - CSD Communications
  - Patents
  - University repositories
  - Theses
- Every entry enriched and annotated by experts
- Discoverability of data and knowledge
- Sustainable for over 56 years
- A CoreTrustSeal repository

Certified as Trustworthy by CoreTrustSeal

CCDC

CSD Refcodes



CSD Refcode - ELOFUJ

What is ELOFUJ?

- A CSD Refcode
- A database reference code
- Containing 6-8 characters
- Used to identify entries in the CSD

Refcode families

- The same substances are assigned the *same* 6 letter code plus an additional 2 numbers
  - Polymorphs
  - New determinations or re-refinements of the same substance
  - Determinations at different temperatures/pressures
- Stereoisomers or different solvates, co-crystals, etc are assigned *different* refcode families

CCDC

Inside the Cambridge Structural Database

The CSD is a database of all the published organic and metal-organic experimental crystal structures

Organic
43%

Metal-Organic
57%

At least one transition metal, lanthanide, actinide or any of Al, Ga, In, Tl, Ge, Sn, Pb, Sb, Bi, Po

Not Polymeric
89%

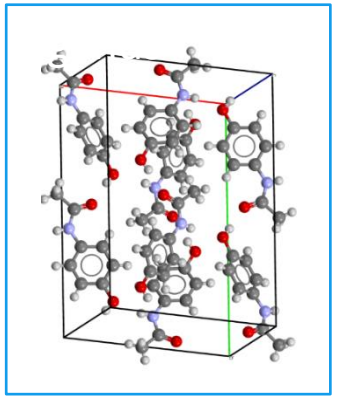
Polymeric: 11%

Single Component
56%

Multi Component
44%

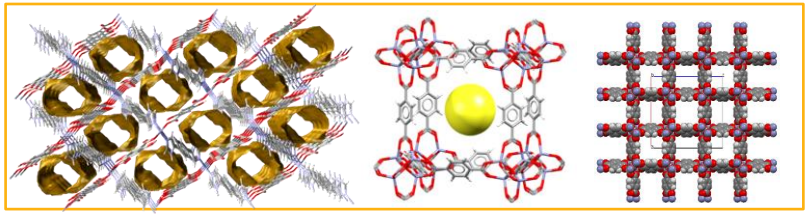
Organic

- Drugs
- Agrochemicals
- Pigments
- Explosives
- Protein ligands



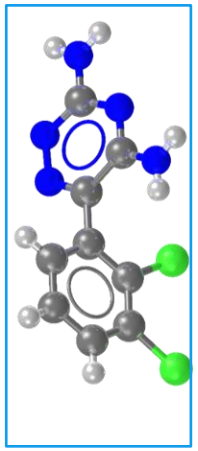
Metal-Organic

- Metal Organic Frameworks
- Models for new catalysts
- Porous frameworks for gas storage
- Fundamental chemical bonding



Additional data

- 11,525 polymorph families
- 171,683 melting points
- 909,992 crystal colours
- 778,663 crystal shapes
- 24,916 bioactivity details
- 11,387 natural source data
- > 250,000 oxidation states



Links and subsets

- DrugBank
- Druglike
- MOFs
- PDB ligands
- PubChem
- ChemSpider
- Pesticide PDB

A wealth of data in the CSD

Growth of co-crystals

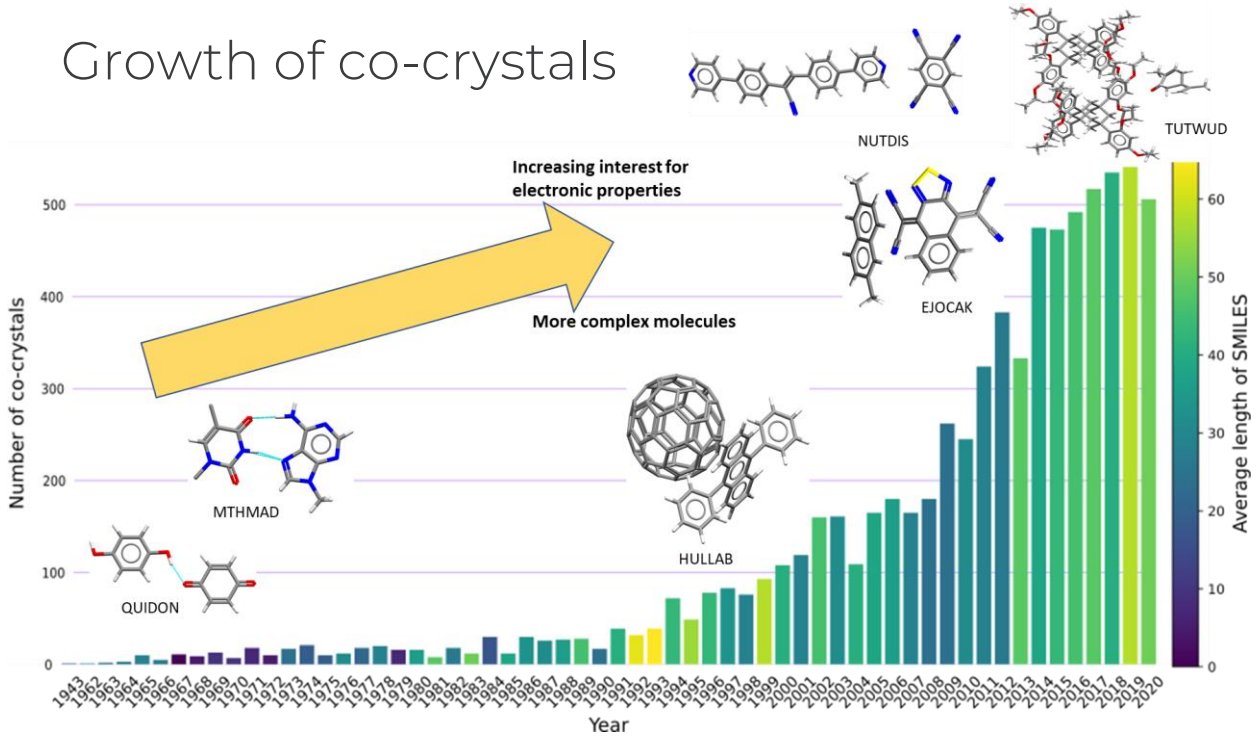
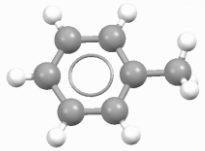
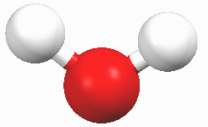


Chart showing the number of co-crystals in the CSD by year and colour coded by the average SMILES length. Image created by Katerina Vriza*, a CCDC sponsored PhD student.

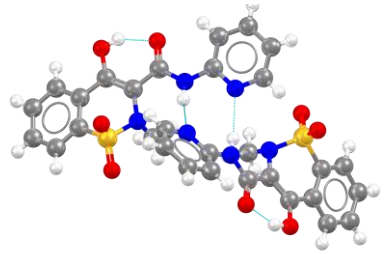


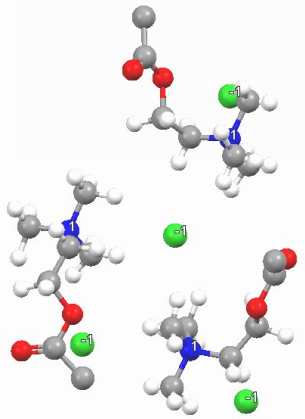
Solvates
>230,000

Hydrates
>140,000

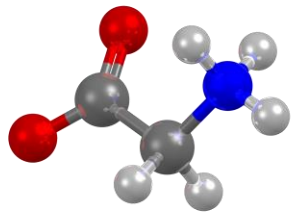


Polymorphs
>11,000 polymorphic families





Salts



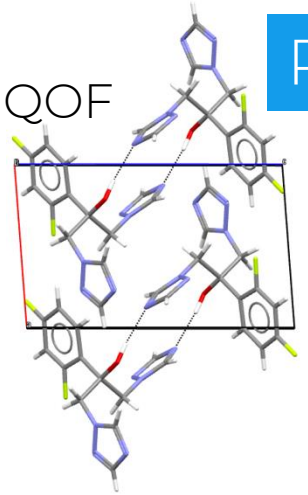
Refcode families
>1,000,000

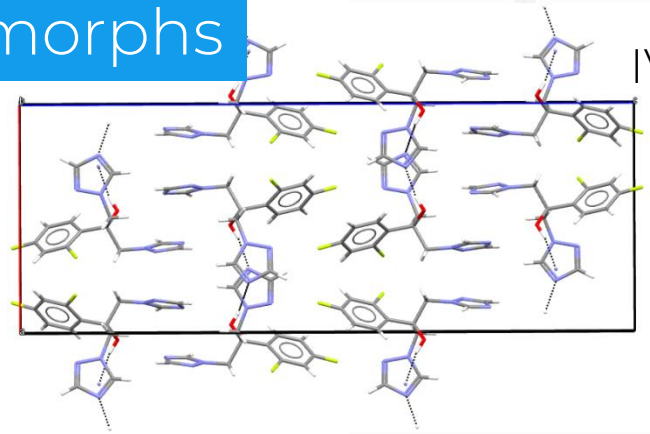
*Katerina Vriza, University of Liverpool, PhD on Data driven discovery of functional molecular co-crystal

A solid form landscape

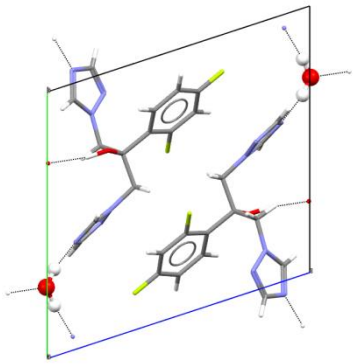
Polymorphs

IVUQOF



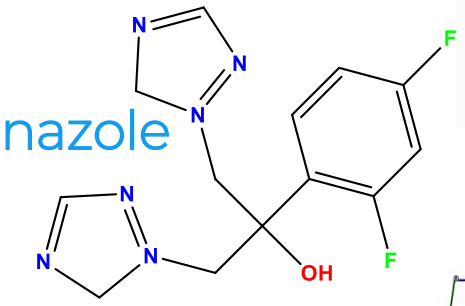


IVUQOF02

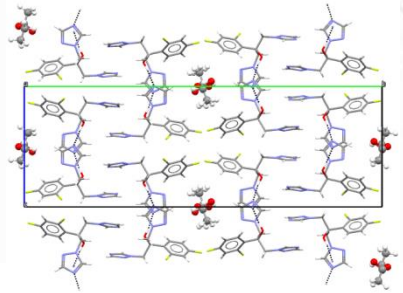


IVUQIZ

Hydrates

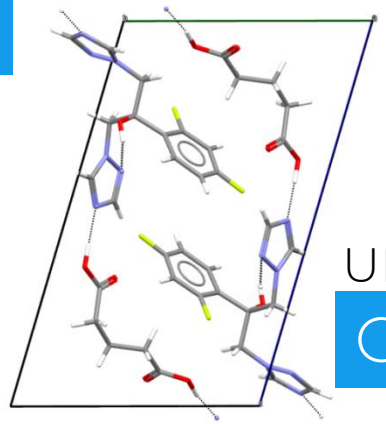


fluconazole



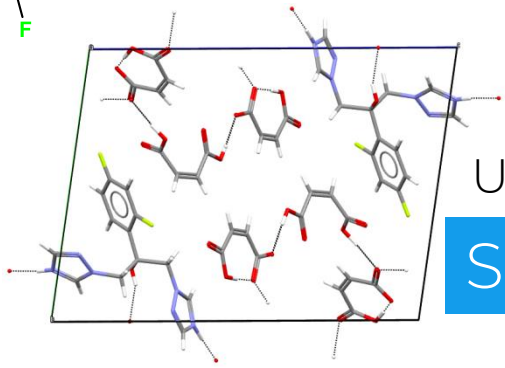
IVUQEV

Solvates



UPOQEW

Co-crystals



UPOQAS

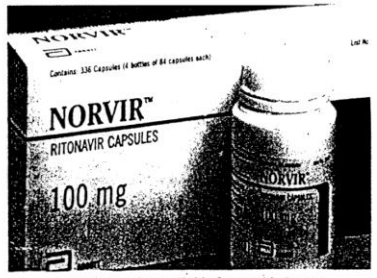
Salts

CCDC

Impact of Hydrogen Bonding on lattice stability -The Ritonavir story

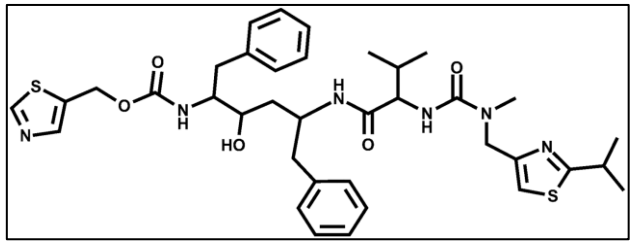
Manufacturing problems hit Abbott's HIV drug ritonavir

Capsules of Abbott Laboratories' protease inhibitor Norvir (ritonavir) are likely to become unavailable by the middle of August. The company has a problem with the manufacture of the anti-HIV capsules which it cannot resolve at present.

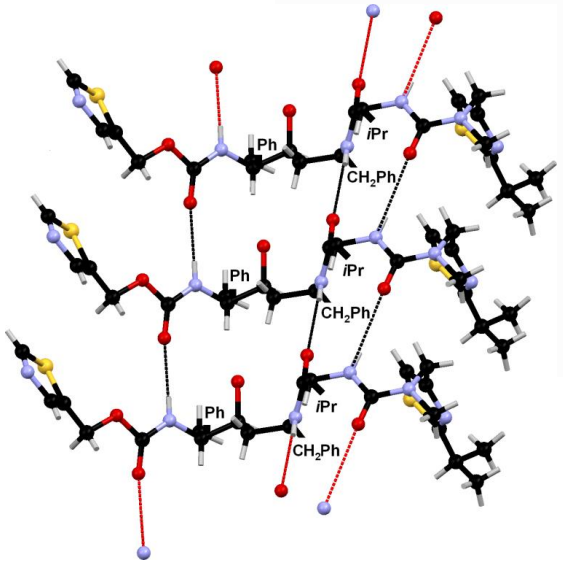


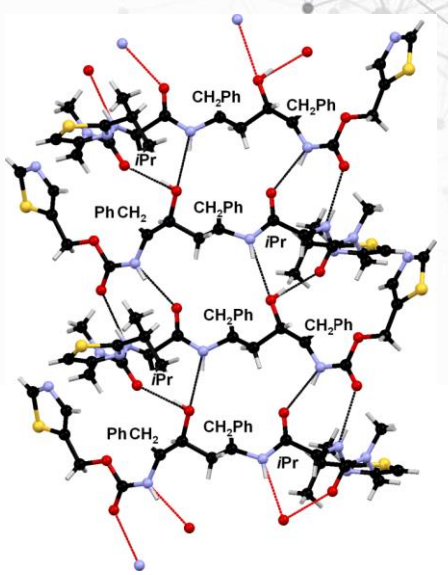
Capsules unlikely to be available from mid-August

The problem relates to "undesirable" crystal formation. Abbott says that a series of samples from a number of marketed batches of capsules were examined and there was no



- ~5-fold decrease in solubility
- Change in Hydrogen Bonding





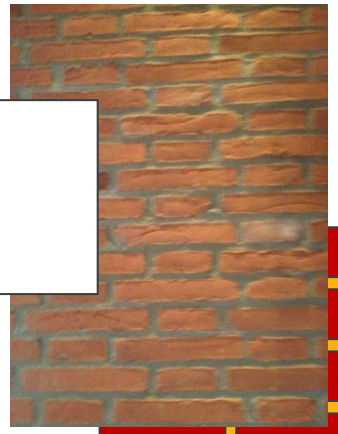
"Since the strength and completeness of the hydrogen bonding has attained the maximum possible in the Form II lattice, it is not thought possible that another undiscovered polymorph of ritonavir would exist with equivalent or lower solubility than that of Form II."

Org. Process Res. Dev. (2000), 4(5), 413-417. Pharmaceutical Research (2001), 18(6), 859-866.

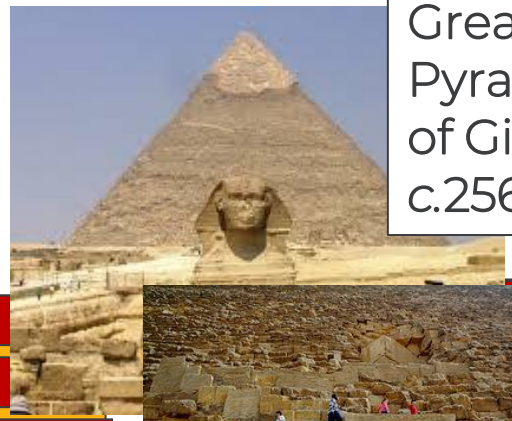
CCDC

Structural Informatics

Which is the stable wall?



The CCDC c.1992



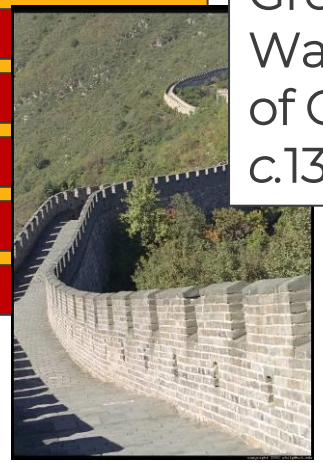
Great Pyramid of Giza c.2560 BC

Hadrian's Wall c.122

A

B

My House c.1967



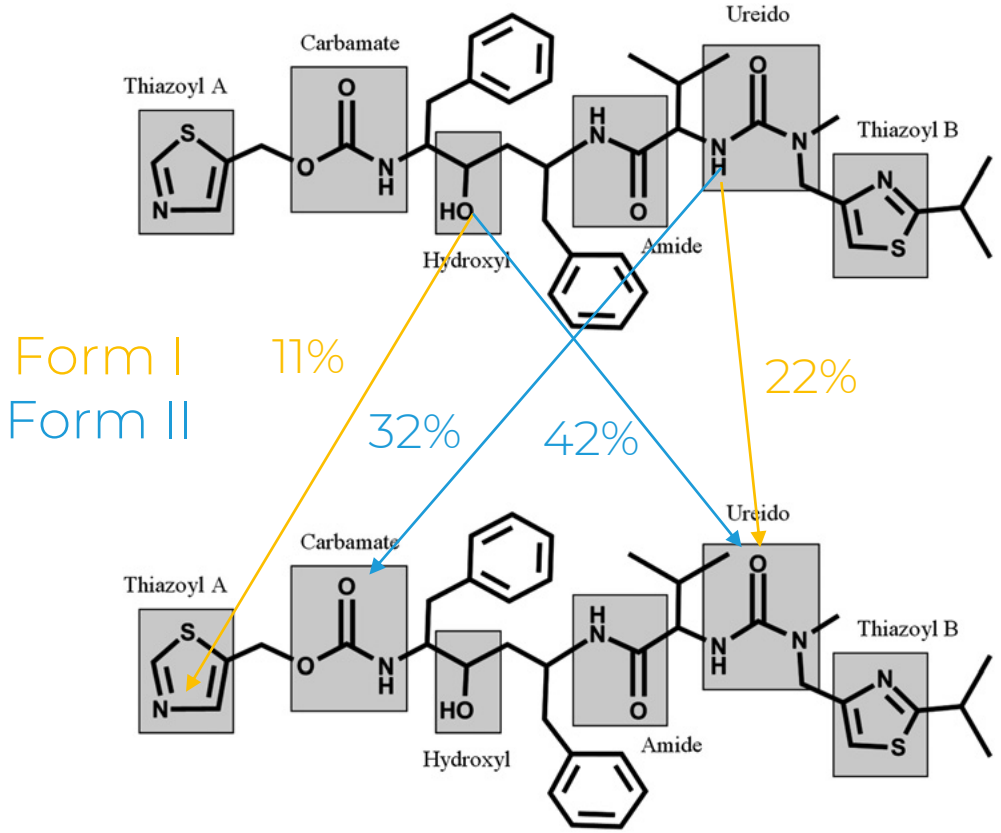
Great Wall of China c.1368

The database of walls indicates that A is the frequently observed arrangement and therefore we can predict it is the most stable form

CCDC

Predicting unlikely interactions

Predictive analytics is used to identify the likelihood of specific molecular interactions occurring from similar crystal structures



The integration of solid-form informatics into solid-form selection

Neil Feeder[a], Elna Pidcock[a], Anthony M. Reilly[a], Ghazala Sadiq[a], Cheryl L. Doherty[b], Kevin R. Back[b], Paul Meenan[c] and Robert Docherty[b]

One in half a million: a solid form informatics study of a pharmaceutical crystal structure

Peter T. A. Galek,*[a] Elna Pidcock,[a] Peter A. Wood,[a] Ian J. Bruno[a] and Colin R. Groom[a]

Navigating the Solid Form Landscape with Structural Informatics

Peter T. A. Galek, Elna Pidcock, Peter A. Wood, Neil Feeder, Frank H. Allen

Book Editor(s): Yuriy A. Abramov

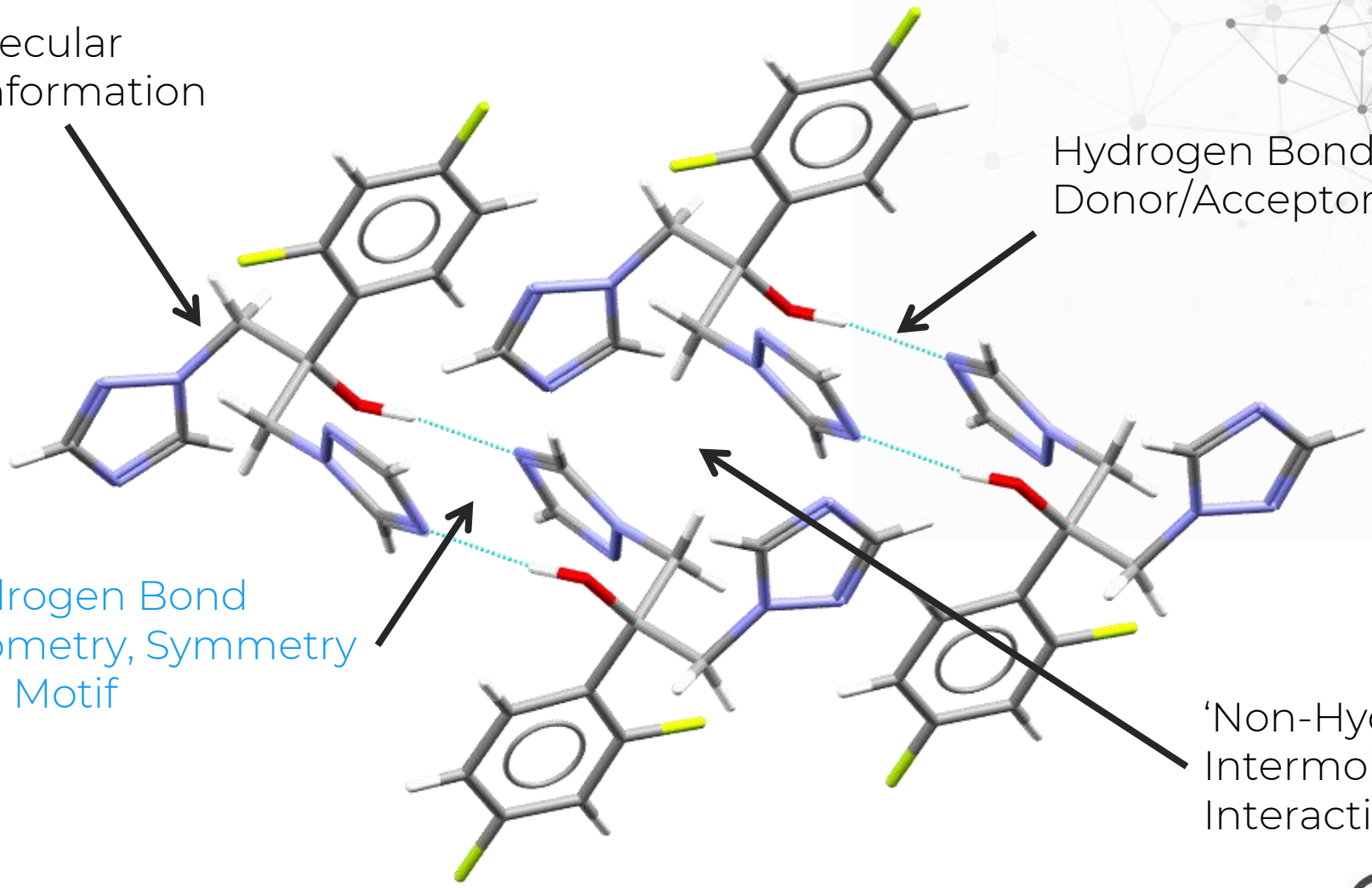
Knowledge-based H-bond prediction to aid experimental polymorph screening

Peter T. A. Galek,*[ab] Frank H. Allen,[a] László Fábián[ab] and Neil Feeder[c]

Galek *et al*, *CrystEngComm*, (2009), 11, 2634 - 2639

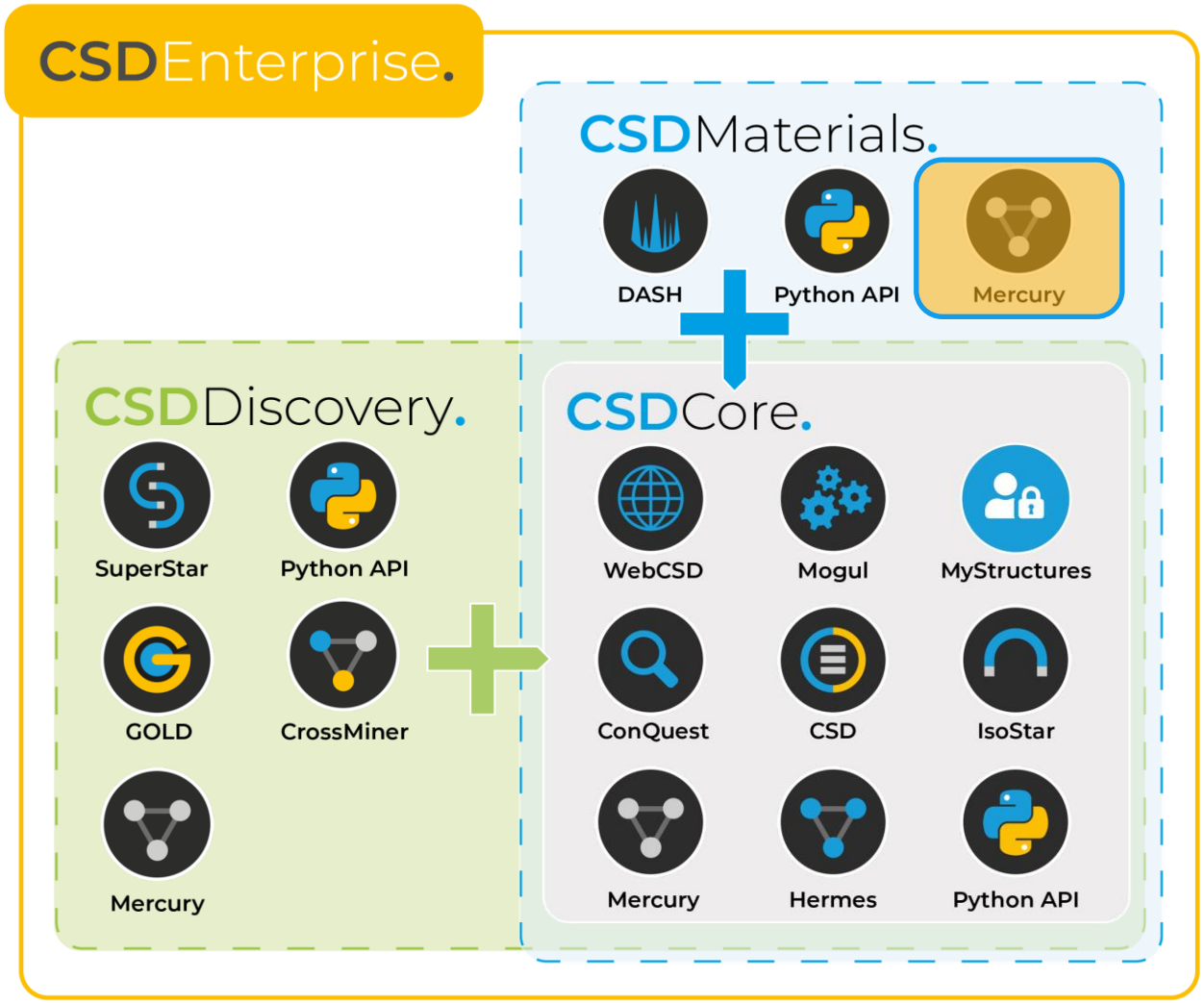
CCDC

Characteristics that influence stability



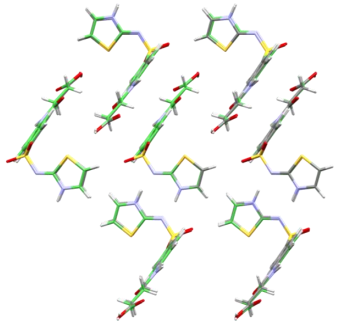
CCDC

The CSD software

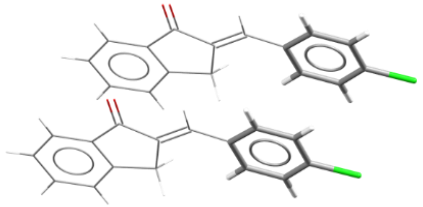


CCDC

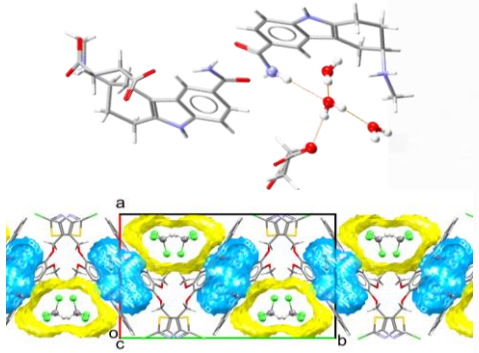
CSD-Materials overview



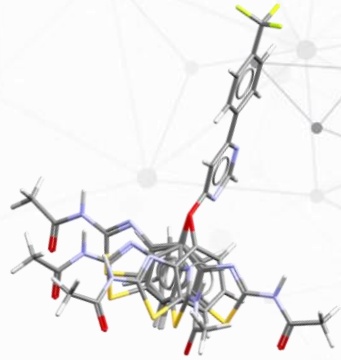
Crystal Packing Similarity



Motif Search & Packing Feature Search



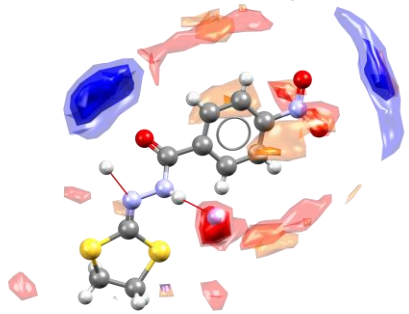
Hydrate Analyser & Solvate Analyser



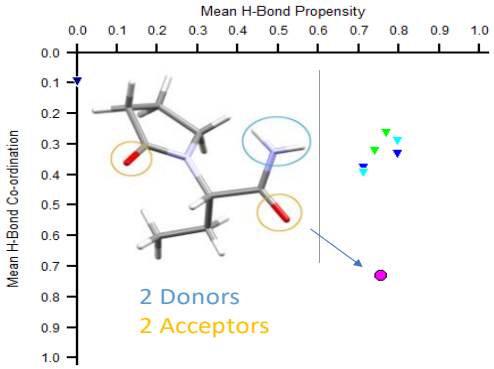
CSD Conformer Generator

Detailed Structural Analysis

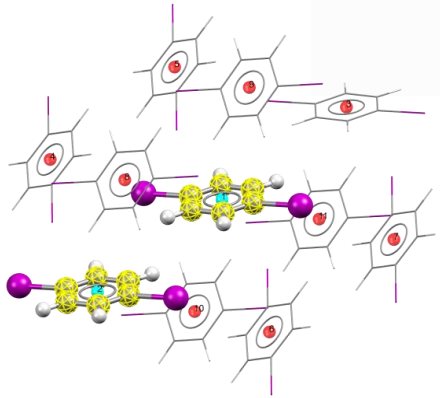
Solid Form Design



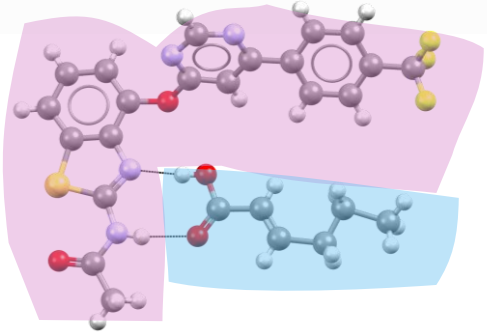
Full Interaction Maps



Hydrogen Bond Propensity



Aromatics Analyser



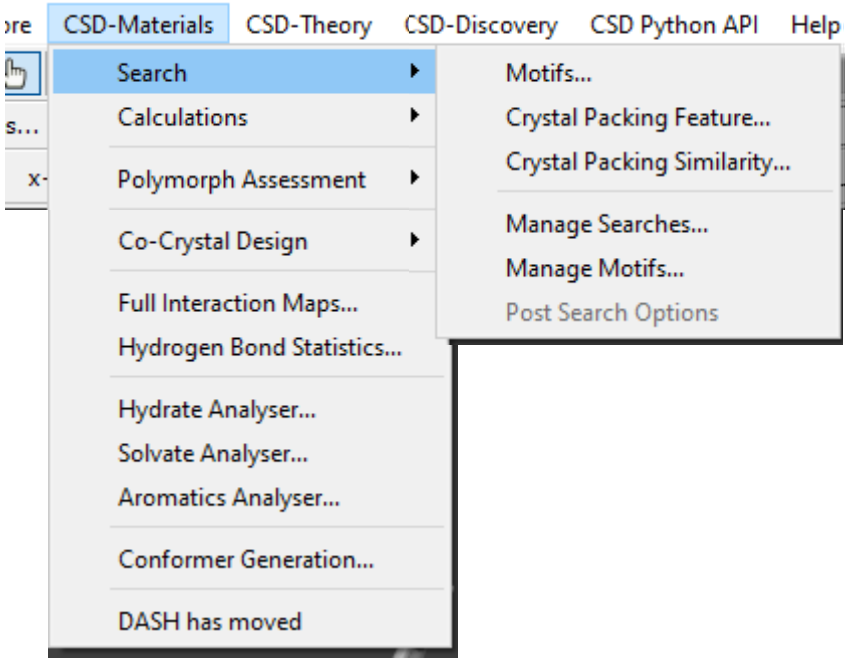
Molecular Complementarity

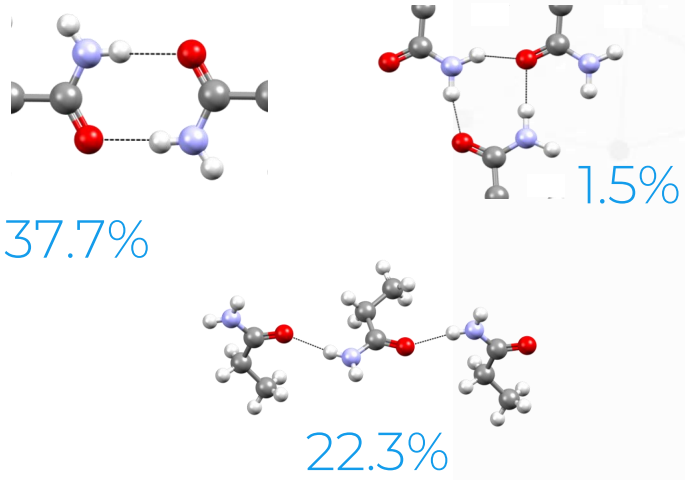
Solid Form Risk Assessment

CCDC

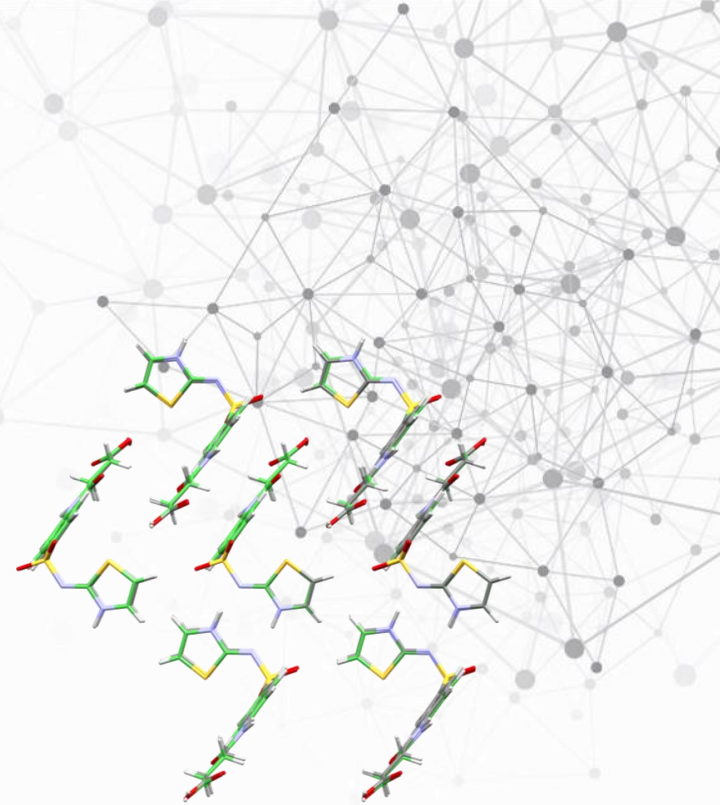
Overview

- Allow the exploration of aspects of a structure to determine how usual a feature is, in context of the CSD.
- Searches are easy to set up and return results in a useful format.

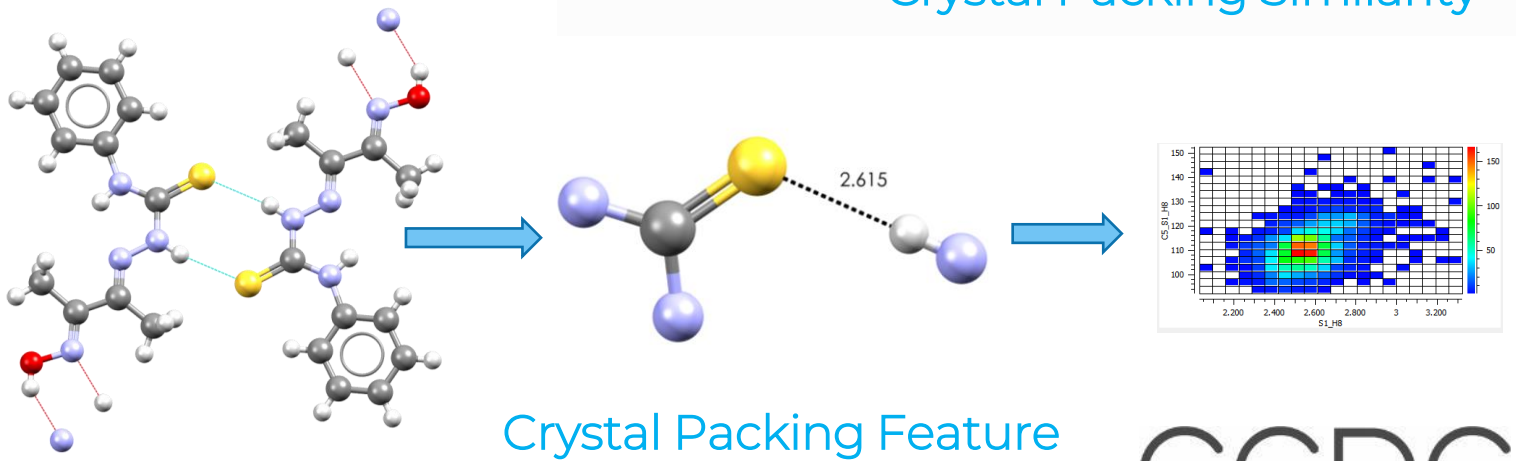




H-bond Motifs



Crystal Packing Similarity

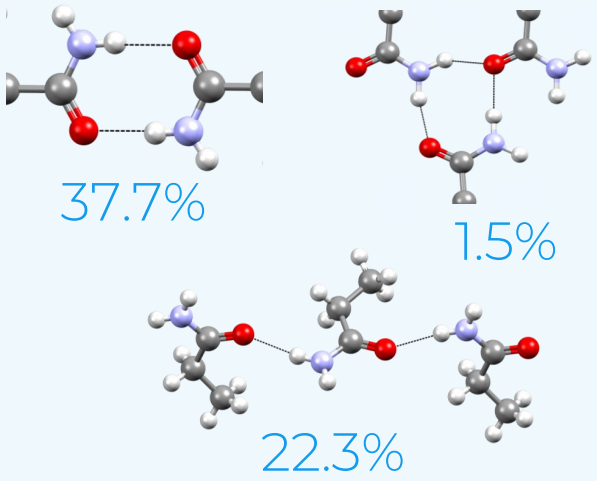


Crystal Packing Feature

CCDC

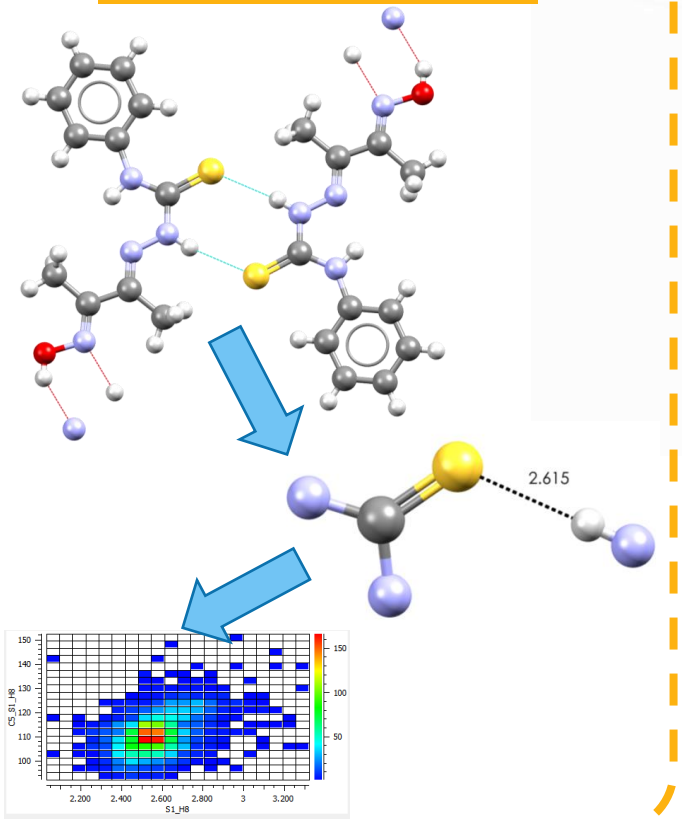
Searching motifs and crystal packing

Motifs

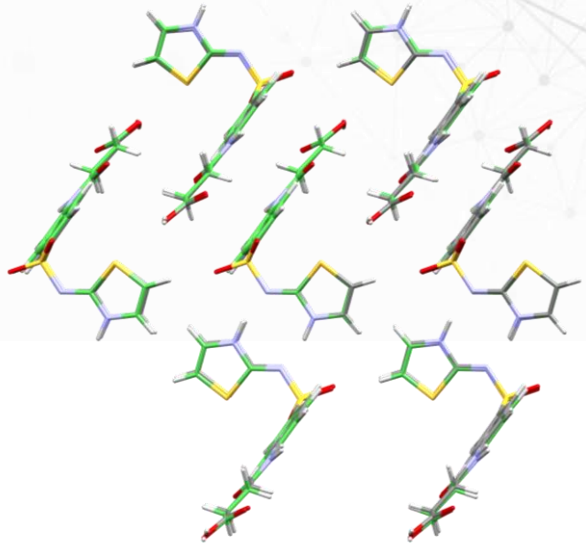


- H-bond motifs

Crystal Packing Feature



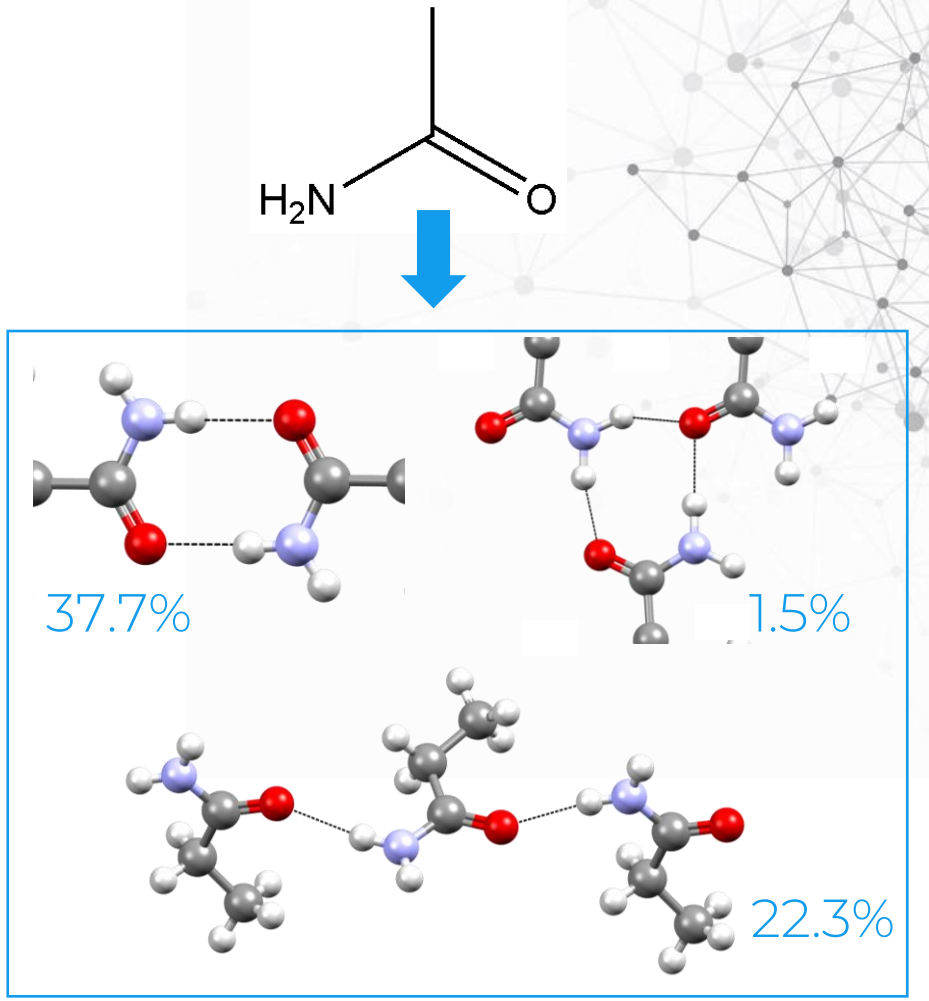
Crystal Packing Similarity



CCDC

Motifs

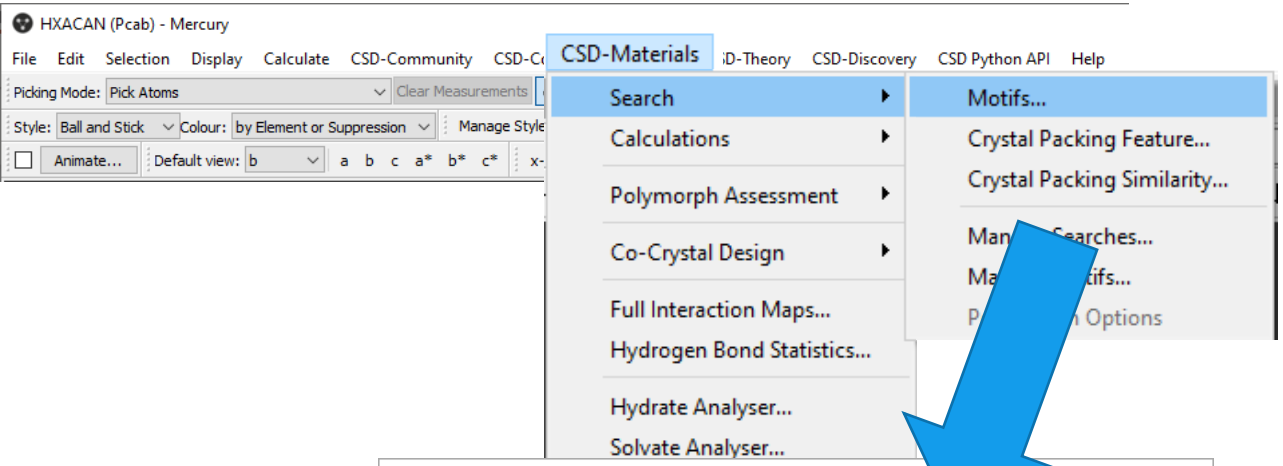
- Determine likely hydrogen bond motifs for a specified set of functional groups.
- Assess motifs by their relative frequency of occurrence in the CSD.
- Search for auto-generated and bespoke motifs.
- Analyse the results of crystal structure prediction runs by identifying the range of predicted motifs.

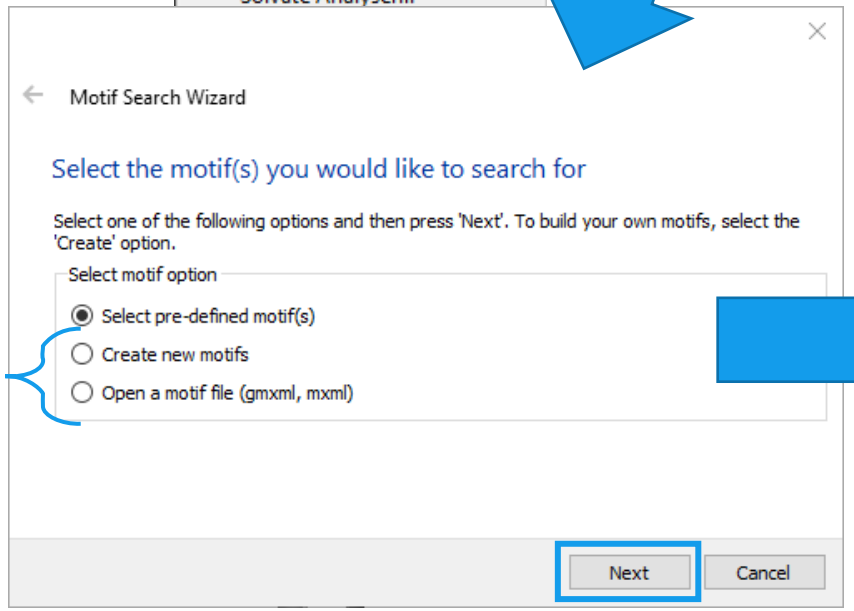


Frequency of occurrence of Primary Amide hydrogen bond motifs

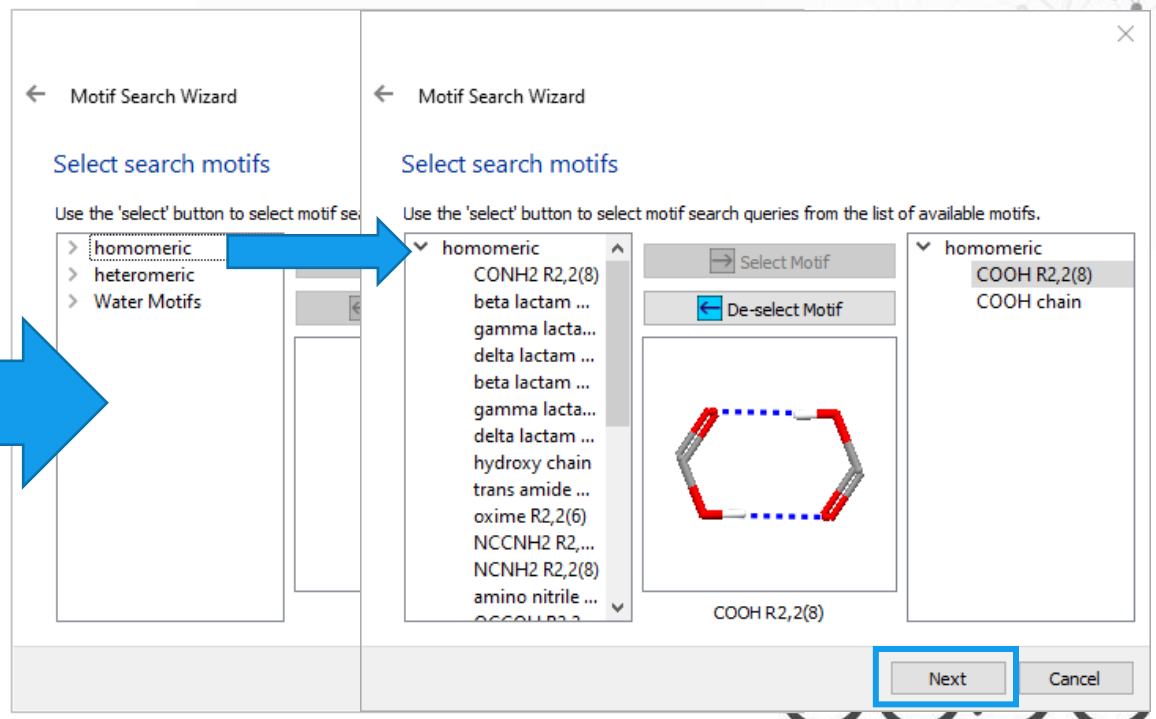
CCDC

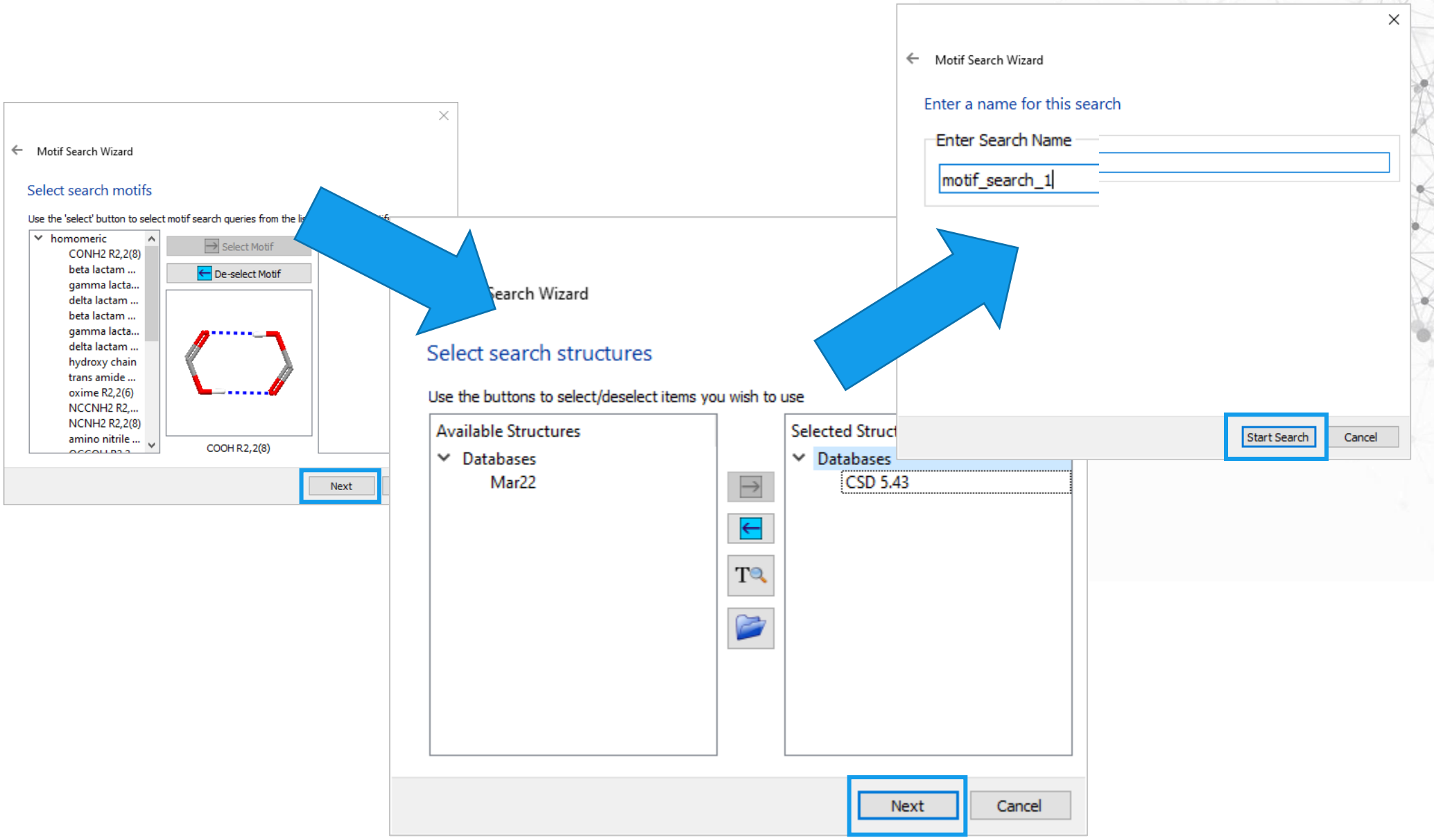
Searching for motifs

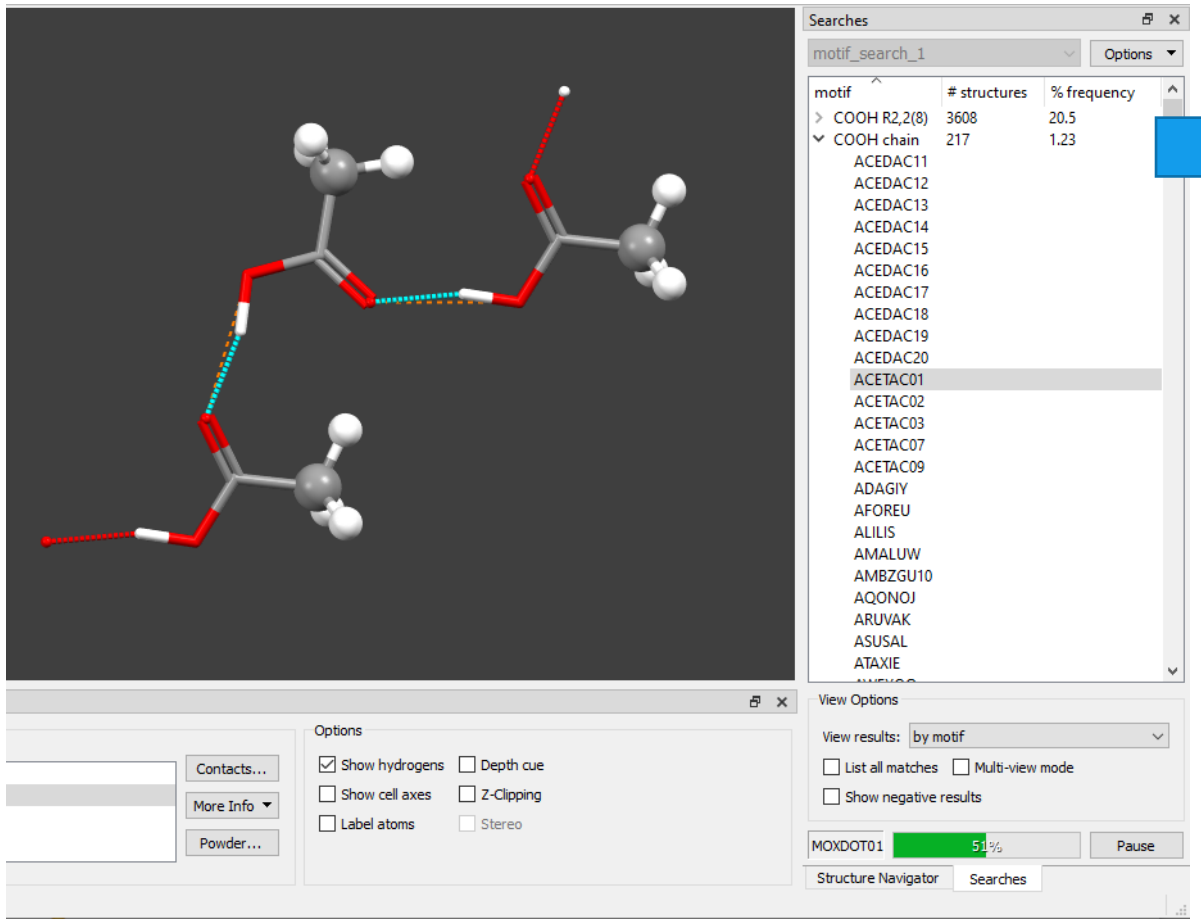


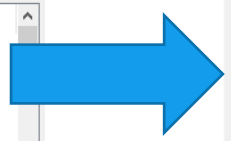


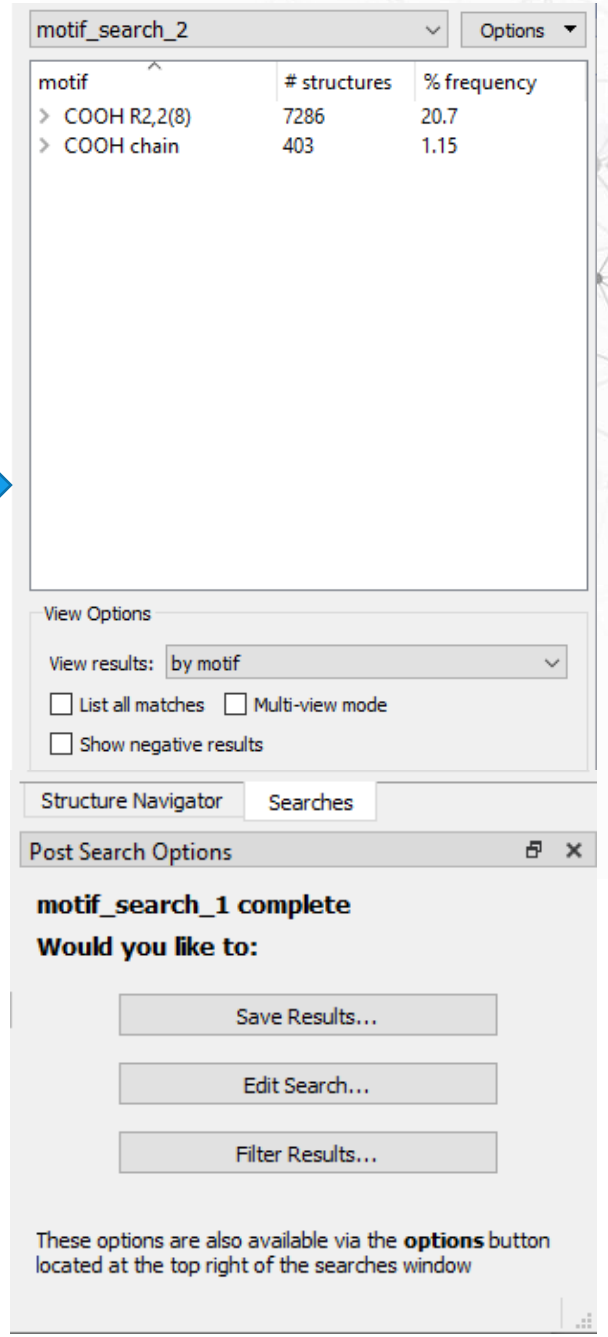
bespoke motifs




Motif Search Wizard
Select search motifs
Use the 'select' button to select motif search queries from the li
homomeric
CONH2 R2,2(8)
beta lactam ...
gamma lacta...
delta lactam ...
beta lactam ...
gamma lacta...
delta lactam ...
hydroxy chain
trans amide ...
oxime R2,2(6)
NCCNH2 R2,...
NCNH2 R2,2(8)
amino nitrile ...
Select Motif
De-select Motif
COOH R2,2(8)
Next
Search Wizard
Select search structures
Use the buttons to select/deselect items you wish to use
Available Structures
Databases
Mar22
Selected Struct
Databases
CSD 5.43
Next
Cancel
Motif Search Wizard
Enter a name for this search
Enter Search Name
motif_search_1
Start Search
Cancel


Searches
motif_search_1
Options
motif
structures
% frequency
COOH R2,2(8) 3608 20.5
COOH chain 217 1.23
ACEDAC11
ACEDAC12
ACEDAC13
ACEDAC14
ACEDAC15
ACEDAC16
ACEDAC17
ACEDAC18
ACEDAC19
ACEDAC20
ACETAC01
ACETAC02
ACETAC03
ACETAC07
ACETAC09
ADAGIY
AFOREU
ALILIS
AMALUW
AMBZGU10
AQONOJ
ARUVAK
ASUSAL
ATAXIE
Contacts...
More Info
Powder...
Options
Show hydrogens
Depth cue
Show cell axes
Z-Clipping
Label atoms
Stereo
View Options
View results: by motif
List all matches
Multi-view mode
Show negative results
MOXDOT01
51%
Pause
Structure Navigator
Searches

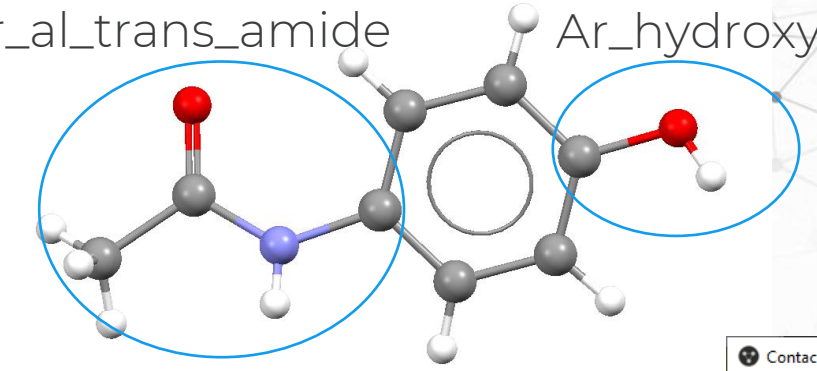


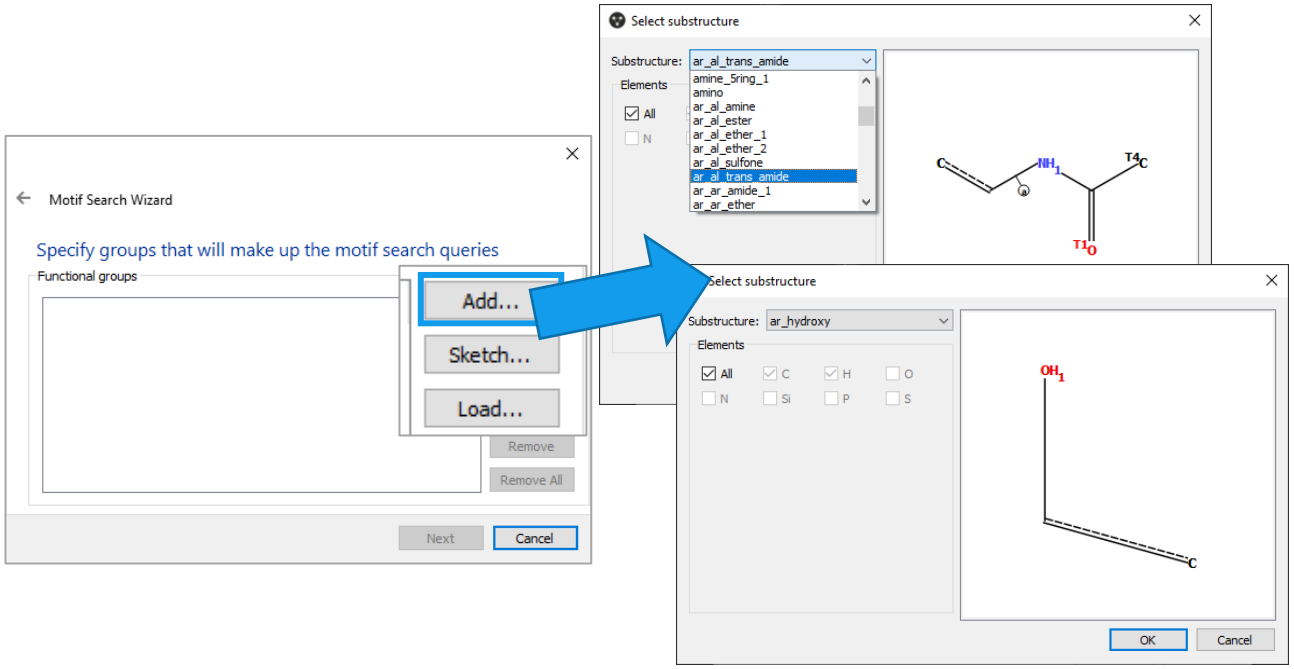

motif_search_2
Options
motif
structures
% frequency
COOH R2,2(8) 7286 20.7
COOH chain 403 1.15
View Options
View results: by motif
List all matches
Multi-view mode
Show negative results
Structure Navigator
Searches
Post Search Options
motif_search_1 complete
Would you like to:
Save Results...
Edit Search...
Filter Results...
These options are also available via the options button located at the top right of the searches window

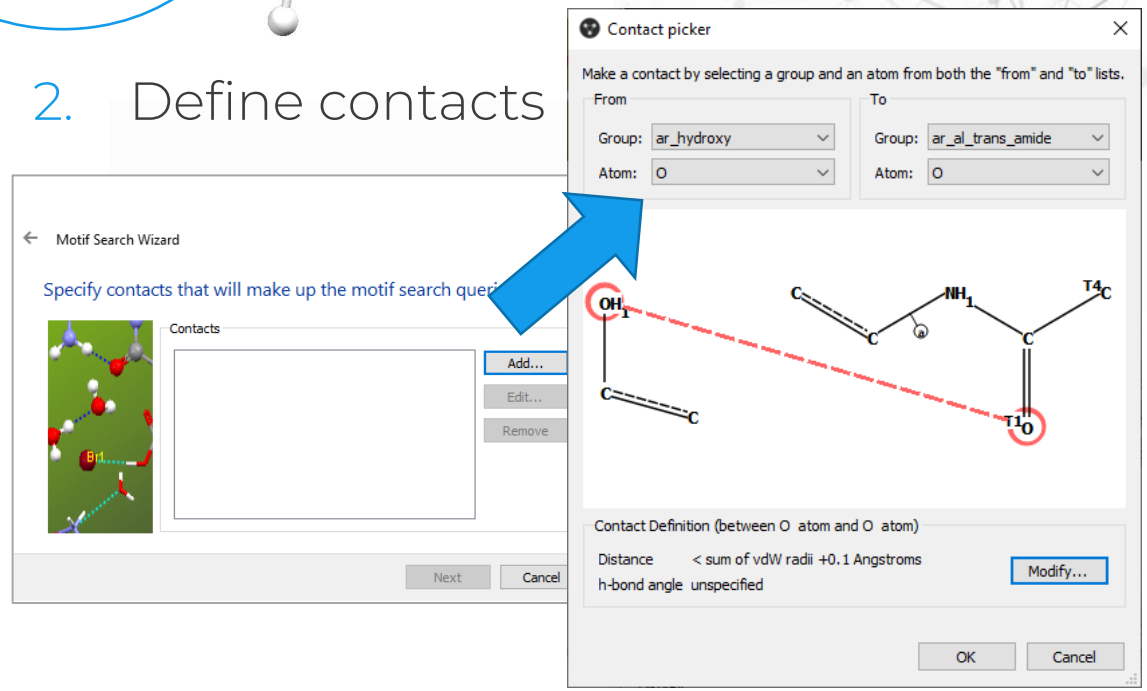
CCDC

Searching for a bespoke motif

- If a motif is not in the pre-defined list you can create it.
- Additional steps:
  1. Specify groups involved
  2. Define contacts

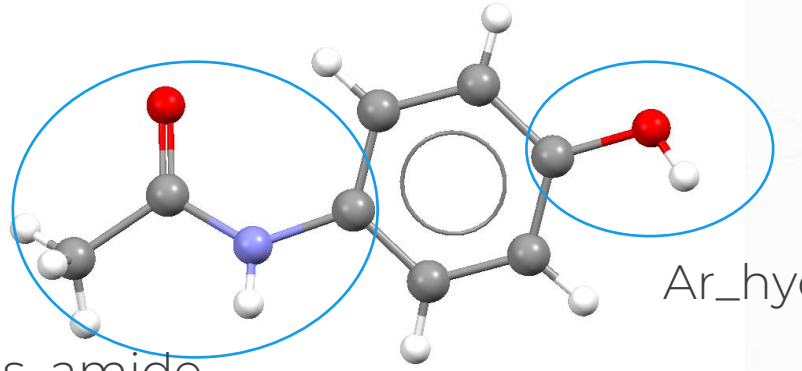




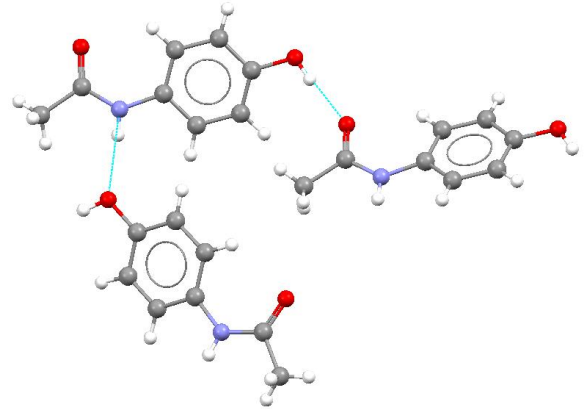


Investigating intermolecular interactions

- HXACAN



Donor	Acceptor	Hits	Frequency of Occurrence
Ar_hydroxy	Ar_al_trans_amide	79	41.6
Ar_hydroxy	Ar_hydroxy	3413	12.3
Ar_al_trans_amide	Ar_hydroxy	37	19.5
Ar_al_trans_amide	Ar_al_trans_amide	1131	32.8



*Searched best representative list

CCDC

Investigating possible co-formers

For the *pure* form, the frequency of occurrence of interactions:

NH of trans amide to OH	20%	Trans-amide to trans-amide	33%
OH to C=O trans amide	42%	OH to OH	12%

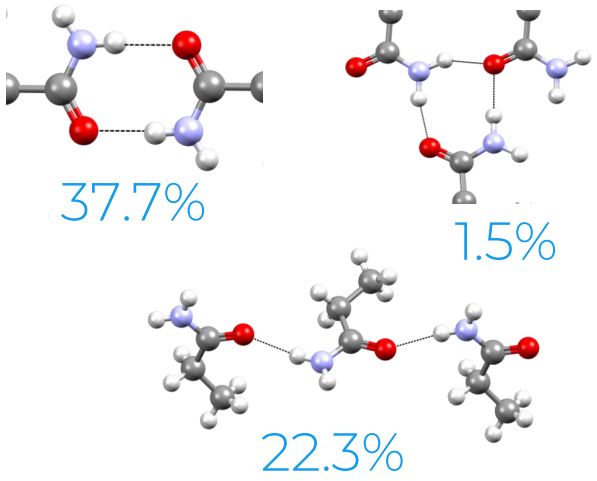
Look at possible co-former groups: COOH, aromatic_nitrogen, saturated ring NH

motif	# structure	% frequency	Observed cocrystals
inter_A-B- (A=NH1 of ar_al_trans_amide B=O of al_cooh_1)	60	35.7	0
inter_A-B- (A=NH1 of ar_al_trans_amide B=N of aromatic_nitrogen)	46	27.4	2
inter_A-B- (A=NH1 of ar_al_trans_amide B=NH1 of saturated_ring_NH)	3	9.09	2
inter_A-B- (A=O of al_cooh_1 B=OH1 of ar_hydroxy)	135	25.1	1
inter_A-B- (A=N of aromatic_nitrogen B=OH1 of ar_hydroxy)	1252	63.8	12
inter_A-B- (A=NH1 of saturated_ring_NH B=OH1 of ar_hydroxy)	93	44.1	4

CCDC

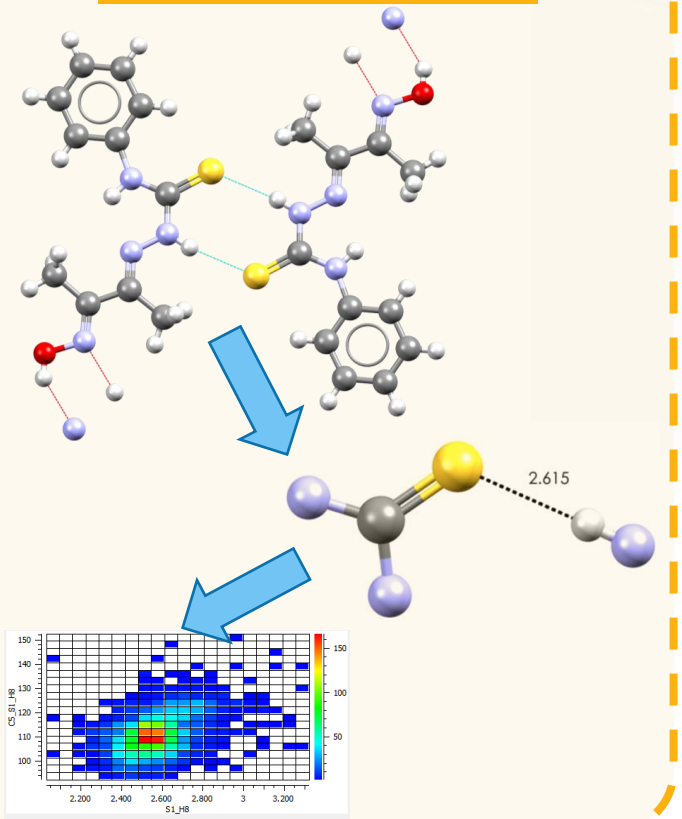
Searching motifs and crystal packing

Motifs

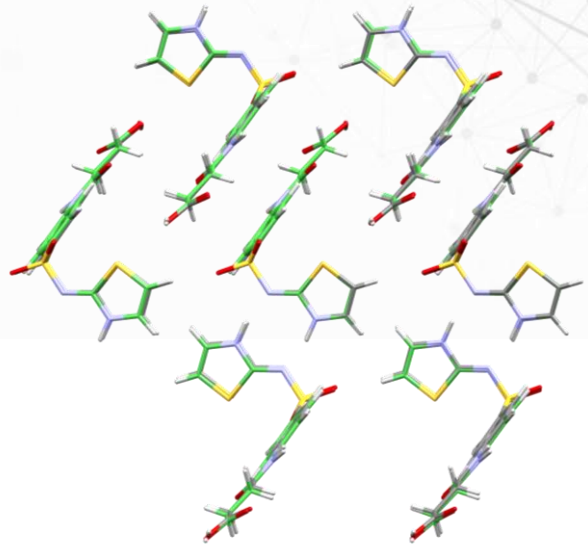


- H-bond motifs

Crystal Packing Feature



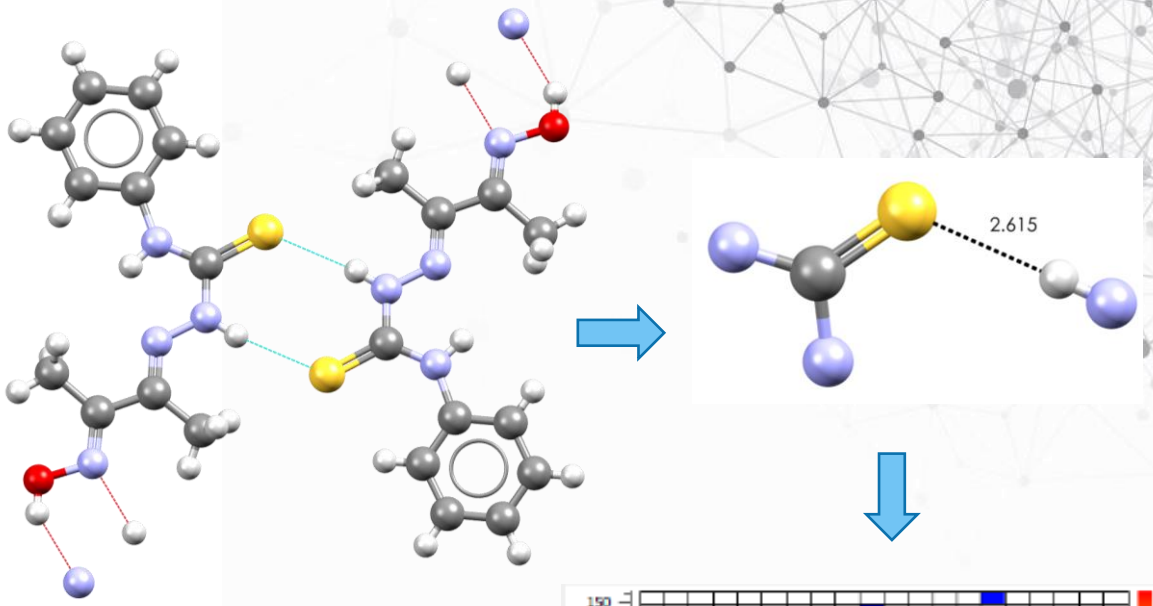
Crystal Packing Similarity



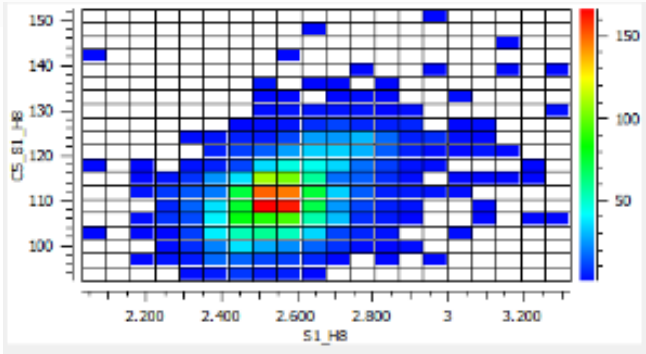
CCDC

Crystal Packing Feature Search

- Perform a substructure search
- Investigate conformations of molecules or bonded fragments
- Search for non-covalent interactions such as π-π or hydrogen bond interactions
- Search for particular spatial arrangements of functional groups
- Search for particular spatial arrangements of molecules

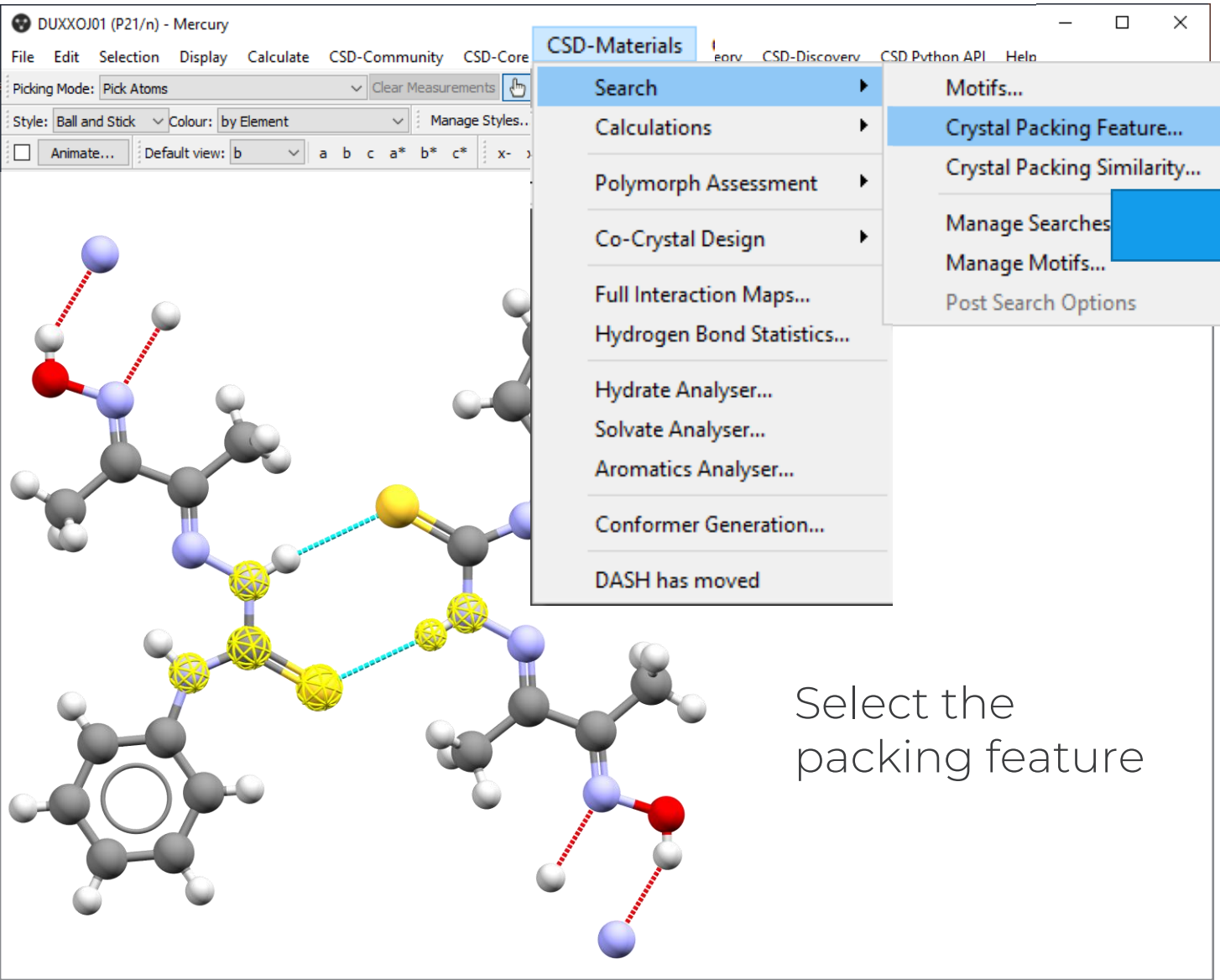


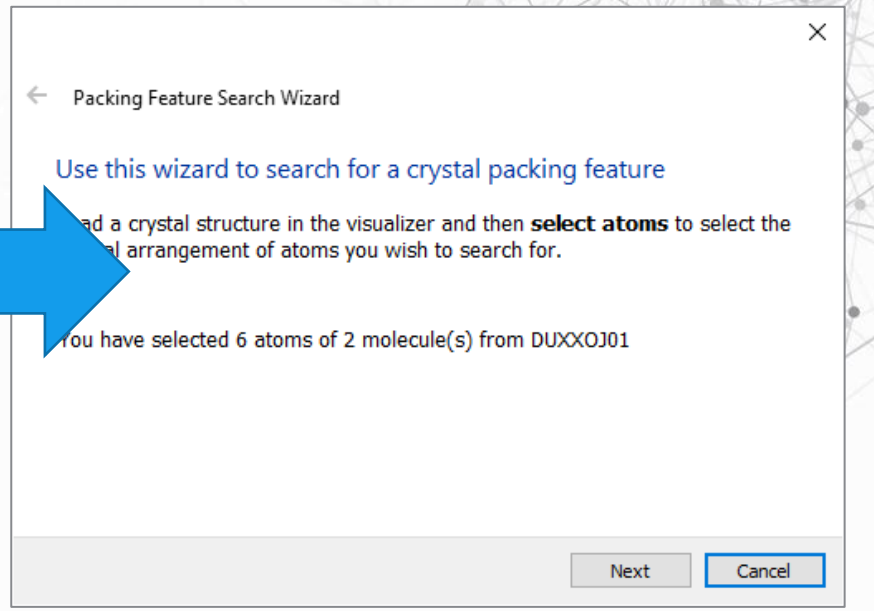
Query for S...H-N interaction as found in a thioamide derivative (DUXXOJ01)



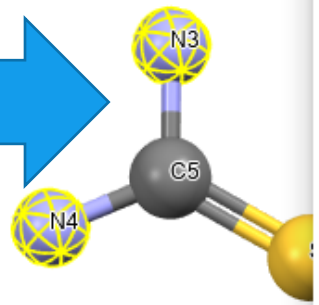
CCDC

Searching for packing features





Select the packing feature



Packing Feature Search Wizard

Allow variable atom and bond types

Select the atoms and bonds you wish to vary and press 'Modify'. Or simply press 'Next'.

atoms | bonds

atom	residue	elements	hydroger	bonds
N4	1	N	1	3
N3	1	N	1	3
C5	1	C	0	3
H8	2	H		1
N3	2	N	1	3
S1	1	S	0	1

Search Options

Modify

- Element
- Hydrogens
- Charge
- Number of bonded atoms
- Cyclicity

Charge

Cyclicity

Next | Cancel

- C
- H
- O
- N
- S
- P
- F
- Cl
- Any
- More
- Other Elements...

- Any Halogen
- Any Metal
- Any Non Metal
- Any Transition Metal
- C or H
- C or N
- Not Hydrogen
- O or S
- Other Elements...

Packing Feature Search Wizard

Allow tolerances on the geometry of the packing feature

Level of Geometric Similarity Required

Very high | High | Medium | Low | Custom

Custom Tolerances

Distances must match those of selected feature to within + or - 30 %

Angles must match those of selected feature to within + or - 20 degrees

Match Options

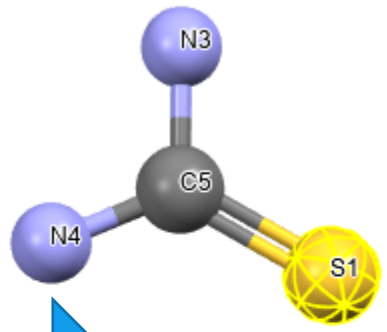
All matched fragments must belong to same molecule

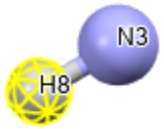
Show distances | Show angles

Next | Cancel

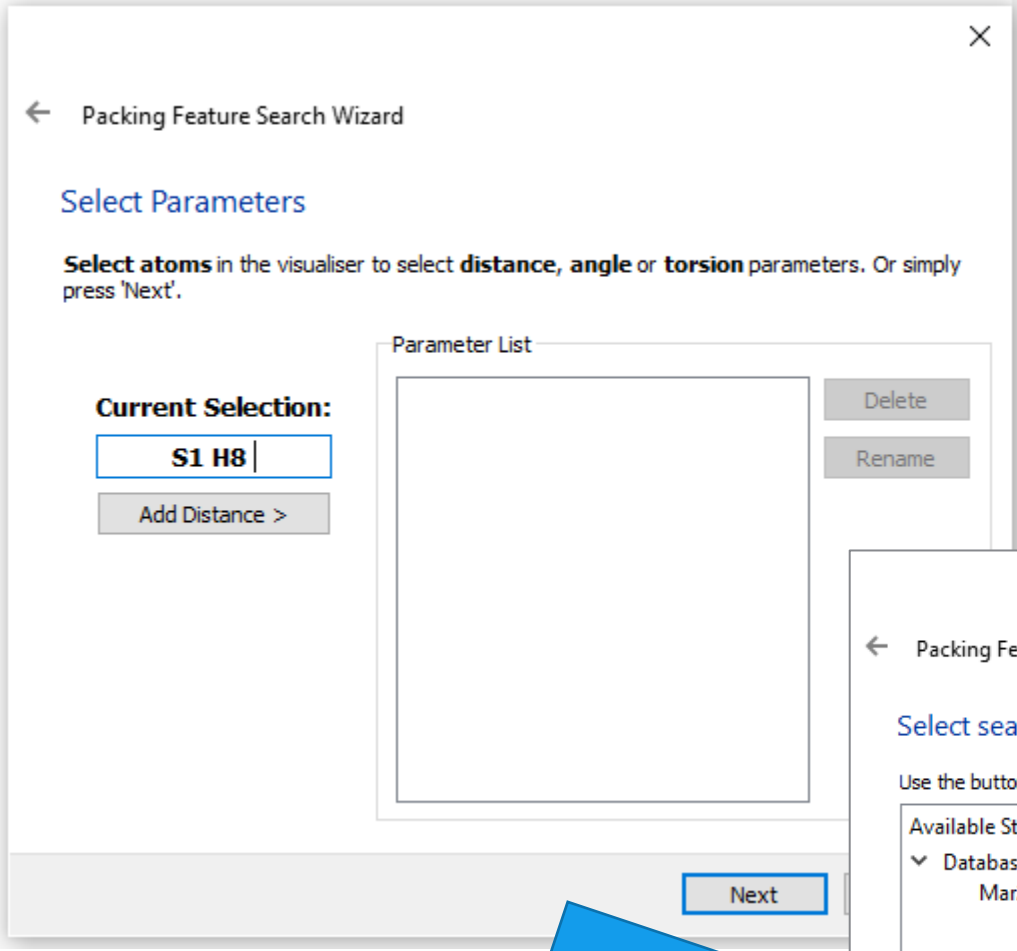
You may set some atoms or bonds to be variable. E.g., the selected ones could be either a C or an N

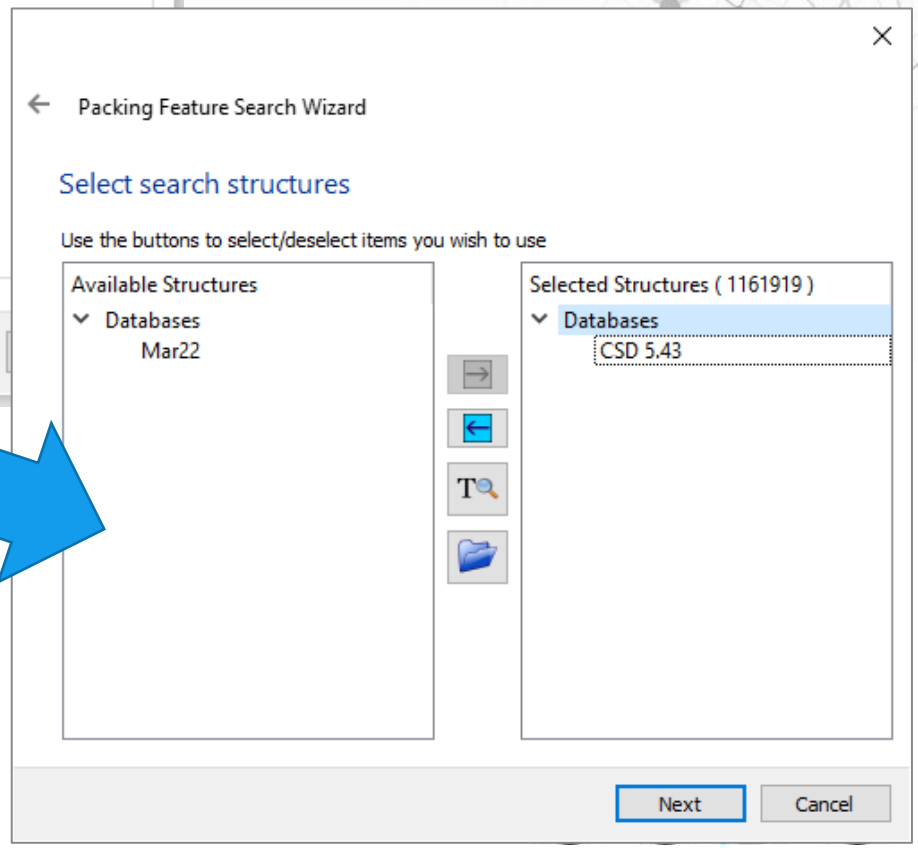
CCDC



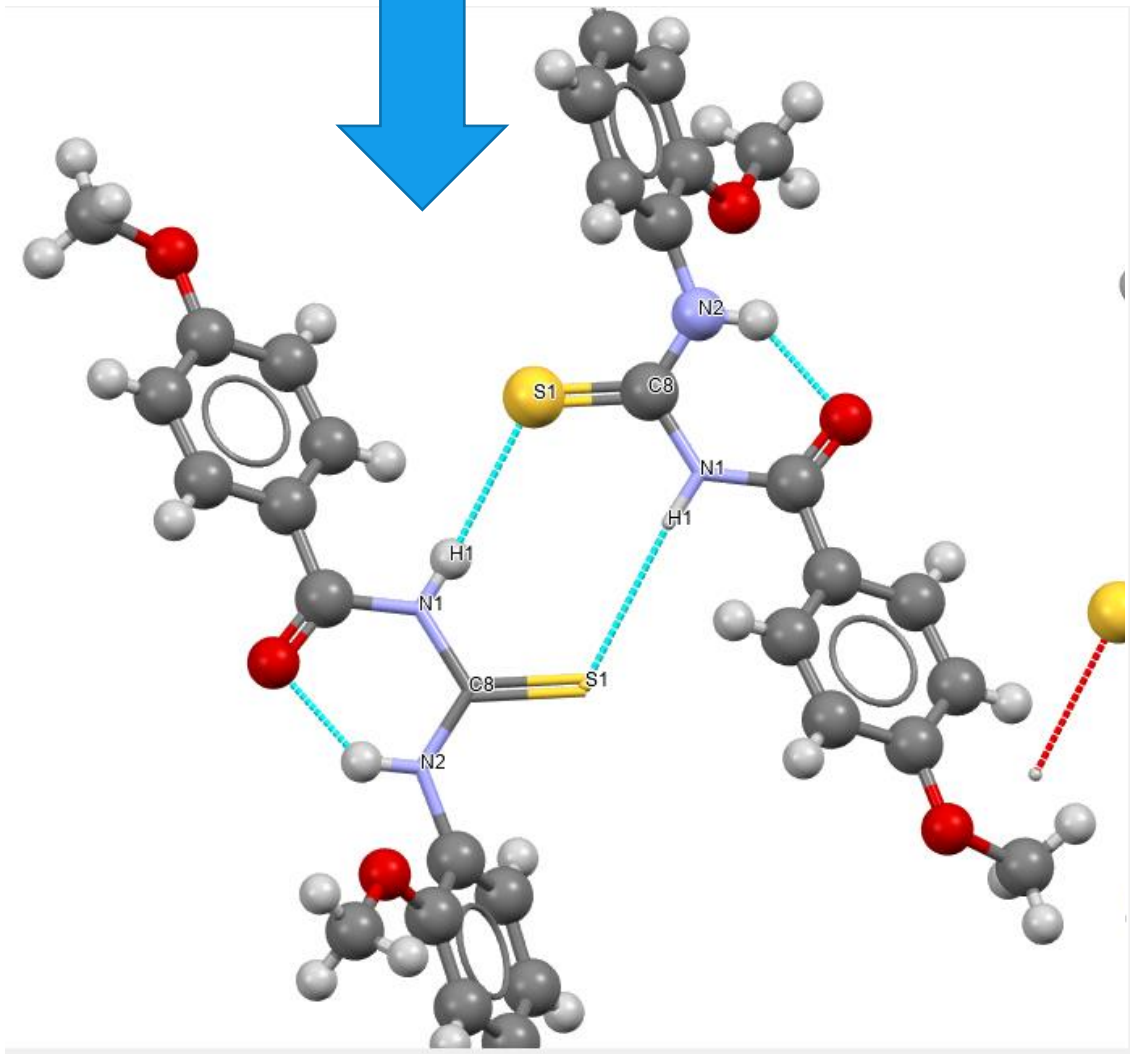


You may set to monitor some parameters such as distances, angles and torsions





Select filters and name the run and start the search!



Searches

Packing Feature_feature_searcl | Options

1762 hits	RMS	S1_H8
ABEFAP	0.048	2.70176
ABESUW	0.148	2.51547
ABOVEW	0.118	2.58532
ABOVOG	0.133	2.56231
ACAFEQ	0.253	2.47826
ACAFEQ01	0.156	2.54067
ACAFIU	0.245	2.50152
ACAFUG	0.24	2.64227
ACESOR	0.231	2.50905
ACEXIQ	0.181	2.69929
ACIZAP	0.146	2.65814
ADAHET	0.192	2.64739
ADAVOR	0.147	2.61819
ADEZAL	0.187	2.54114
ADEZAL02	0.198	2.52302
ADISIR	0.077	2.77197
AFUGOZ	0.236	2.59472
AFULOE	0.1	2.60527
AGAVIP	0.209	2.57117

View Options

- Show packing feature
- Show negative results
- List all matches

Spreadsheet...

Structure Navigator | Searches

Post Search Options

Packing Feature_feature_search complete

Would you like to:

Analyse the results

Data Analysis

File Options

Packing Feature_feature_search Spreadsheet 1

File Tools Descriptors Display Selection Plots Statistics

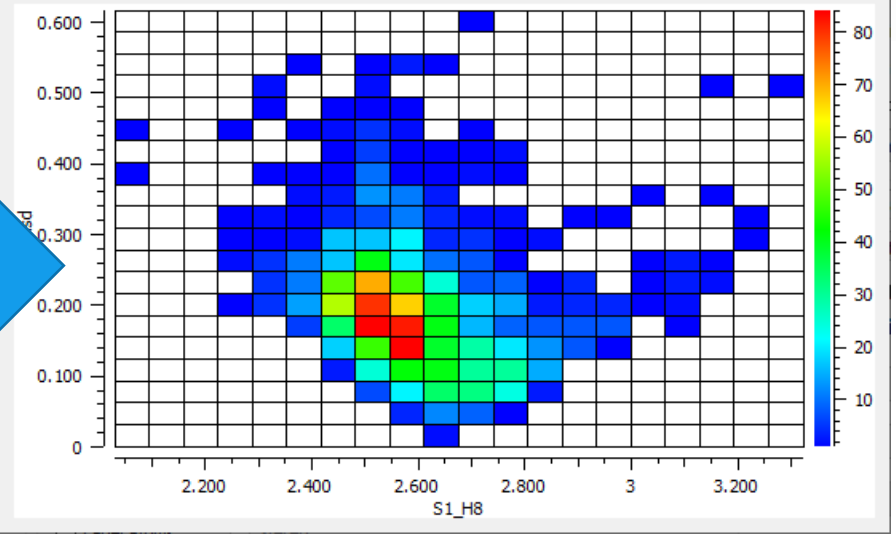
- Histogram
- Polar histogram
- Scatterplot
- Polar scatterplot
- Heat plot

Find identifier

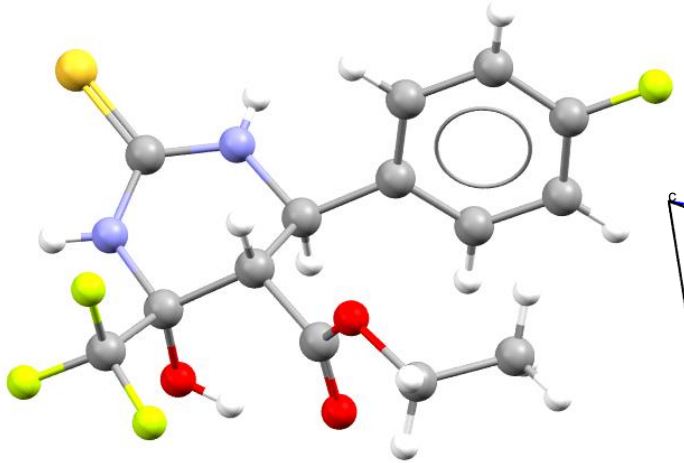
Identifier			
Packing Feature_feature_search|ABEFAP|0	ABEFAP	0.0480	2.7018
Packing Feature_feature_search|ABESUW|1	ABESUW	0.1480	2.5155
Packing Feature_feature_search|ABOVEW|2	ABOVEW	0.1180	2.5853
Packing Feature_feature_search|ABOVOG|3	ABOVOG	0.1330	2.5623
Packing Feature_feature_search|ACAFEQ|4	ACAFEQ	0.2530	2.4783
Packing Feature_feature_search|ACAFEQ01|5	ACAFEQ01	0.1560	2.5407
Packing Feature_feature_search|ACAFIU|6	ACAFIU	0.2450	2.5015
Packing Feature_feature_search|ACAFUG|7	ACAFUG	0.2400	2.6423
Packing Feature_feature_search|ACESOR|8	ACESOR	0.2310	2.5090
Packing Feature_feature_search|ACEXIQ|9	ACEXIQ	0.1810	2.6993
Packing Feature_feature_search|ACIZAP|10	ACIZAP	0.1460	2.6581

Packing Feature_feature_search Heat plot 1

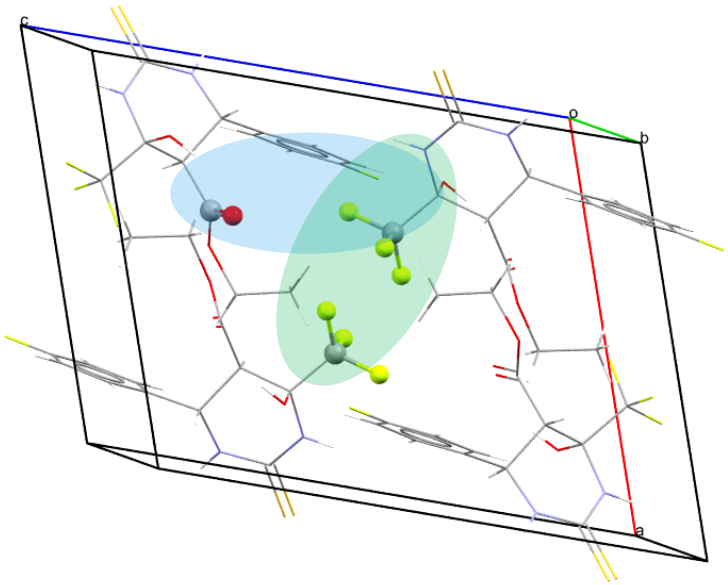
File Axes Mouse Display Selection Plots Statistics

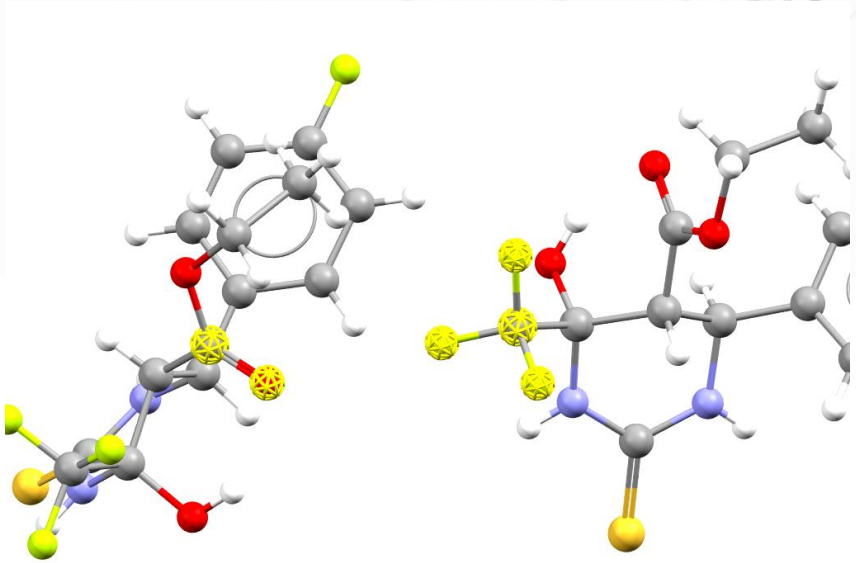


Interactions that are not H-bonds



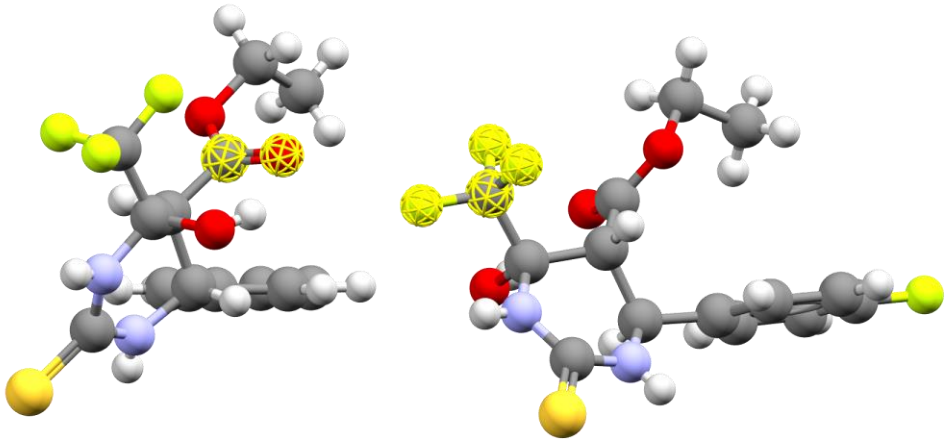
LAQKEV





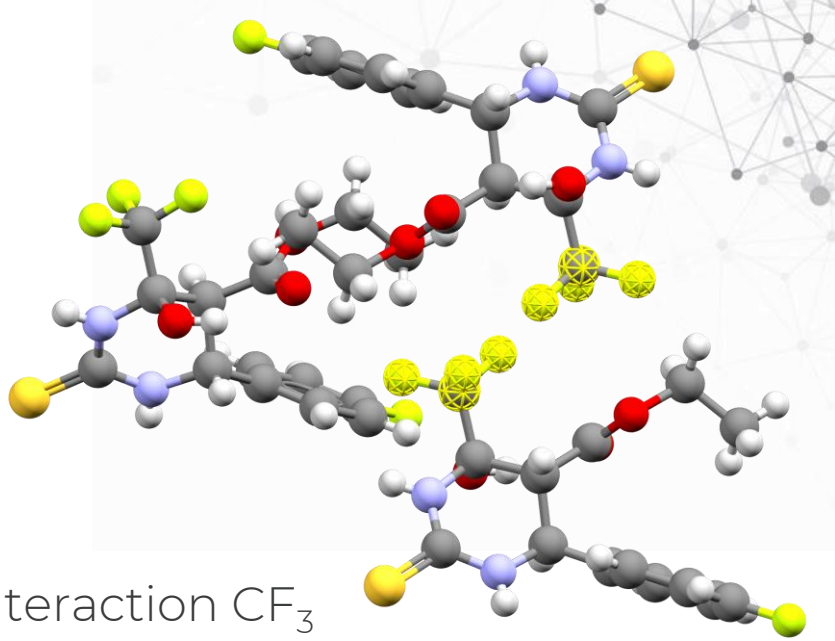
CCDC

Interactions that are not H-bonds



CF_3 interaction with carbonyl
Packing Feature search
367 hits (5735 structures searched) 6.4%

Rf values for F….O interactions ~0.3



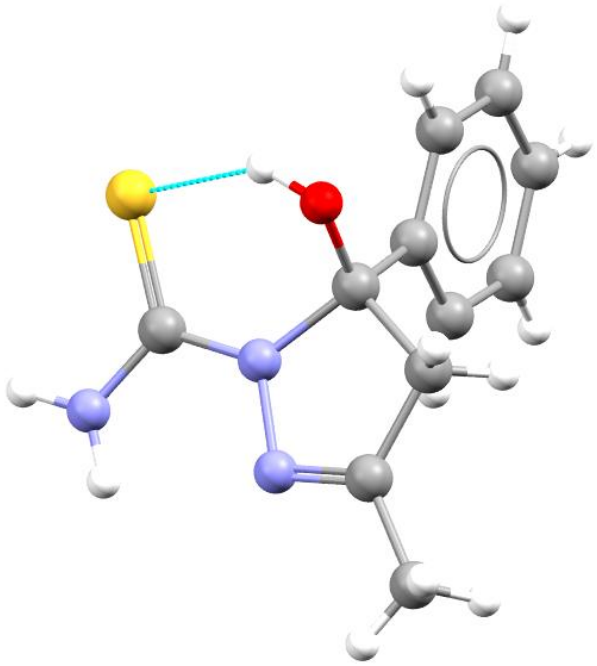
CF_3 interaction CF_3
Packing Feature search
2114 hits (12406 structures searched) 17.0%

Rf value for CF_3-CF_3 up to 1.26

Cole, J. C.; Taylor, R. Intermolecular Interactions of Organic Fluorine Seen in Perspective. *Crystal Growth & Design* **2022**, *22* (2), 1352–1364. https://doi.org/10.1021/acs.cgd.1c01315.

CCDC

Intramolecular geometry

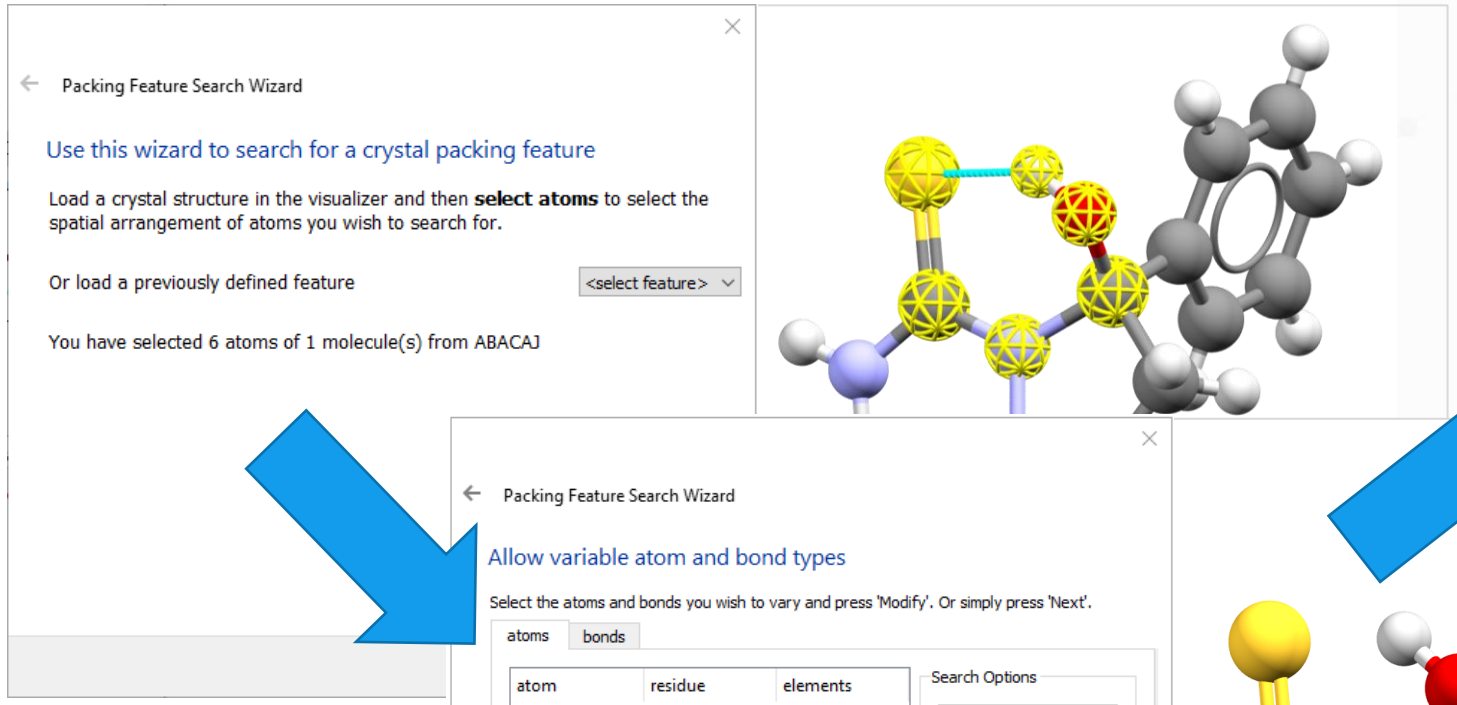


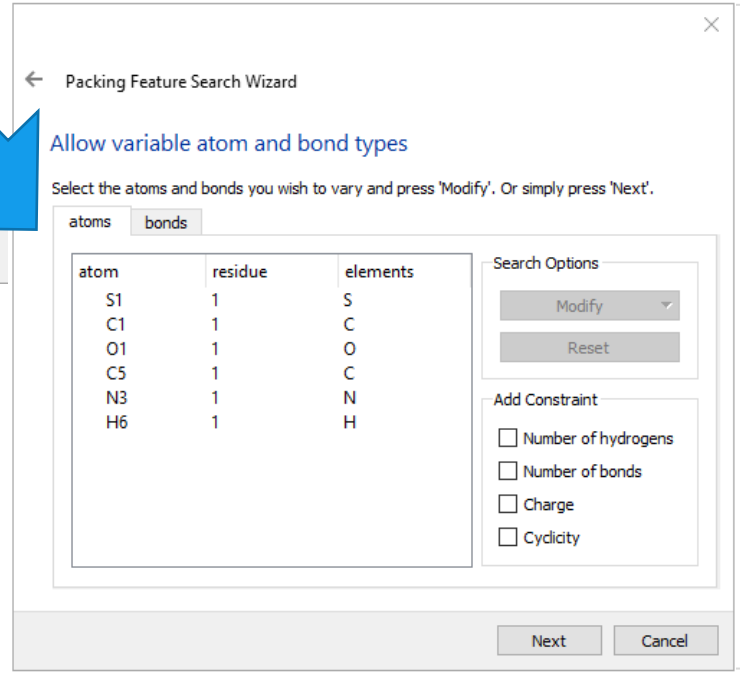
- Intramolecular bond
- How common is this?

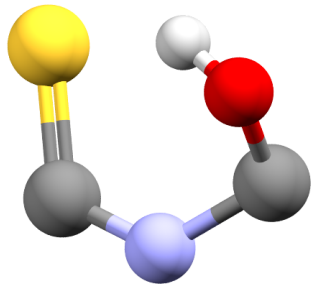
ABACAJ

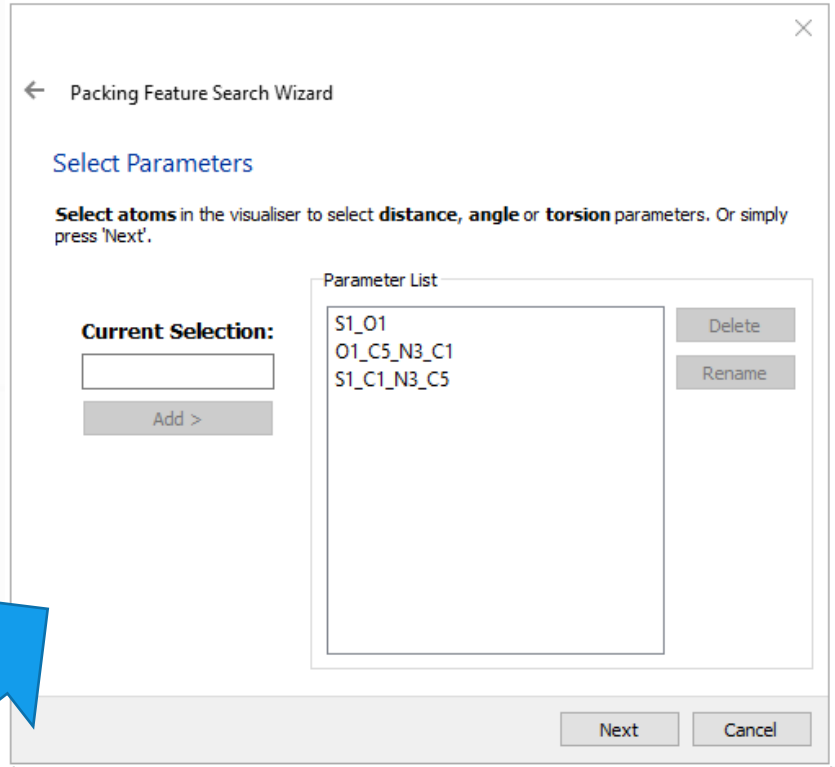
CCDC

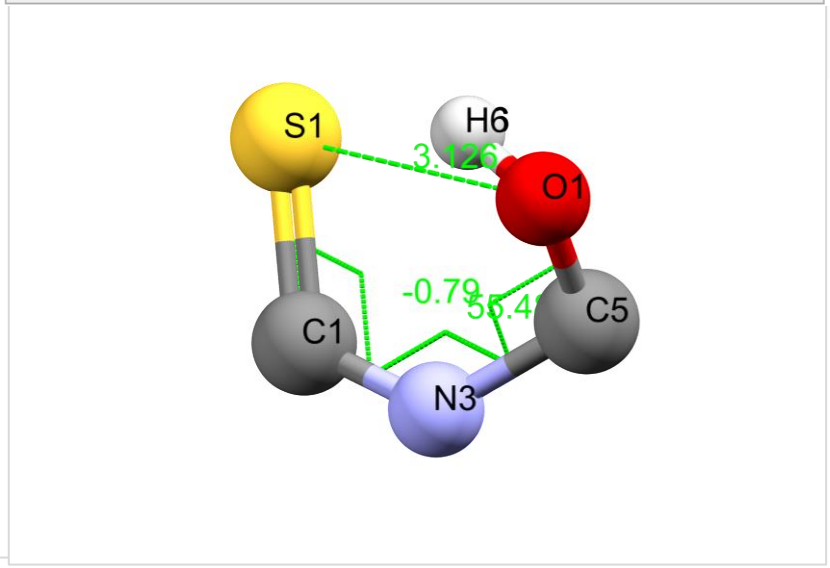
Intramolecular geometry

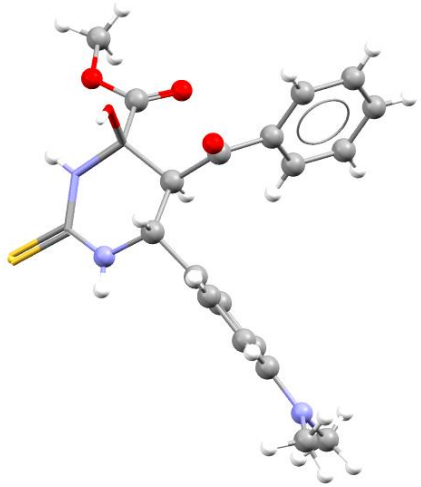


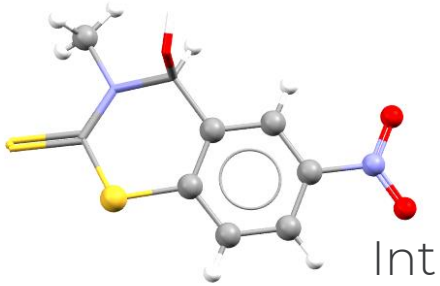




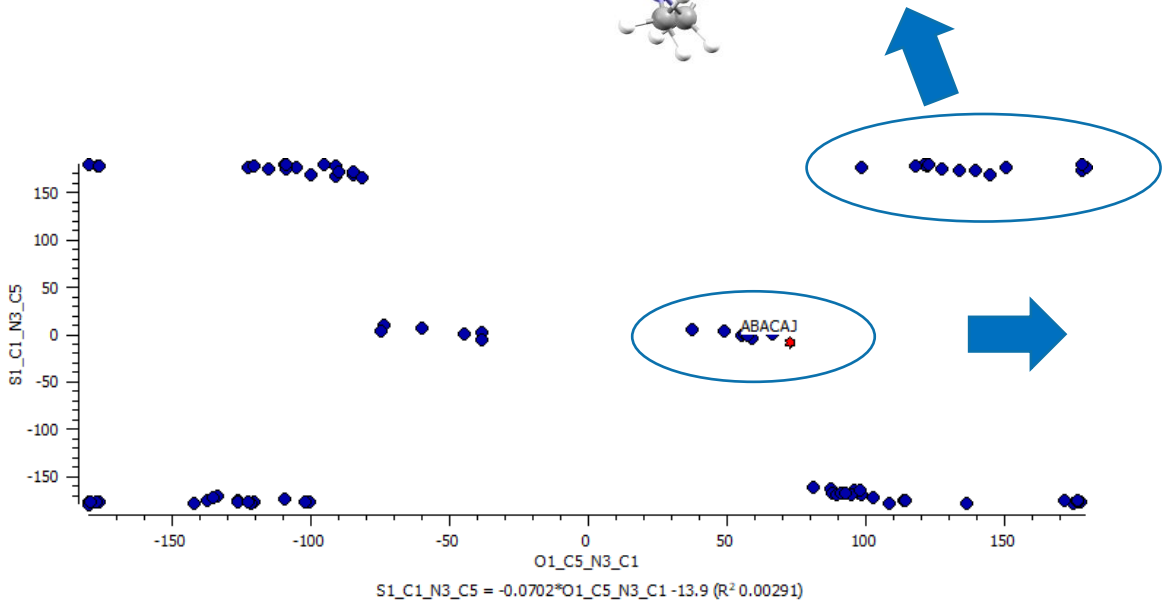


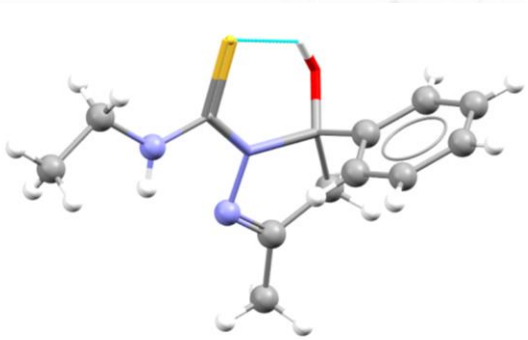


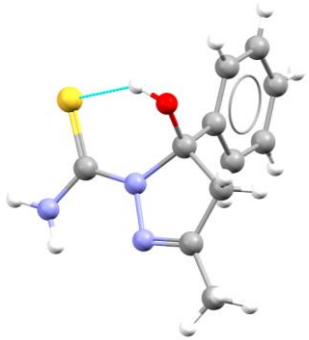




Intramolecular interaction not possible due to chemistry of fragment





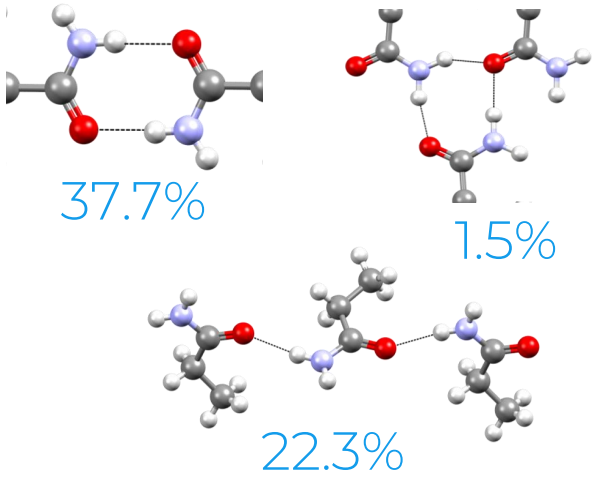


When physically allowed, intra H bond, observed frequently

CCDC

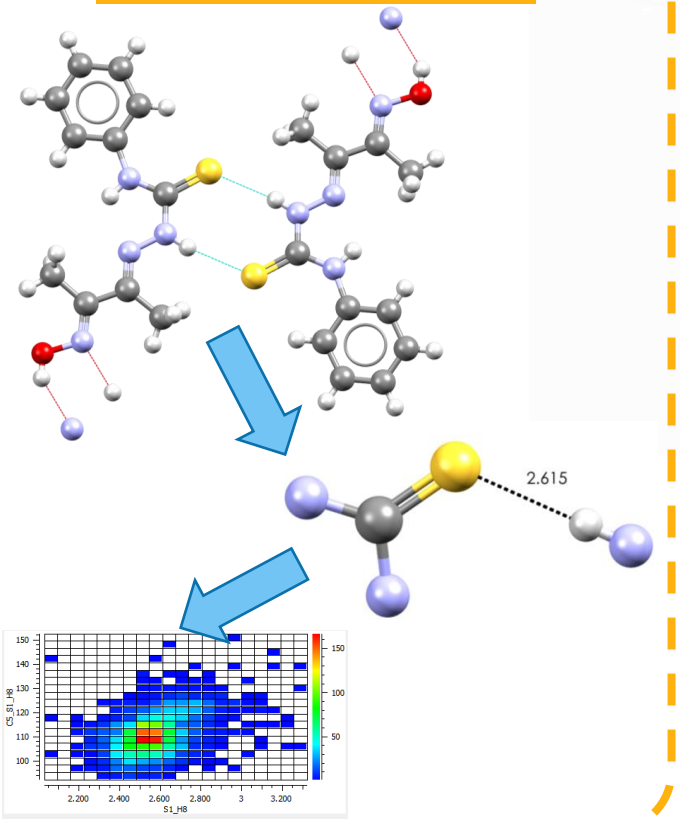
Searching motifs and crystal packing

Motifs

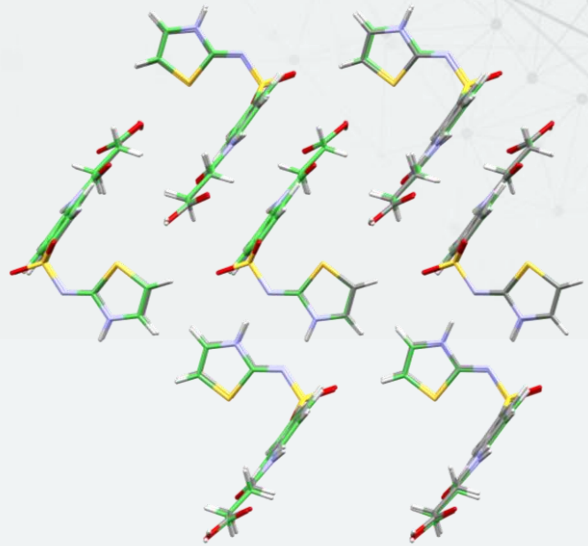


- H-bond motifs

Crystal Packing Feature



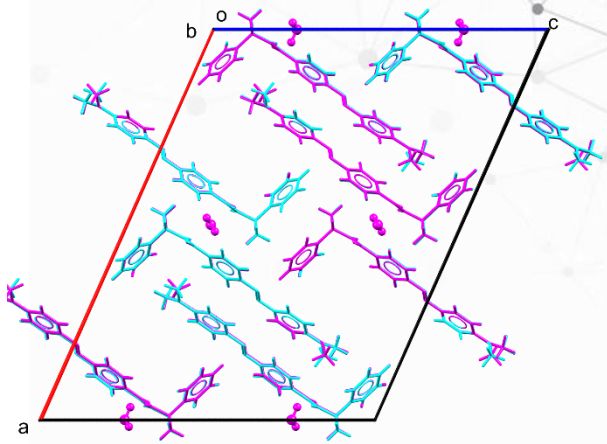
Crystal Packing Similarity



CCDC

Crystal Packing Similarity

- Determine whether two crystal structures are the same
- Identify, from a list of structures, the number of distinct polymorphic forms
- Identify iso-structurality in solvates, salts, hydrates & co-crystals
- Find an experimentally observed structure within a prediction list
- Quantify similarity between polymorphs, solvates, salts, hydrates & co-crystals



Packing Similarity analysis for:

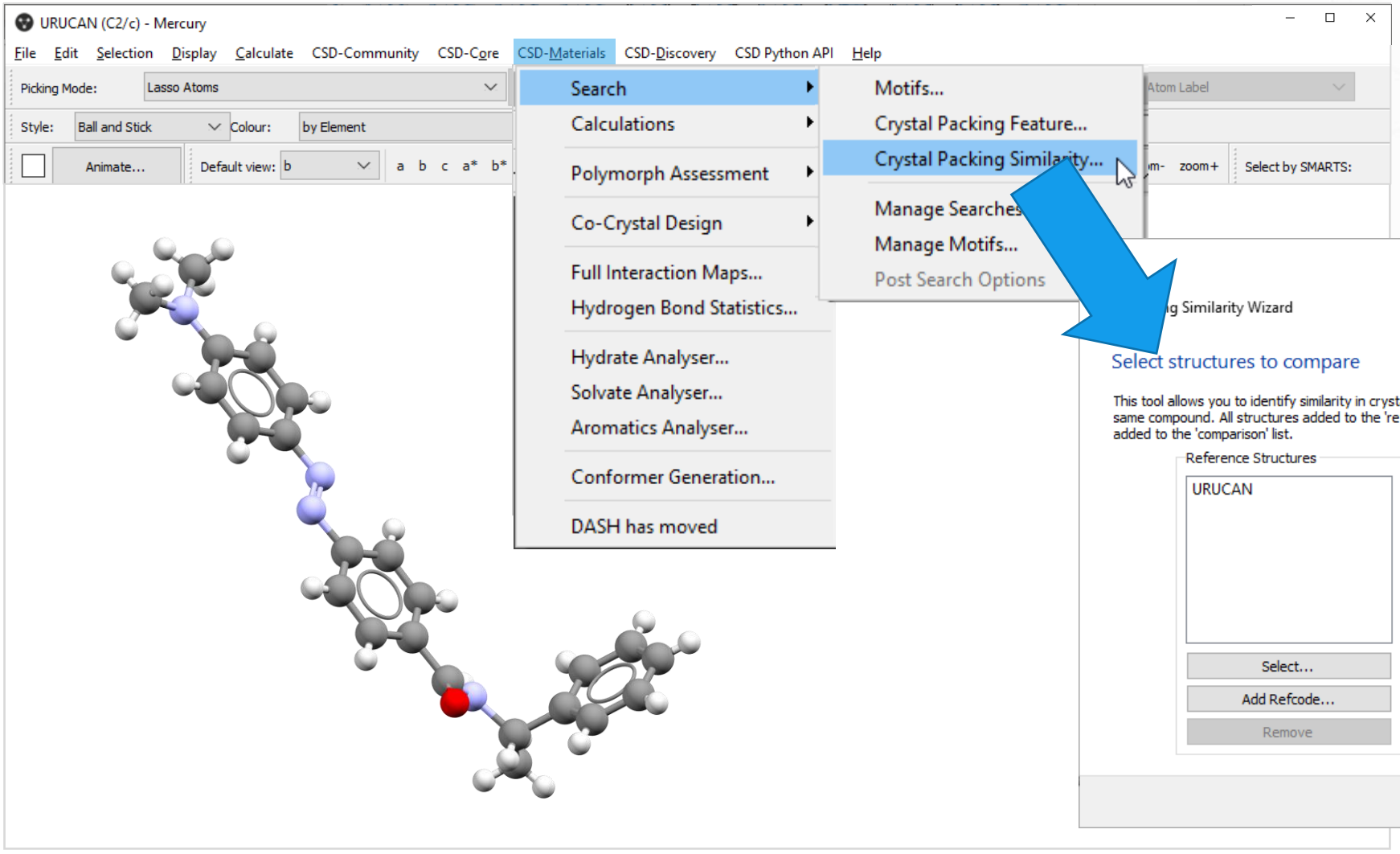
- trans-rac-4-((4-(dimethylamino)phenyl)diazenyl)-N-(1-phenylethyl)benzamide (Refcode: URUCAN)
- and its hydrated structure (Refcode: URUCEL)

Molecules depicted in:

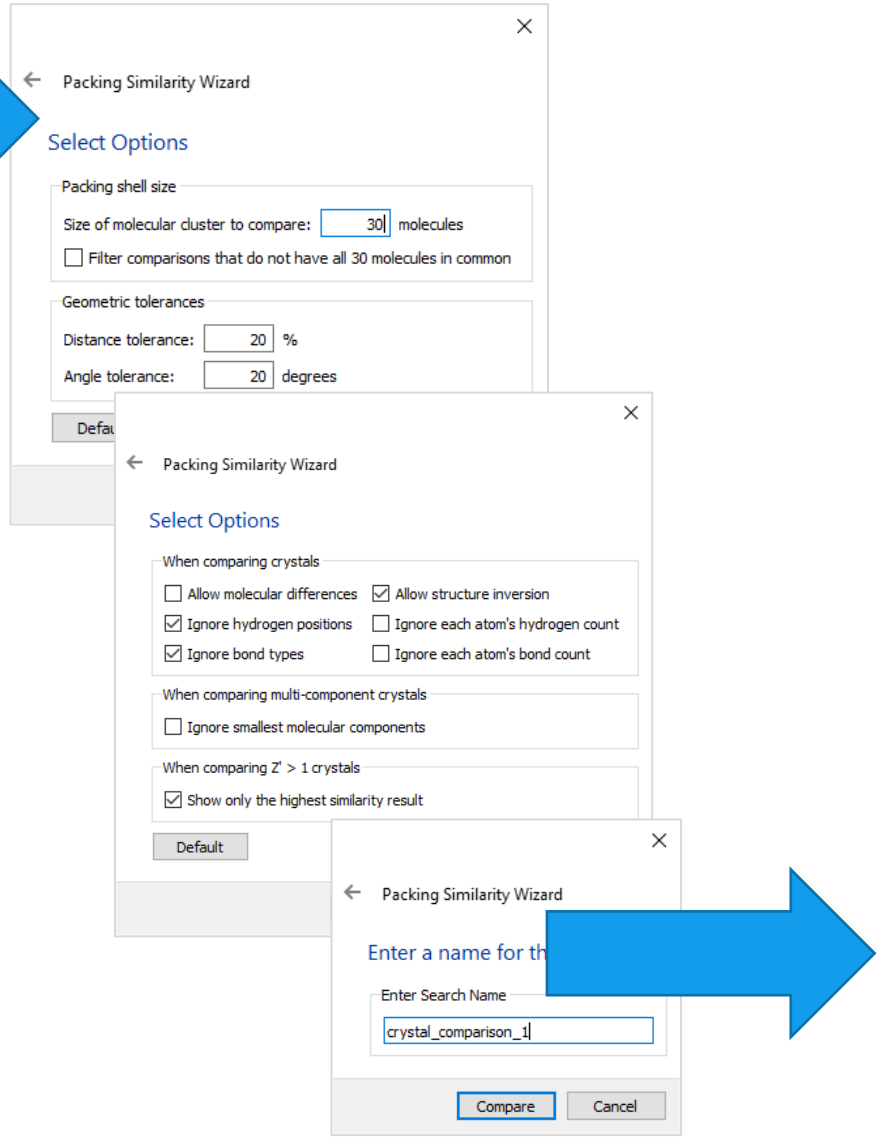
- magenta show the hydrated structure
- blue show the anhydrated structure

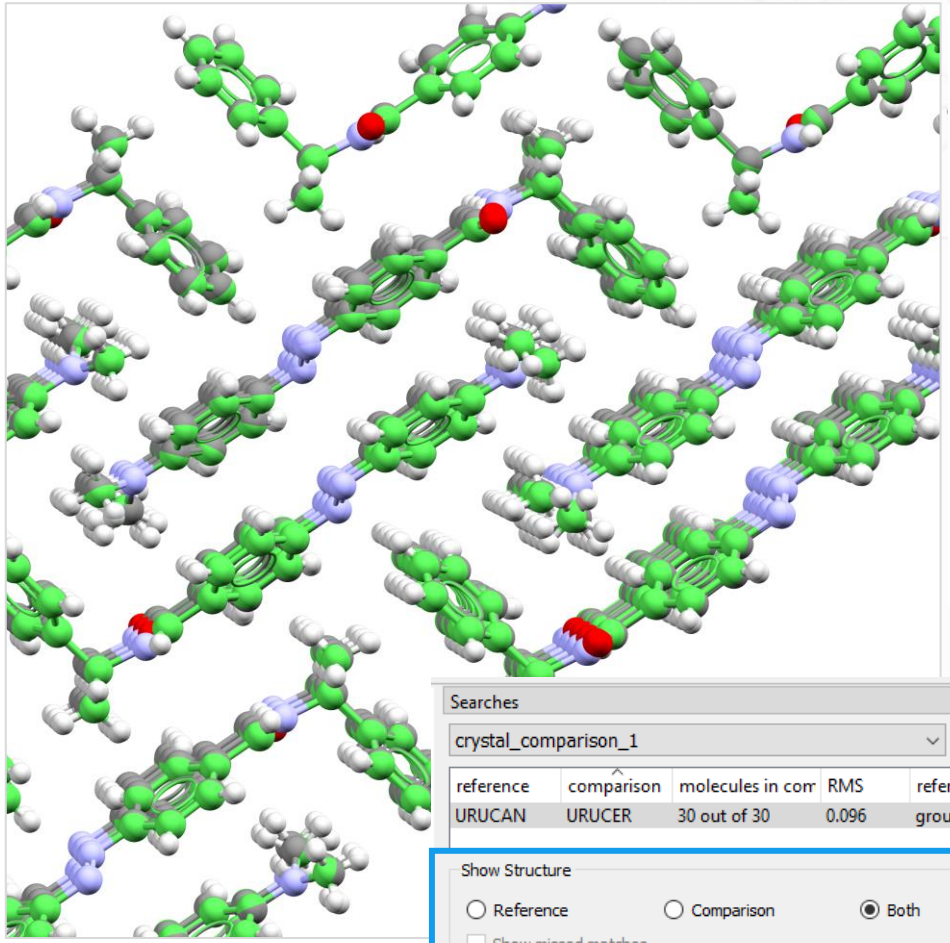
CCDC

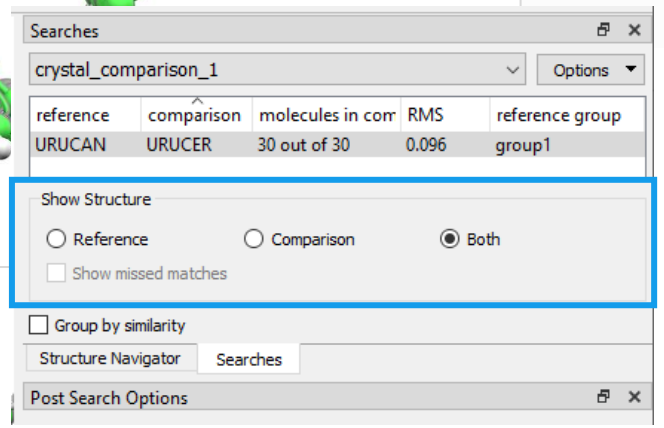
Searching for packing similarity



CCDC







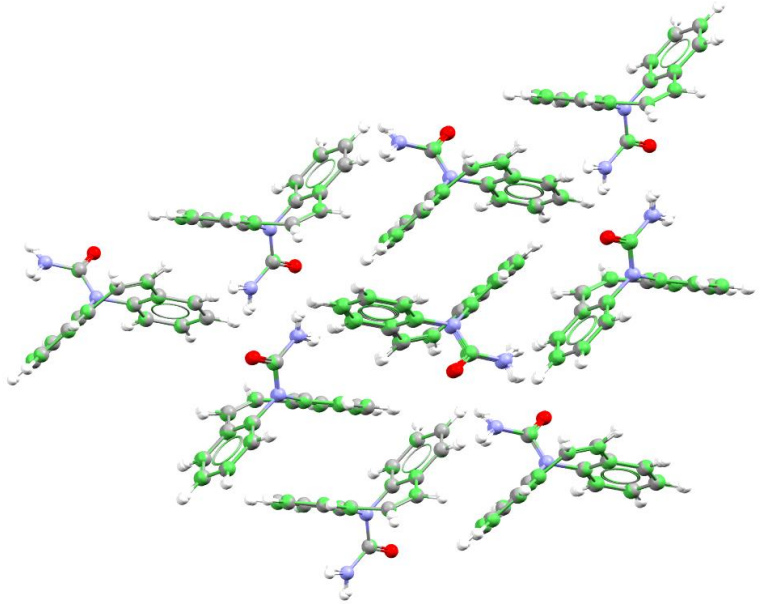
- Reference structure: **grey** carbon atoms.
- Comparison structure: **green** C if in matching molecules, **red** C if missed matches.

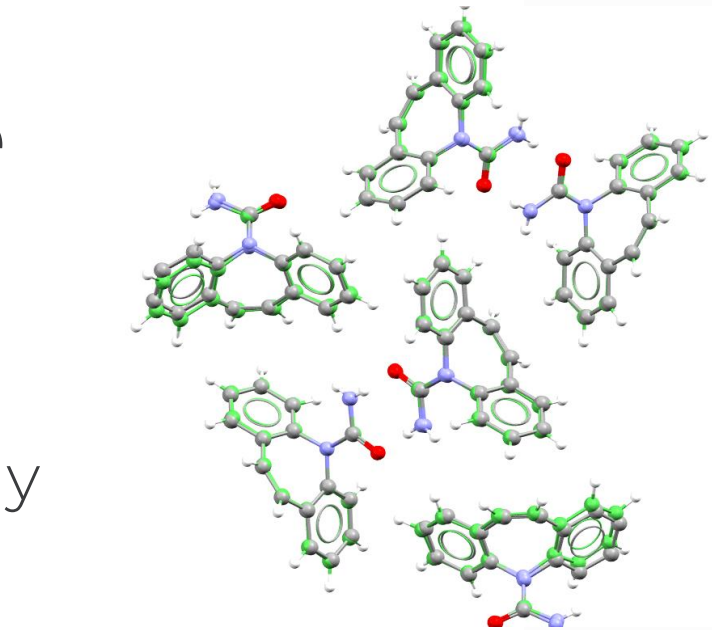
Select what to visualise

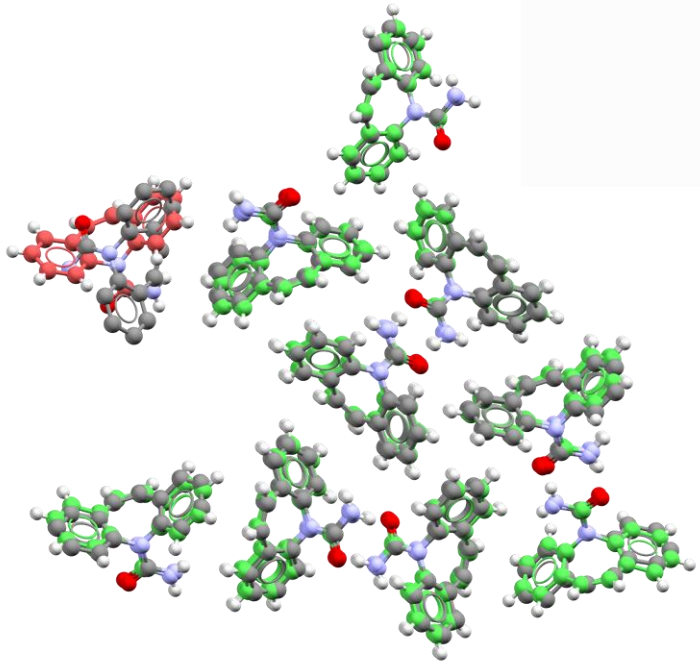
CCDC

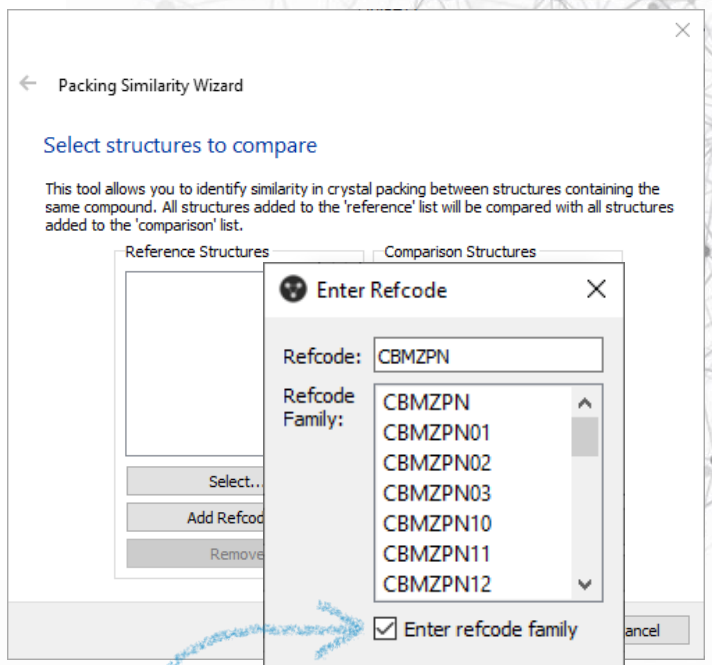
Carbamazepine

- CBMZPN
- Quickly identify similarly packed structures/ polymorph families







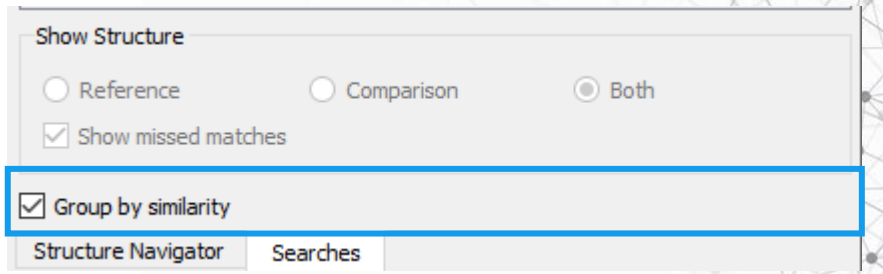


Tick Enter refcode family to add all the structures of carbamazepine

CCDC

Carbamazepine

- Quickly identify similarly packed structures/ polymorph families



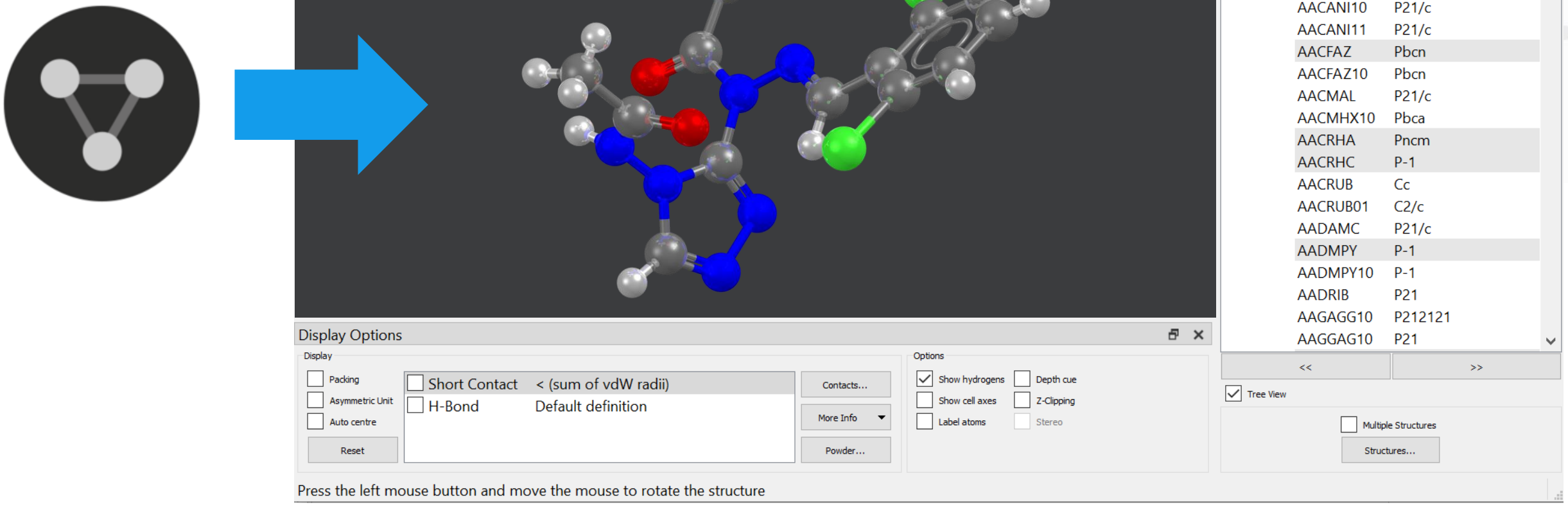
reference	comparison	molecules in com	RMS	reference	PXRD similarity
CBMZPN01	CBMZPN02	15 out of 15	0.011	group2	1
CBMZPN01	CBMZPN03	2 out of 15	0.288	group2	0.803
CBMZPN01	CBMZPN10	15 out of 15	0.008	group2	1
CBMZPN01	CBMZPN11	2 out of 15	0.306	group2	0.889
CBMZPN01	CBMZPN12	2 out of 15	0.127	group2	0.919
CBMZPN01	CBMZPN13	2 out of 15	0.274	group2	0.888
CBMZPN01	CBMZPN14	15 out of 15	0.028	group2	0.999
CBMZPN01	CBMZPN16	1 out of 15	0.074	group2	0.919
CBMZPN01	CBMZPN17	15 out of 15	0.084	group2	0.993
CBMZPN01	CBMZPN18	15 out of 15	0.084	group2	0.993
CBMZPN01	CBMZPN19	15 out of 15	0.084	group2	0.993
CBMZPN01	CBMZPN20	15 out of 15	0.015	group2	1
CBMZPN01	CBMZPN21	15 out of 15	0.078	group2	0.994
CBMZPN01	CBMZPN22	15 out of 15	0.09	group2	0.992
CBMZPN01	CBMZPN23	15 out of 15	0.09	group2	0.992
CBMZPN01	CBMZPN27	15 out of 15	0.059	group2	0.995
CBMZPN01	CBMZPN28	15 out of 15	0.073	group2	0.998
CBMZPN01	CBMZPN29	15 out of 15	0.057	group2	0.998
CBMZPN01	CBMZPN30	15 out of 15	0.229	group2	0.982
CBMZPN01	CBMZPN31	15 out of 15	0.093	group2	0.993
CBMZPN01	CBMZPN32	15 out of 15	0.161	group2	0.997
CBMZPN01	CBMZPN33	15 out of 15	0.07	group2	0.994
CBMZPN02	CBMZPN03	2 out of 15	0.291	group2	0.804
CBMZPN02	CBMZPN10	15 out of 15	0.012	group2	1
CBMZPN02	CBMZPN11	2 out of 15	0.309	group2	0.89
CBMZPN02	CBMZPN12	2 out of 15	0.13	group2	0.919

group		spacegroup	density	a	b	c	alpha	beta	gamma
group1									
	CBMZPN13	P-1	1.335	5.1856	20.57...	22.24...	84.19...	87.97...	85.1053
	CBMZPN03	R-3	1.235	35.454	35.454	5.253	90	90	120
	CBMZPN11	P-1	1.339	5.1705	20.574	22.245	84.124	88.008	85.187
group2									
	CBMZPN17	P21/n	1.38	7.4874	11.04...	13.77...	90	92.939	90
	CBMZPN18	P21/n	0	7.4874	11.04...	13.77...	90	92.939	90
	CBMZPN19	P21/n	1.38	7.487	11.041	13.775	90	92.9	90
	CBMZPN20	P21	1.342	7.542	11.155	13.919	90	92.885	90
	CBMZPN21	P21/n	1.374	7.498	11.058	13.789	90	92.838	90
	CBMZPN22	P21/n	1.382	7.4893	11.03...	13.764	90	92.953	90
	CBMZPN23	P21/n	1.382	7.4893	11.03...	13.764	90	92.953	90
	CBMZPN27	P21/n	1.373	7.494...	11.06...	13.80...	90	92.91...	90
	CBMZPN28	P21/n	1.326	7.578	11.176	13.991	90	93.08	90
	CBMZPN29	P21/n	1.327	7.576	11.188	13.967	90	87.03	90
	CBMZPN30	P21/n	1.283	7.68	11.44	13.92	90	91.22	90
	CBMZPN31	P21/n	1.386	7.46	11.04	13.76	90	92.61	90
	CBMZPN32	P21/n	1.344	7.534	11.15	13.917	90	92.94	90
	CBMZPN33	P21/n	1.376	7.4907	11.058	13.78...	90	92.903	90
	CBMZPN01	P21/c	1.347	7.529	11.148	15.47	90	116.17	90
	CBMZPN02	P21/n	1.344	7.534	11.15	13.917	90	92.94	90
	CBMZPN10	P21/n	1.343	7.537	11.156	13.912	90	92.86	90
	CBMZPN14	P21/n	1.333	7.55	11.186	13.954	90	92.938	90
group3									
	CBMZPN12	C2/c	1.296	26.609	6.9269	13.957	90	109.7...	90
group4									
	CBMZPN16	Pbca	1.326	9.1245	10.45...	24.82...	90	90	90

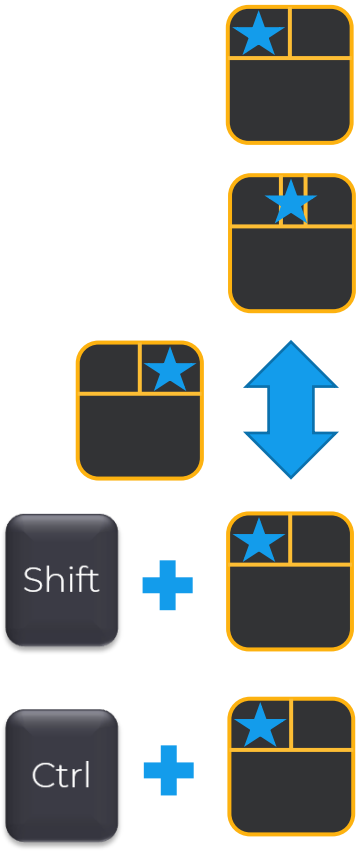
- Show missed matches
- Ignore smallest component
- Allow molecular differences

CCDC

Review of basics of Mercury



The 3D window basics

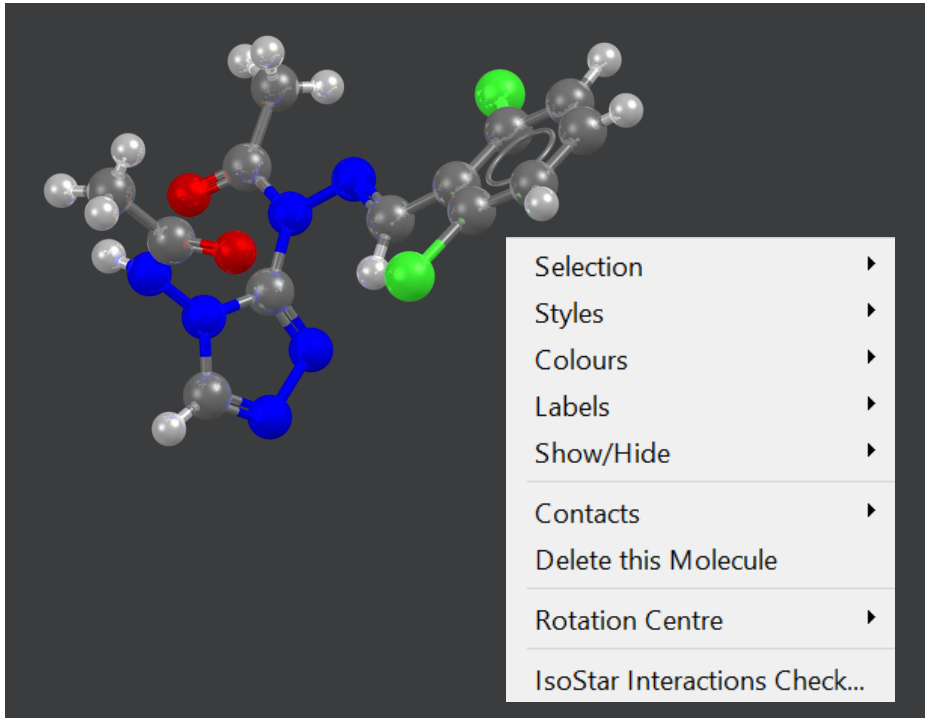


- Left mouse button and move – rotate molecules
- Middle Mouse wheel – move molecules up and down
- Right mouse button and move up and down – zoom in and out of molecules
- Shift + Left mouse button and move - rotate in the plane molecules
- Ctrl + Left mouse button and move - translate molecules

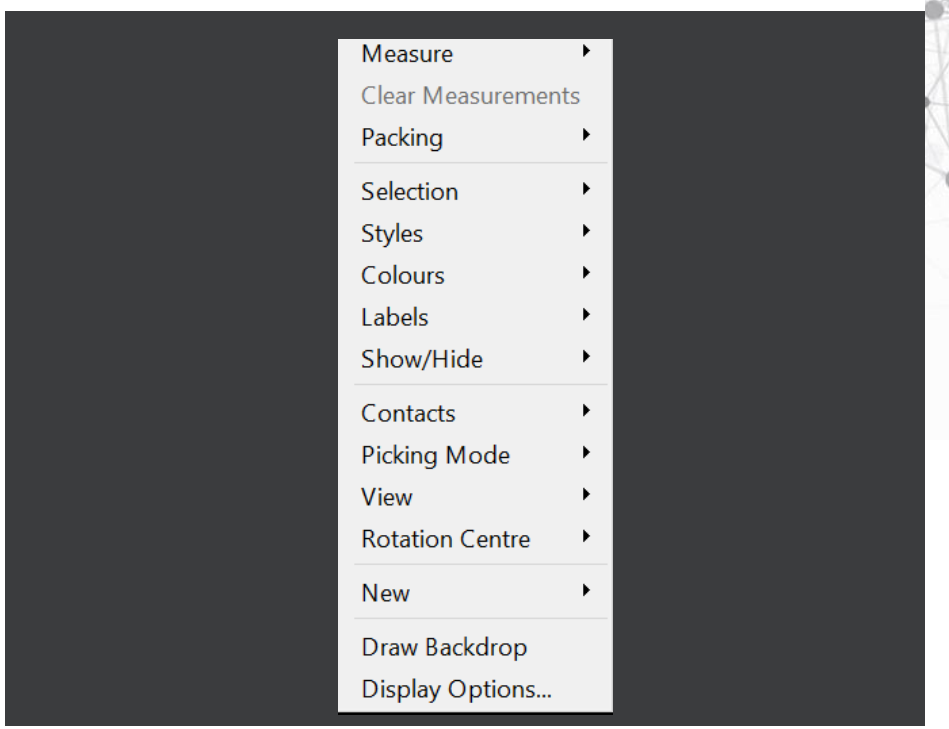
CCDC

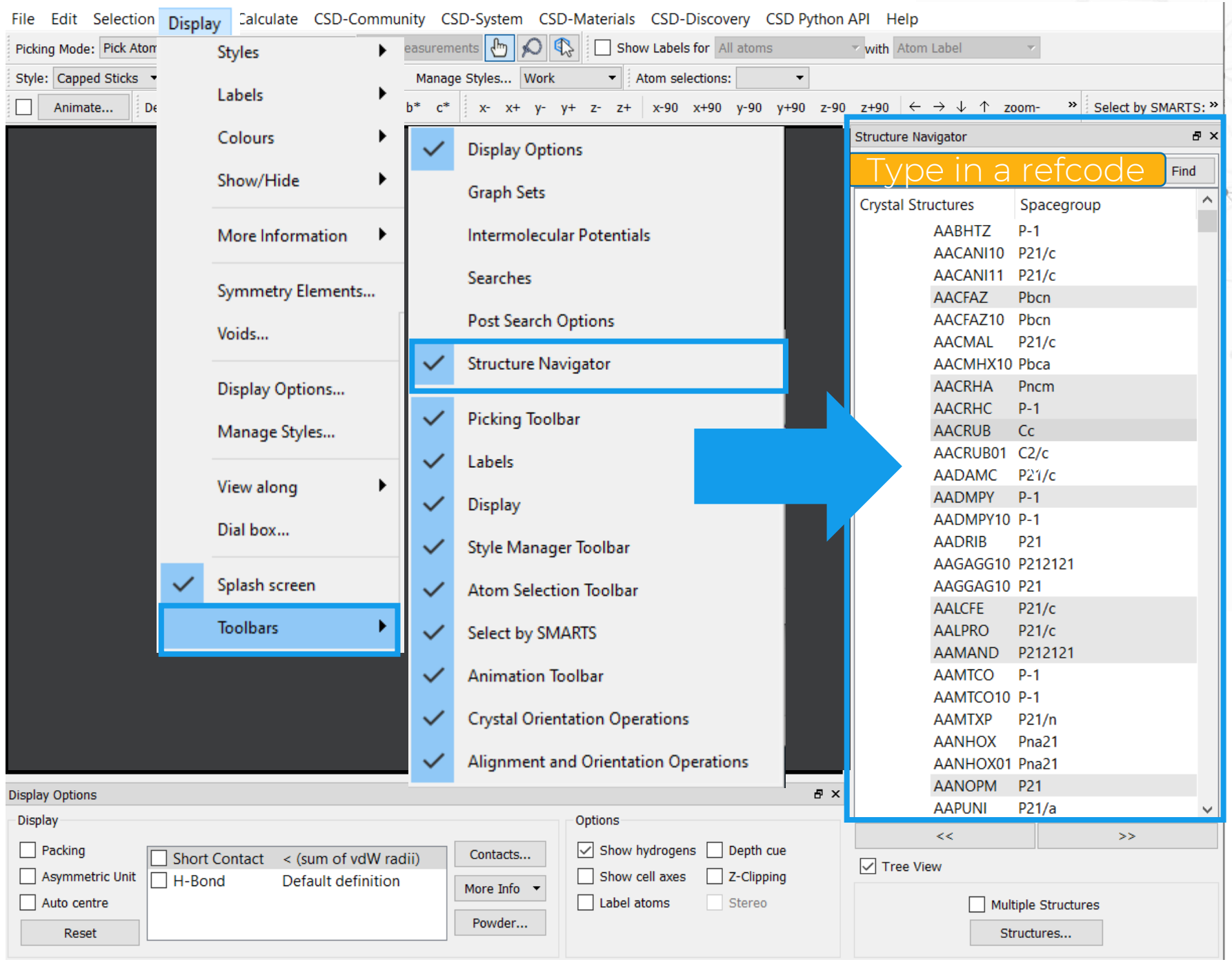
Right mouse click

Near a molecule



Away from a molecule



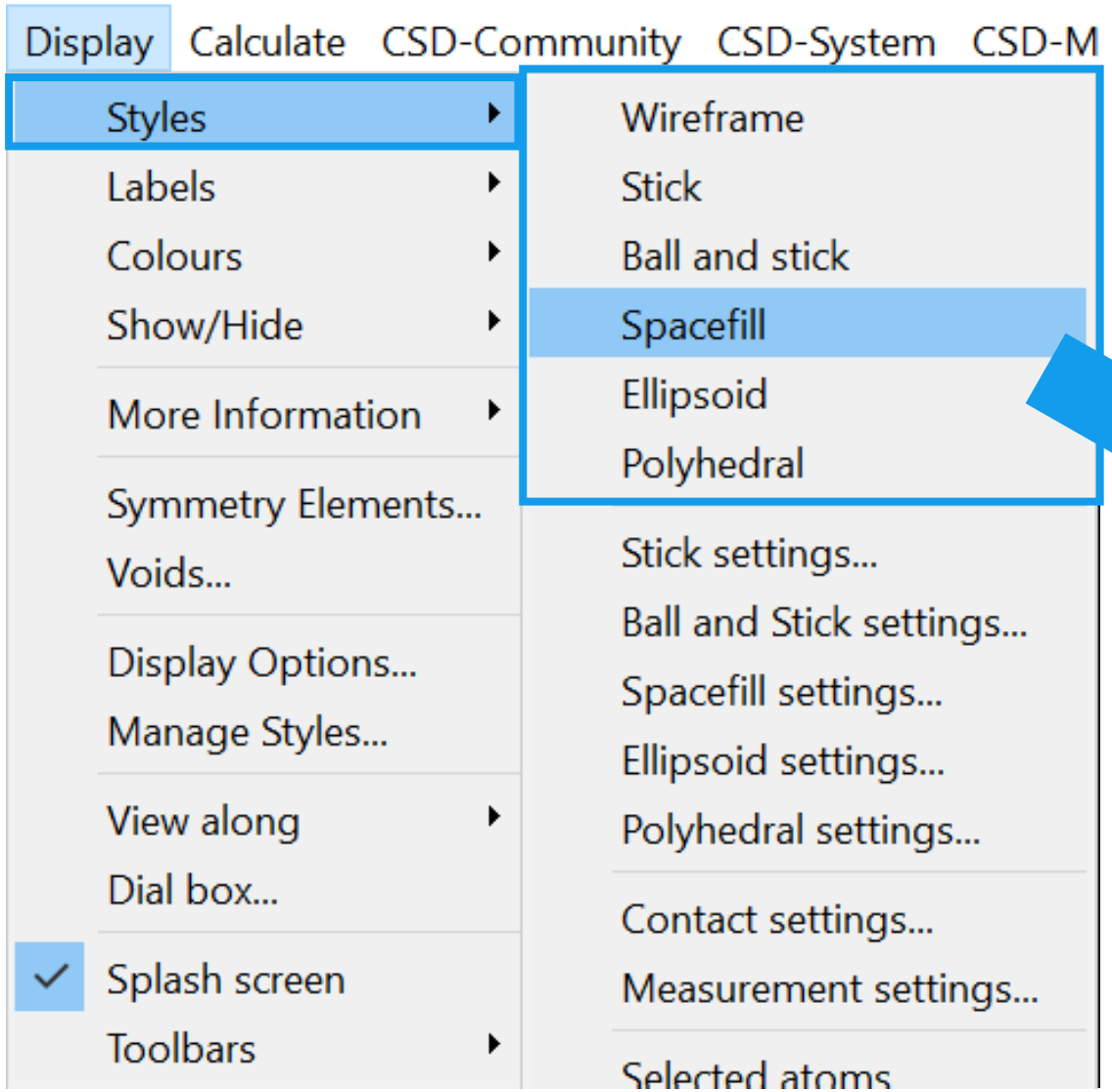


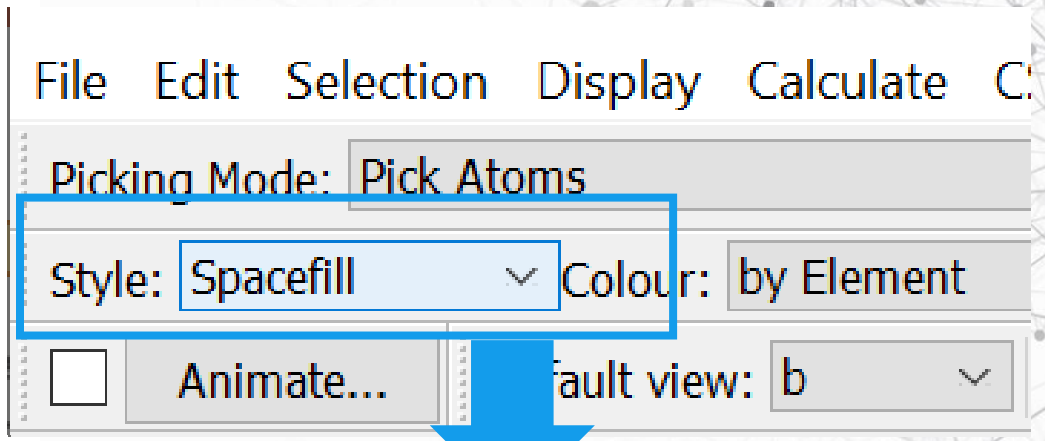
Handy tips:

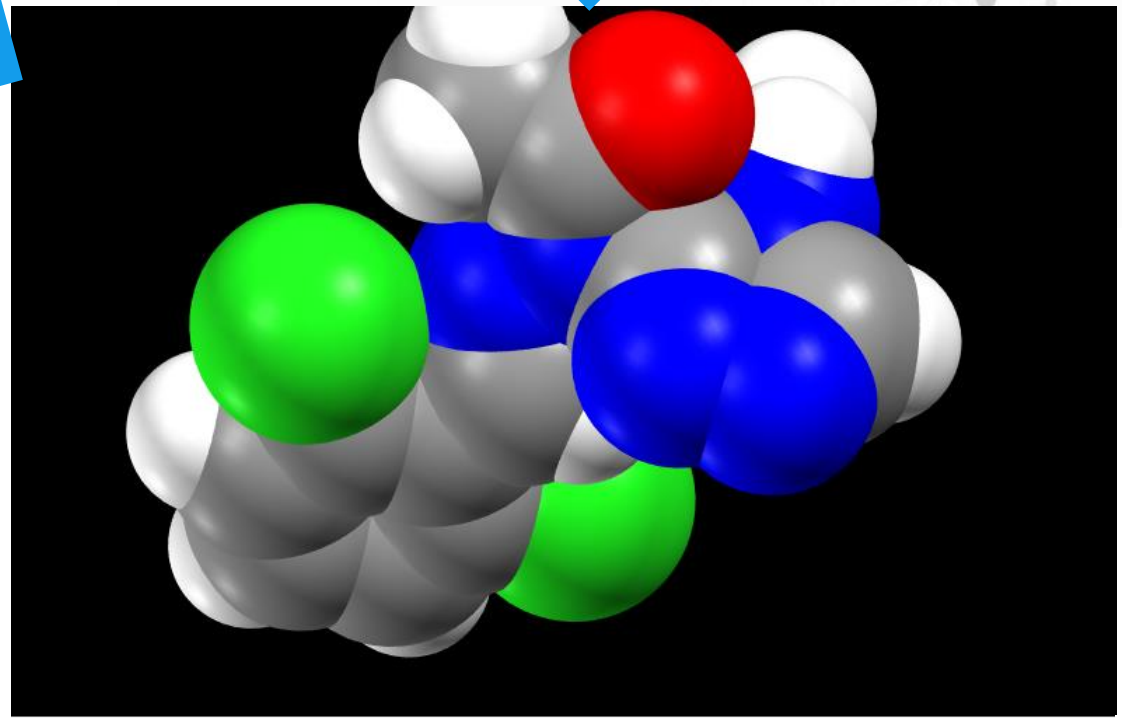
- Recovering Toolbars
- Resetting view

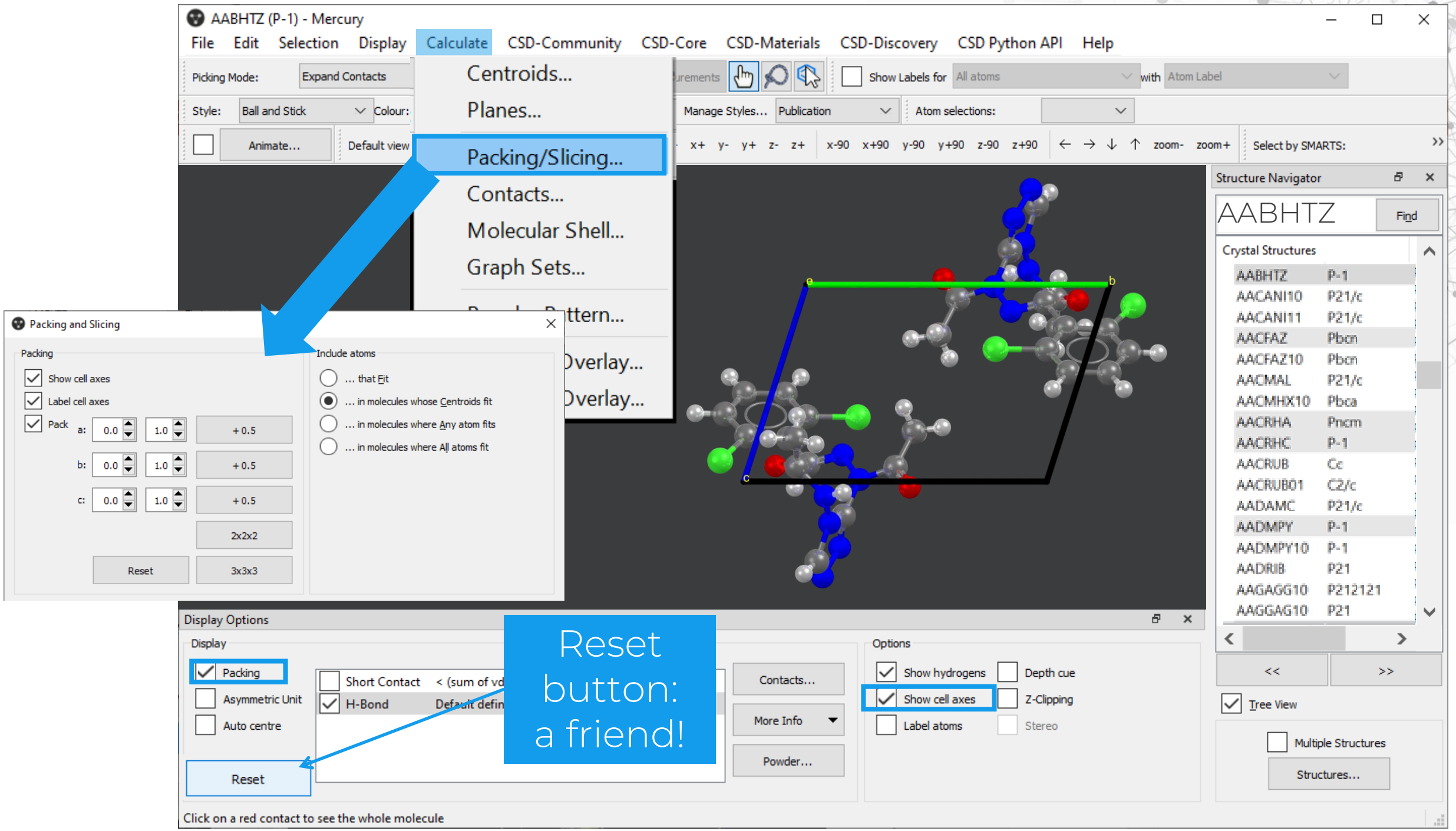
CCDC

Changing display - Style

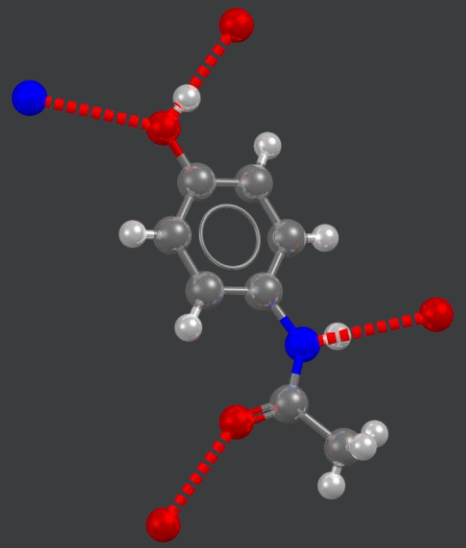




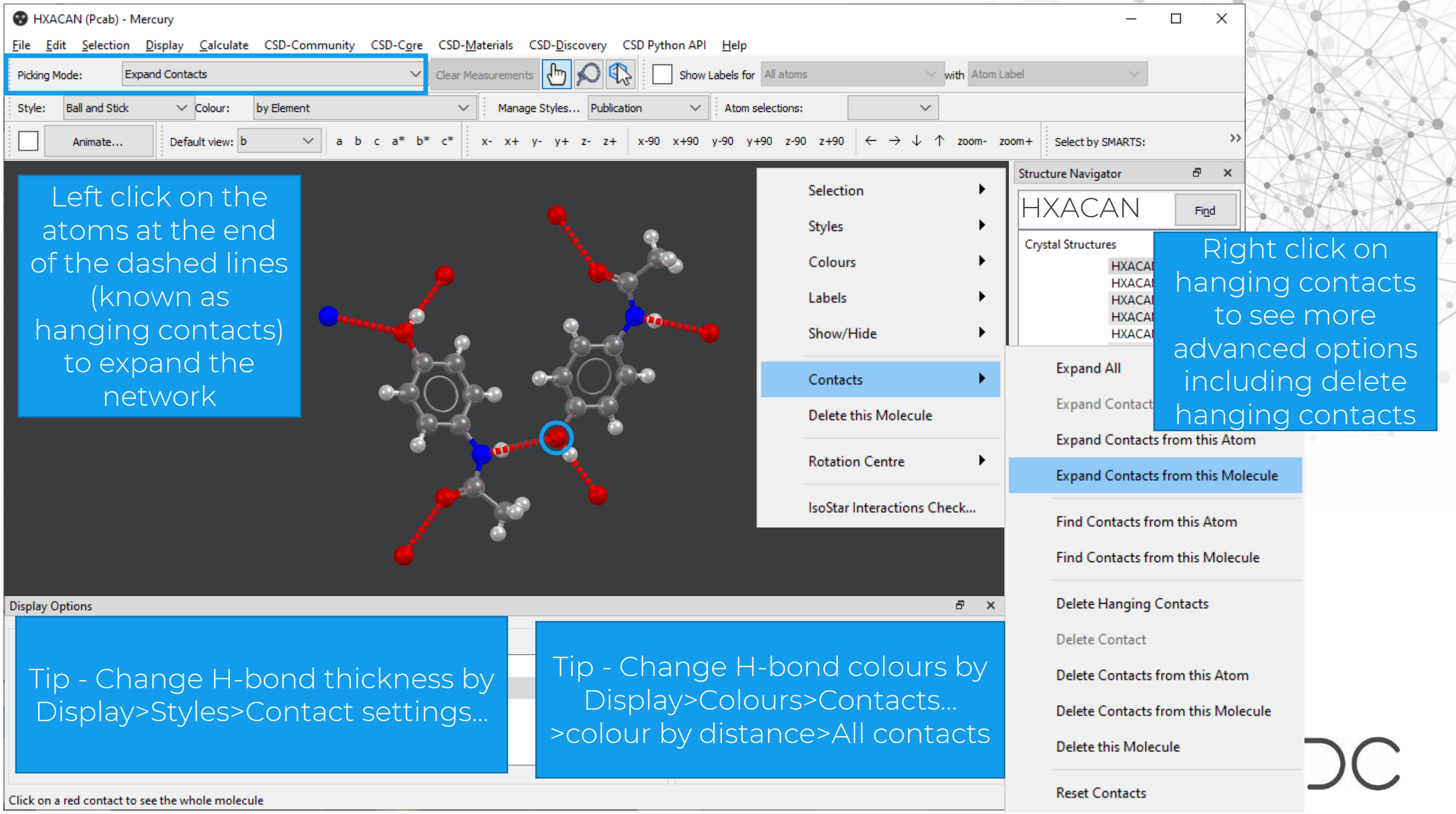



AABHTZ (P-1) - Mercury
File Edit Selection Display Calculate CSD-Community CSD-Core CSD-Materials CSD-Discovery CSD Python API Help
Centroids...
Planes...
Packing/Slicing...
Contacts...
Molecular Shell...
Graph Sets...
Overlay...
Overlay...
Packing and Slicing
Packing
Show cell axes
Label cell axes
Pack
Include atoms
... that Fit
... in molecules whose Centroids fit
... in molecules where Any atom fits
... in molecules where All atoms fit
+0.5
2x2x2
3x3x3
Reset
Structure Navigator
AABHTZ
Find
Crystal Structures
Display Options
Display
Packing
Asymmetric Unit
Auto centre
Short Contact
H-Bond
Contacts...
More Info
Powder...
Options
Show hydrogens
Depth cue
Show cell axes
Z-Clipping
Label atoms
Stereo
Tree View
Multiple Structures
Structures...
Reset button: a friend!
Click on a red contact to see the whole molecule

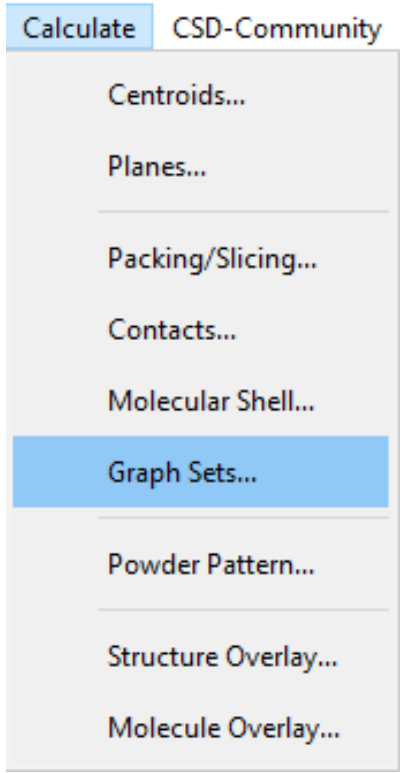
Visualising hydrogen bonds

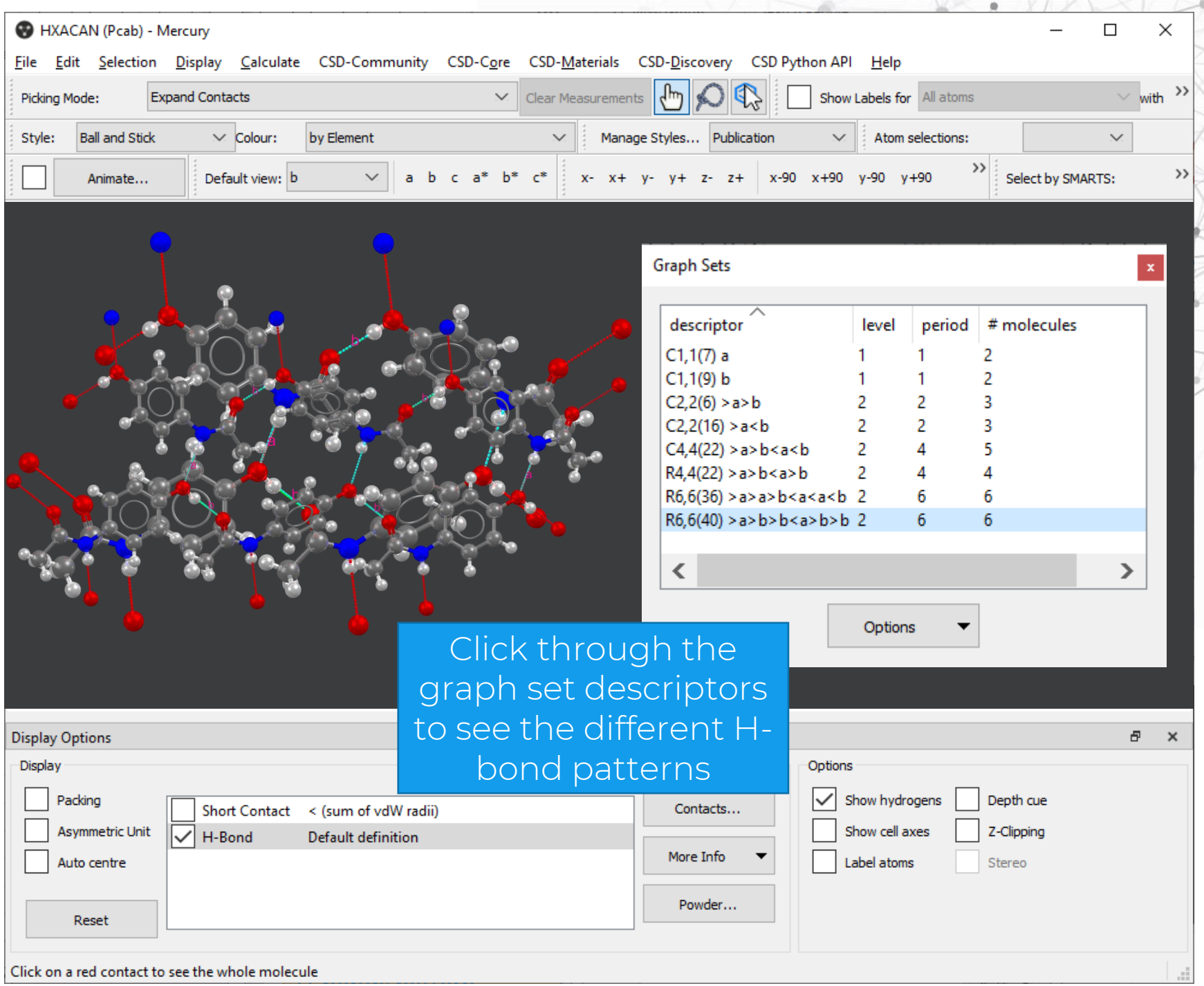


Click on “Default definition” to change the default H-bond definition


HXACAN (Pcab) - Mercury
File Edit Selection Display Calculate CSD-Community CSD-Core CSD-Materials CSD-Discovery CSD Python API Help
Picking Mode: Expand Contacts
Clear Measurements
Show Labels for All atoms with Atom Label
Style: Ball and Stick
Colour: by Element
Manage Styles... Publication
Atom selections:
Animate...
Default view: b
a b c a* b* c*
x- x+ y- y+ z- z+ x-90 x+90 y-90 y+90 z-90 z+90
zoom- zoom+
Select by SMARTS:
Left click on the atoms at the end of the dashed lines (known as hanging contacts) to expand the network
Selection
Styles
Colours
Labels
Show/Hide
Contacts
Delete this Molecule
Rotation Centre
IsoStar Interactions Check...
Structure Navigator
HXACAN
Find
Crystal Structures
Right click on hanging contacts to see more advanced options including delete hanging contacts
Expand All
Expand Contact
Expand Contacts from this Atom
Expand Contacts from this Molecule
Find Contacts from this Atom
Find Contacts from this Molecule
Delete Hanging Contacts
Delete Contact
Delete Contacts from this Atom
Delete Contacts from this Molecule
Delete this Molecule
Reset Contacts
Display Options
Tip - Change H-bond thickness by Display>Styles>Contact settings...
Tip - Change H-bond colours by Display>Colours>Contacts... >colour by distance>All contacts
Click on a red contact to see the whole molecule

Graph Sets





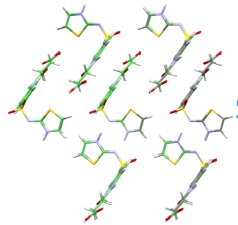
Click through the graph set descriptors to see the different H-bond patterns

Graph Sets describe the **H-bonds pattern**.

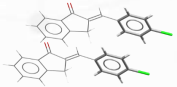
Learn more in the *Glossary* on the handout.

Explore More: Tips and tricks

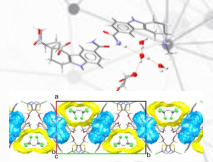
- What else can you do with CSD-Materials to investigate your solid form?
- What else can we learn from the 1.1 million structures in the CSD to guide solid form design and risk assessment?



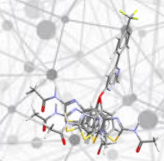
Crystal Packing Similarity



Motif Search & Packing Feature Search



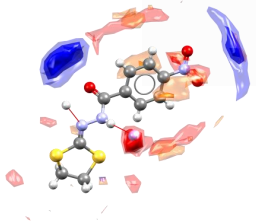
Hydrate Analyser & Solvate Analyser



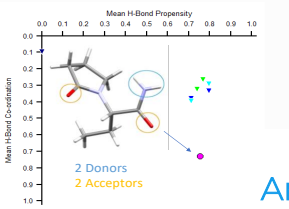
CSD Conformer Generator

Detailed Structural Analysis

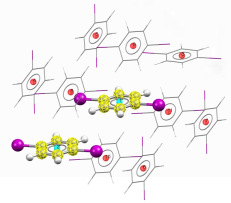
Solid Form Design



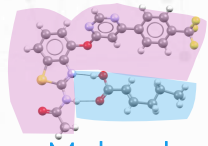
Full Interaction Maps



Hydrogen Bond Propensity



Aromatics Analyser



Molecular Complementarity

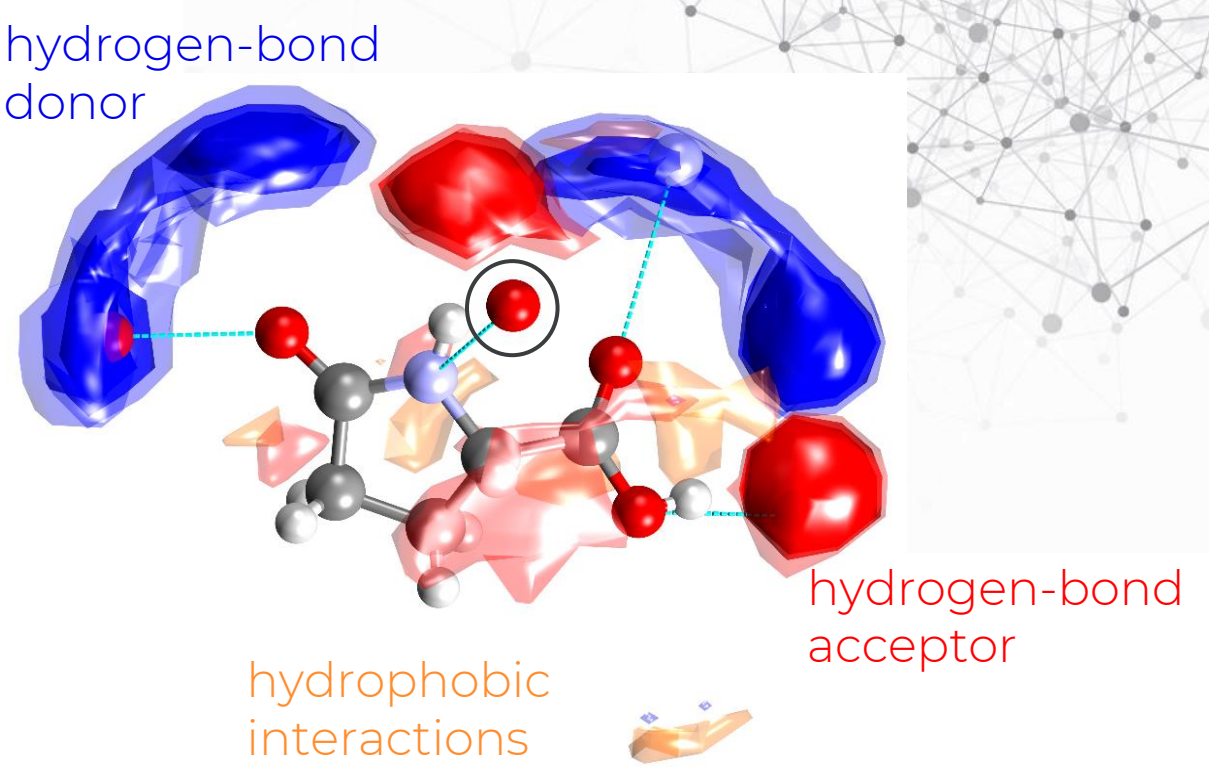
Solid Form Risk Assessment

The session has resumed recording

CCDC

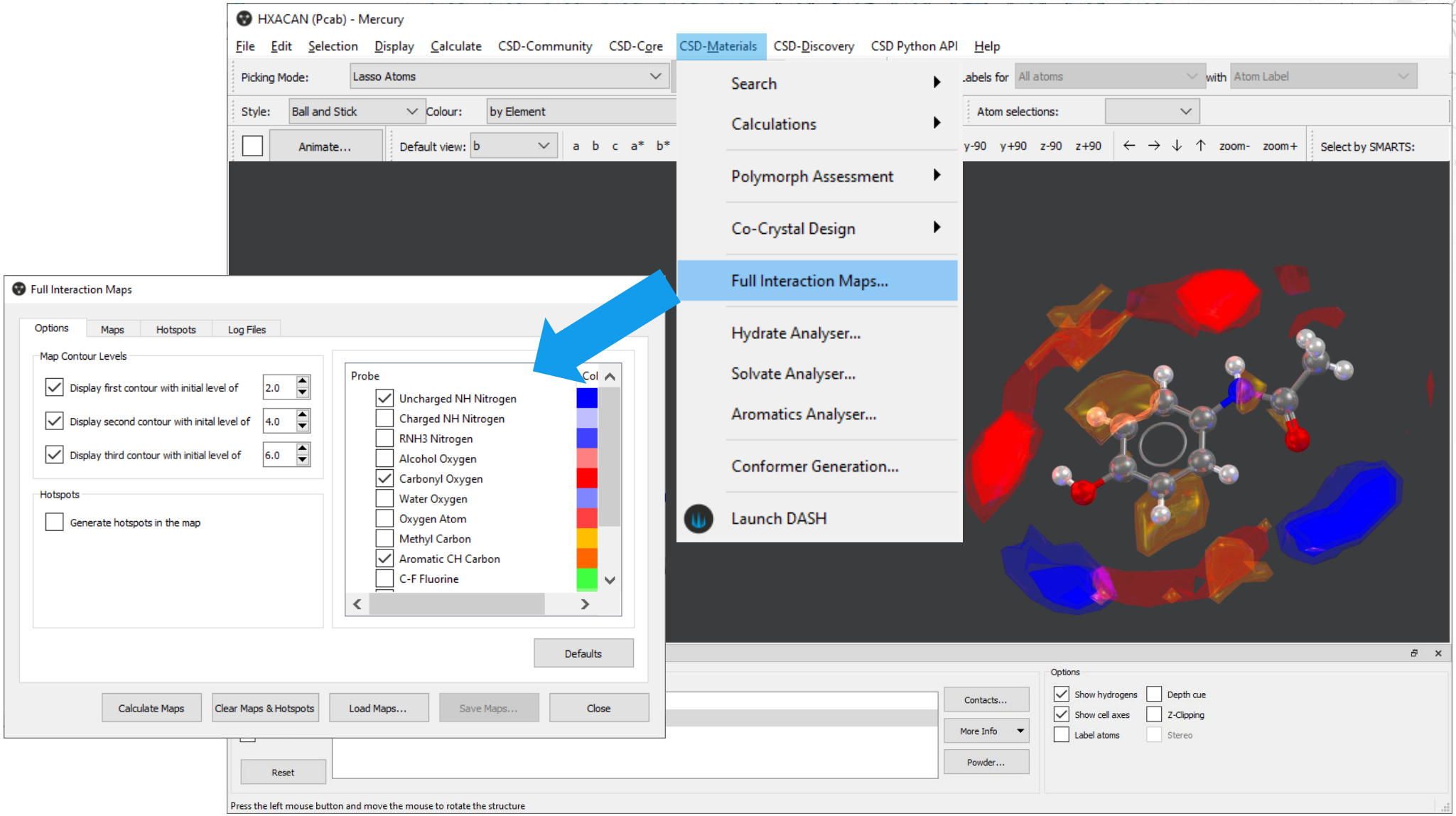
Full Interaction Maps (FIMs)

- FIMs enable you to generate a 3D interaction map around a molecule representing regions of higher probability to find interactions with certain functional groups.
- Visualise observed atom-atom contacts with respect to likely geometries in 3D space.
- Identify interaction hot-spots around chemical groups.



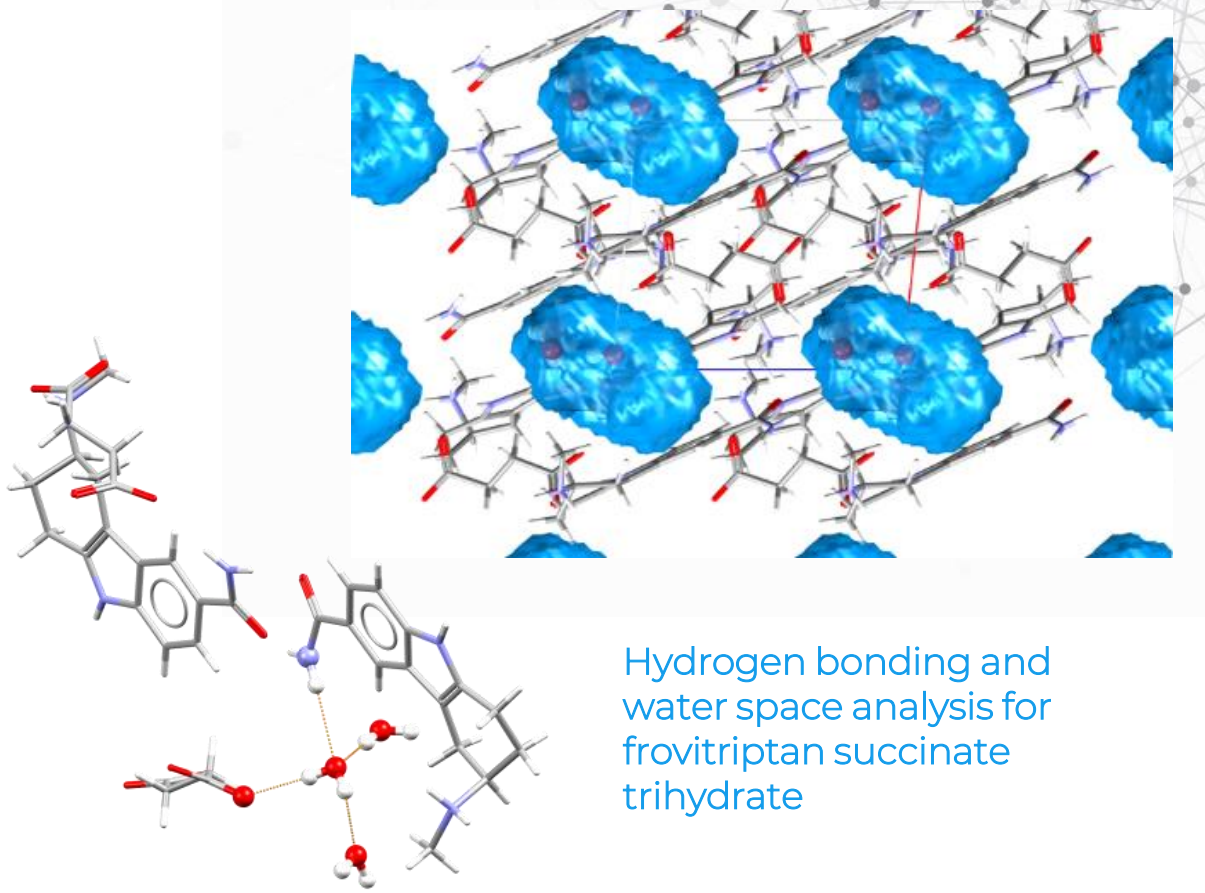
CCDC

Creating a FIM



Hydrate Analyser

- Quickly obtain a summary describing features of a hydrated phase
- Assess hydrogen bonding motifs involving water molecules
- Interpret the space occupied in the lattice by water molecules
- Analyse extended structural features such as coordination polymers including water
- Obtain reports on the assessments made

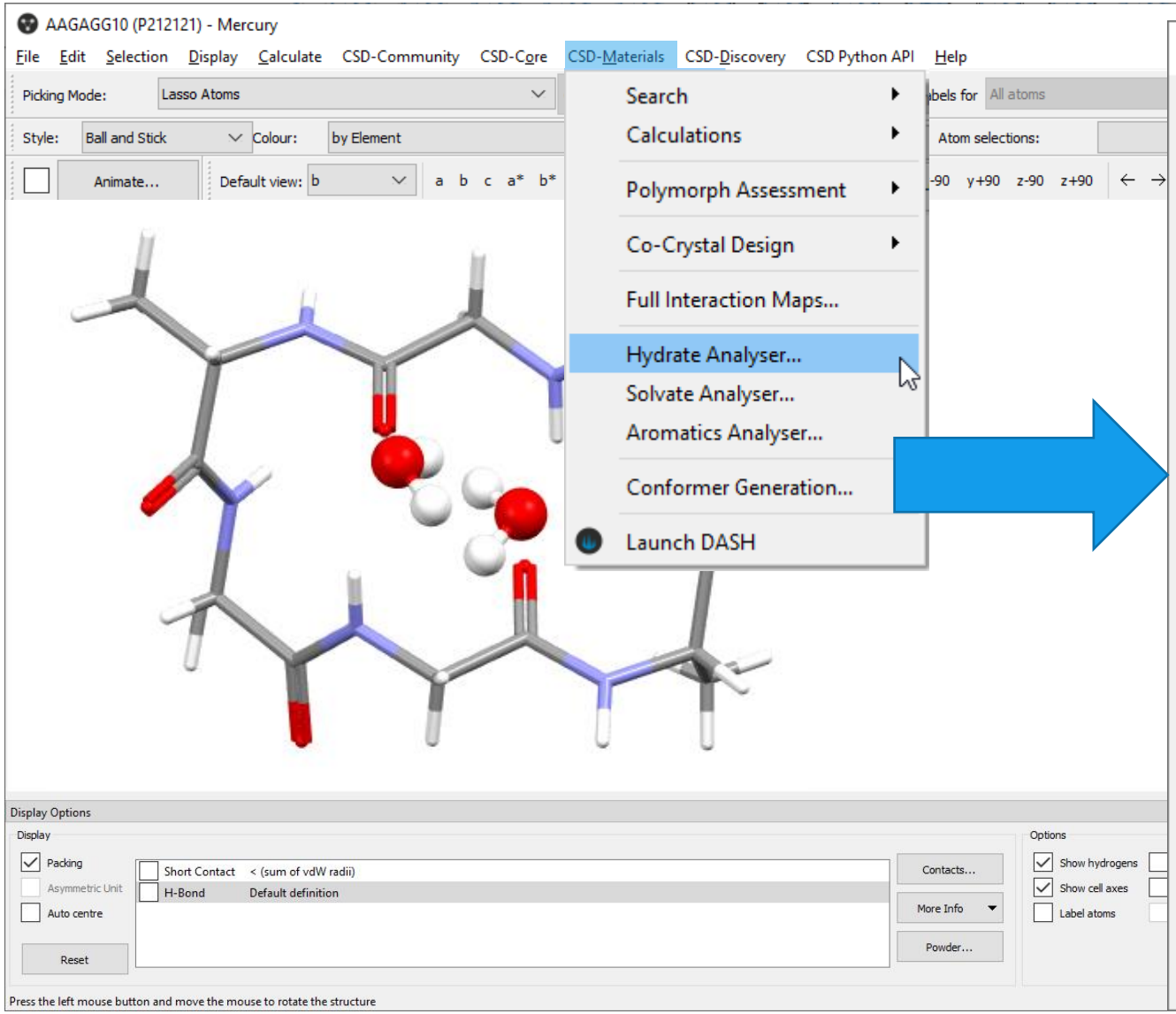


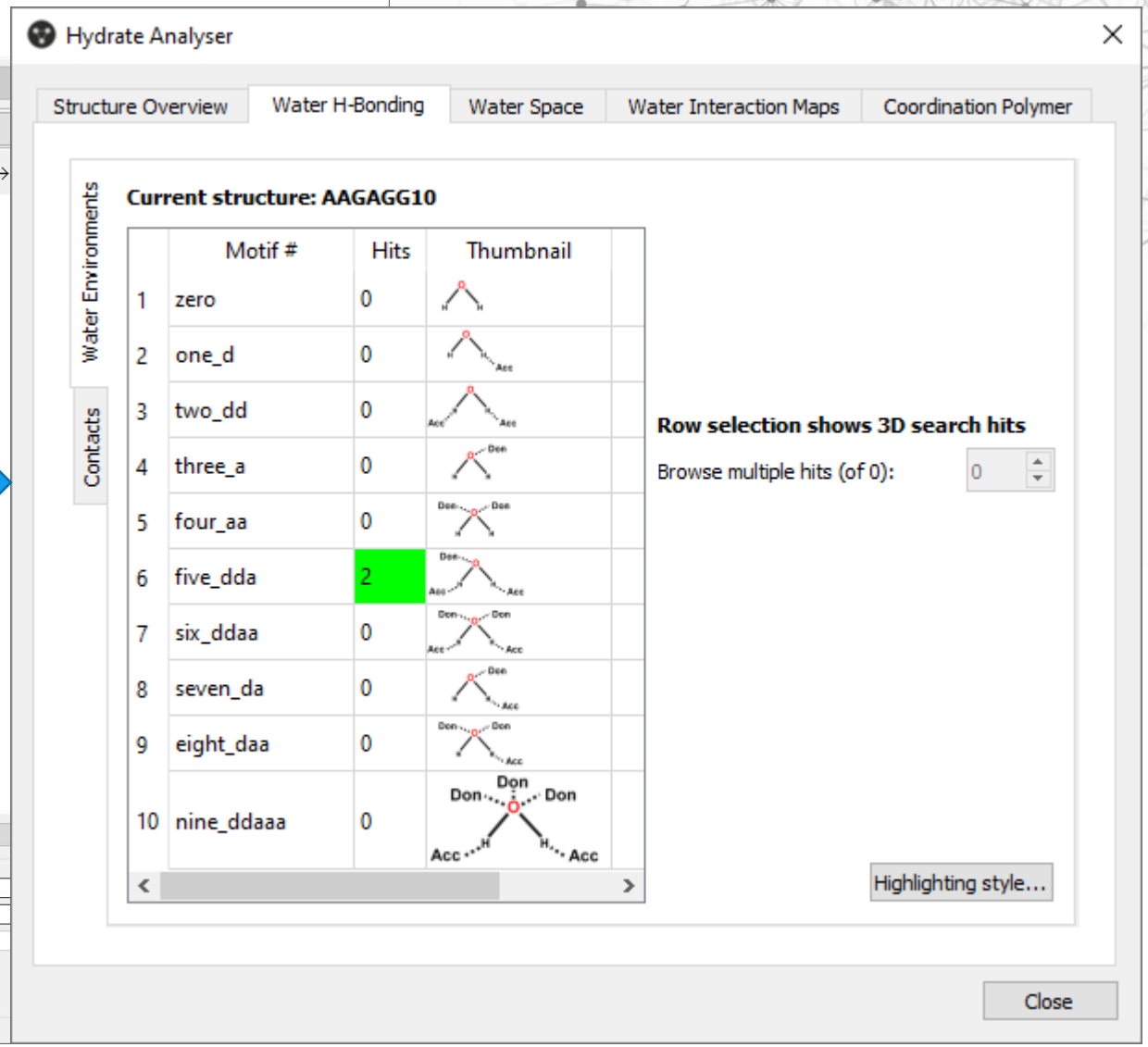
Hydrogen bonding and water space analysis for frovitriptan succinate trihydrate

CCDC

Using the Hydrate Analyser

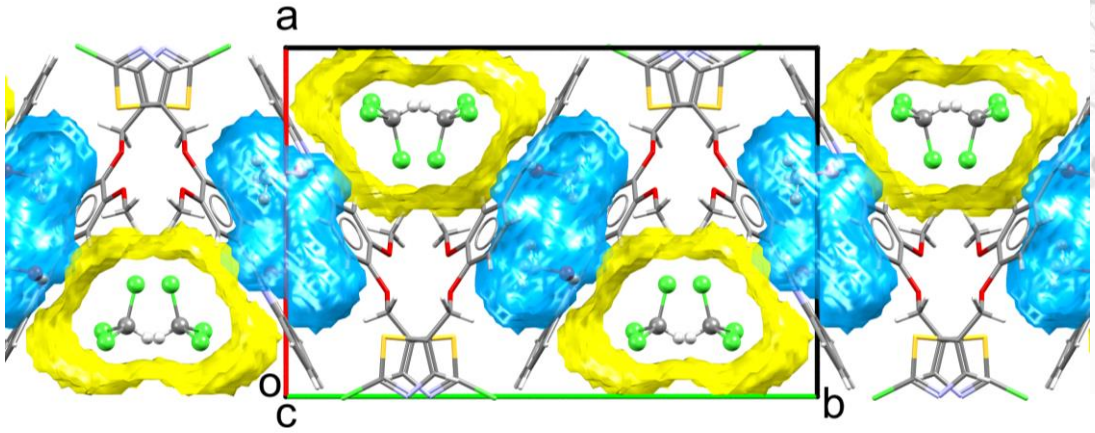
CSD Refcode: AAGAGG10





Solvate Analyser

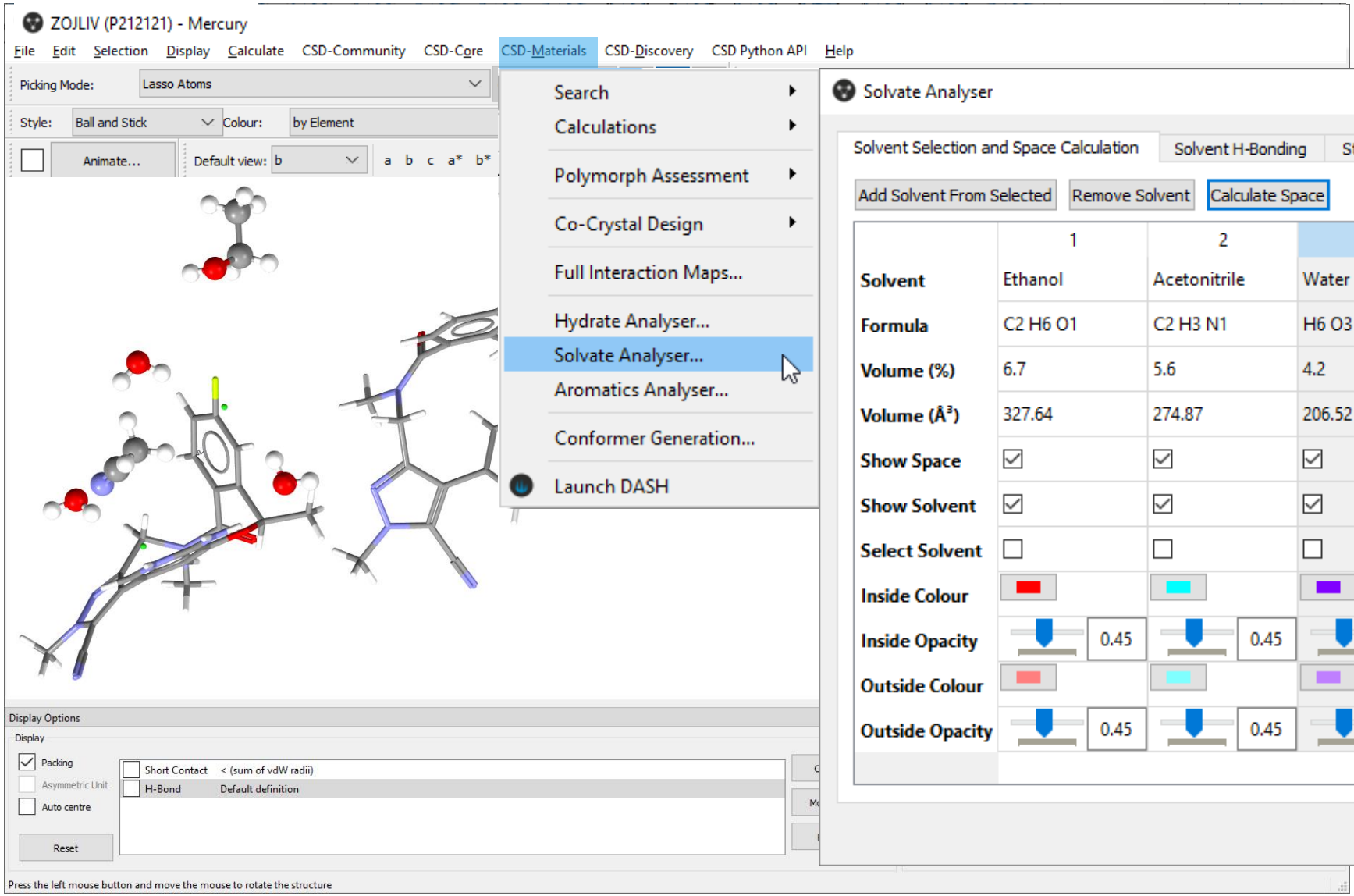
- Quick analysis of the structures containing one or more solvent molecules
- Easy selection of solvent molecule(s)
- Assessment of any hydrogen bonding motifs to/from the solvents
- Calculation and display of the space occupied by each of the different solvent molecules
- Obtain reports on the assessments made

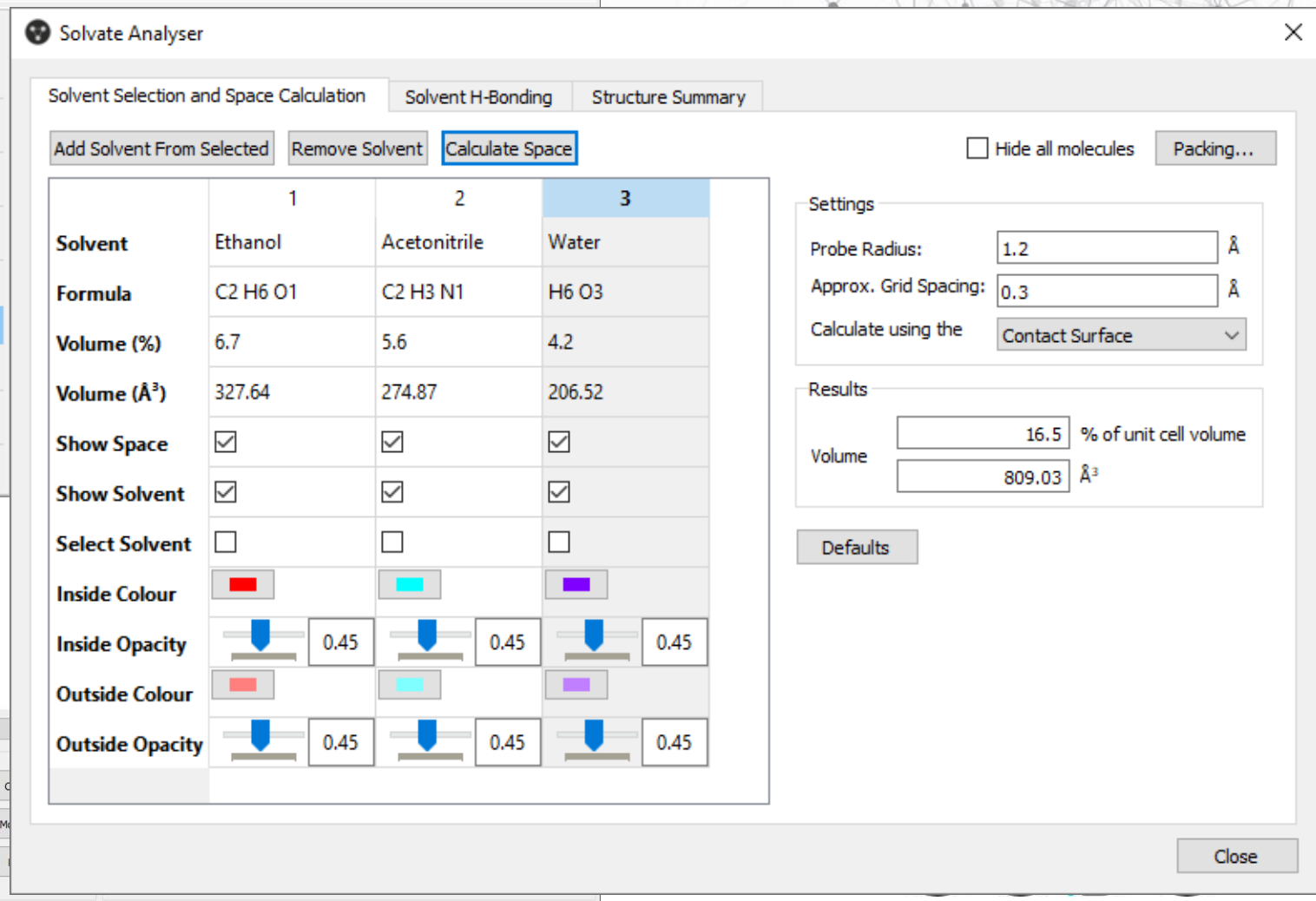


CCDC

Using the Solvate Analyser

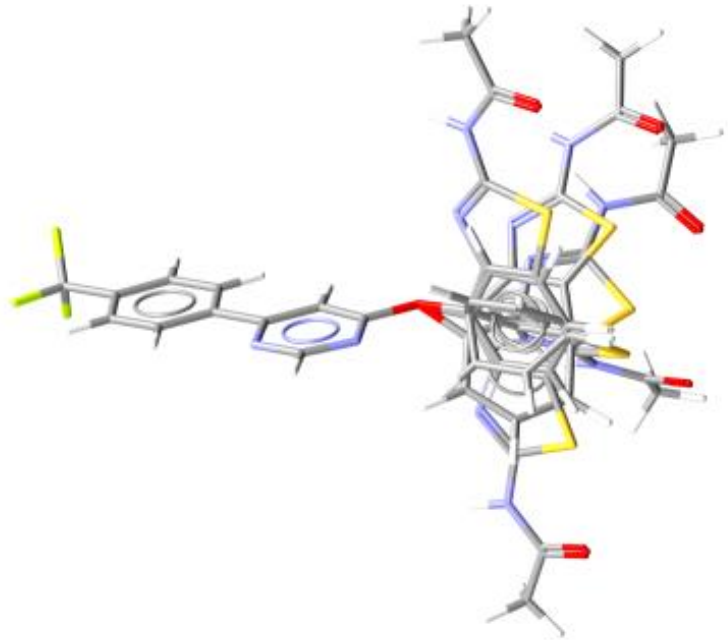
CSD Refcode: ZOJLIV





CSD Conformer Generator

- Generate conformers based on geometrical statistics from the CSD
- Visualise overlays of generated conformers
- Export conformers and further analyse e.g. in a co-crystal screen

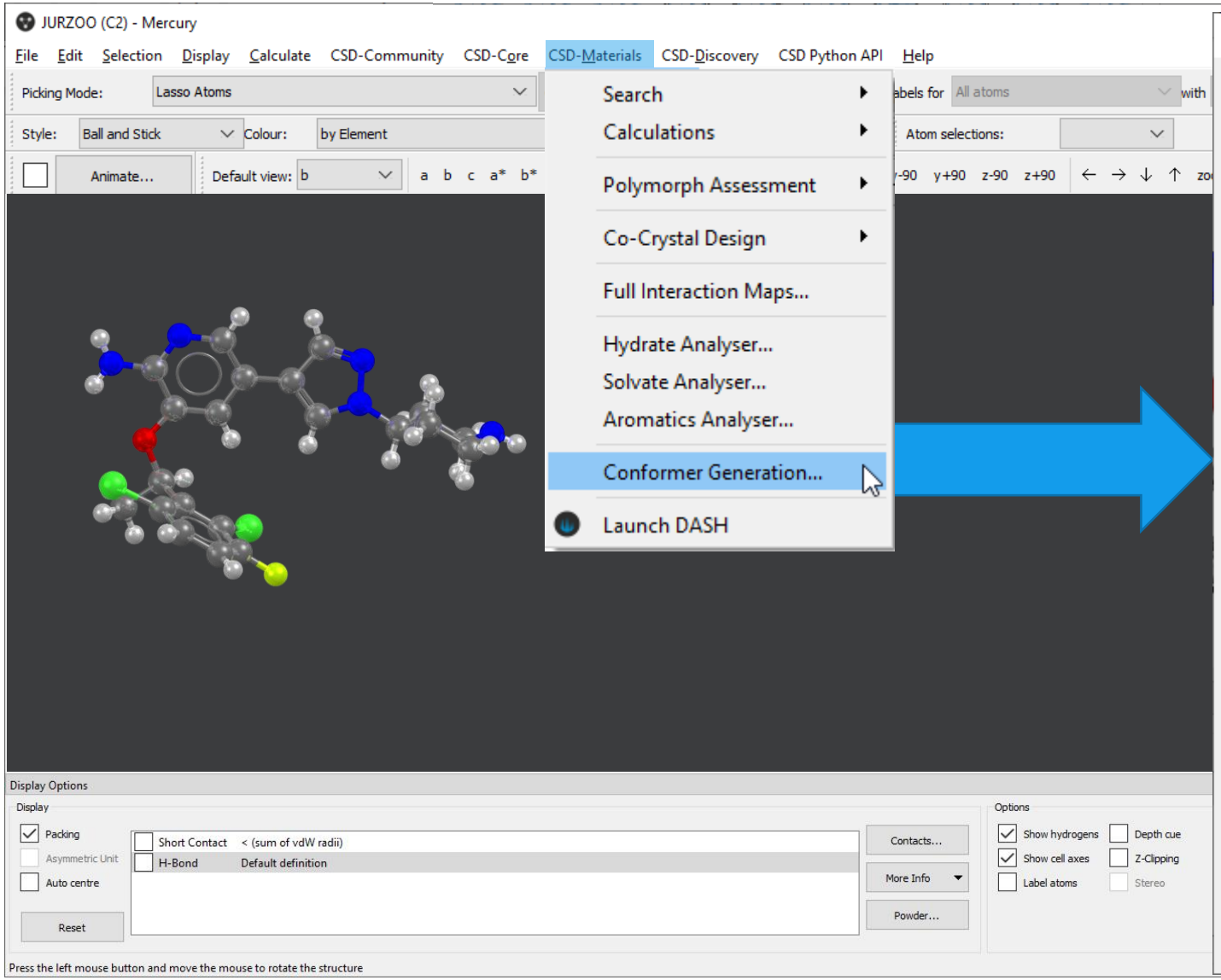


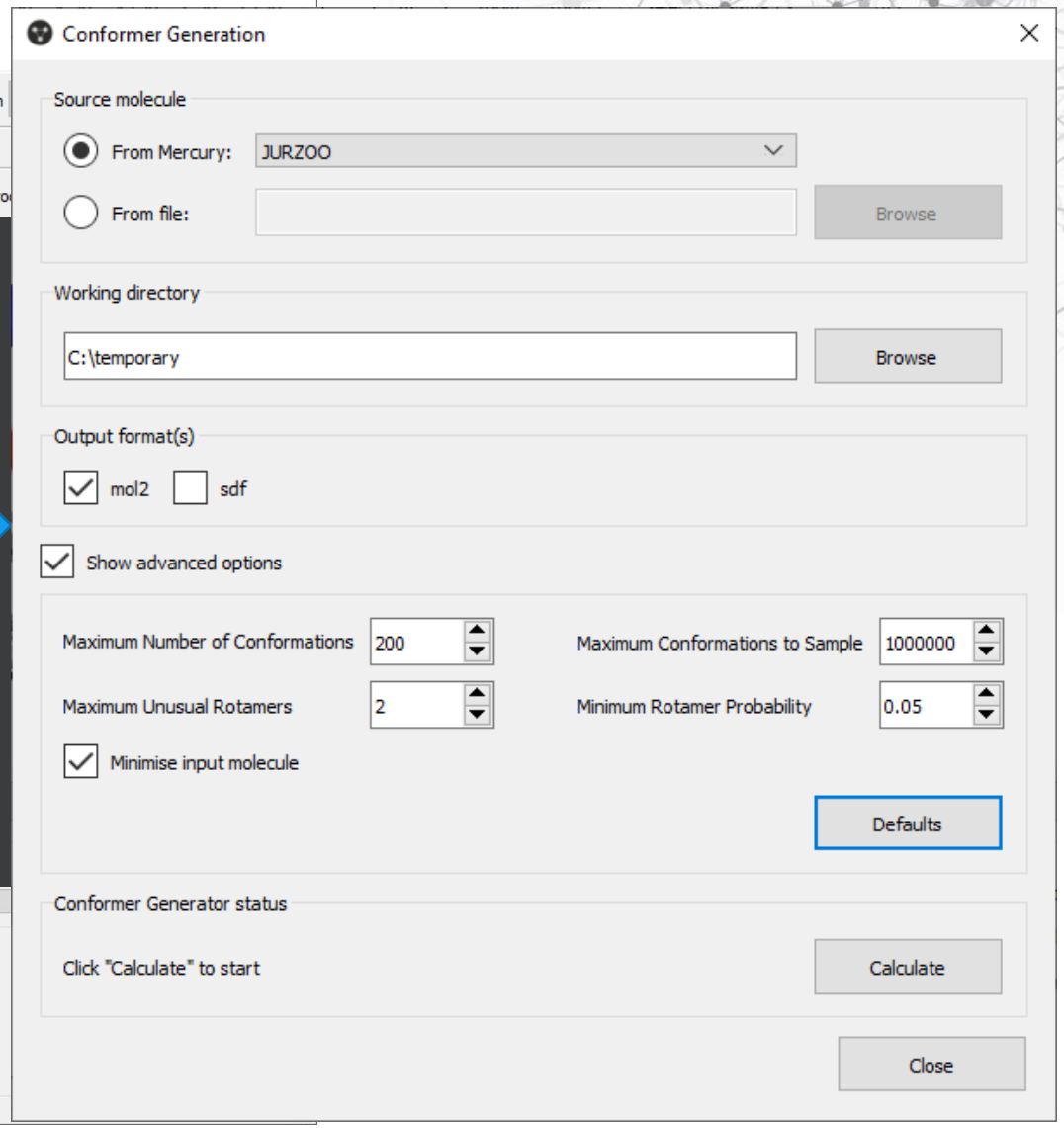
Ensemble of diverse conformations of AMG517 generated by the CSD conformer generator.

CCDC

Generating conformers

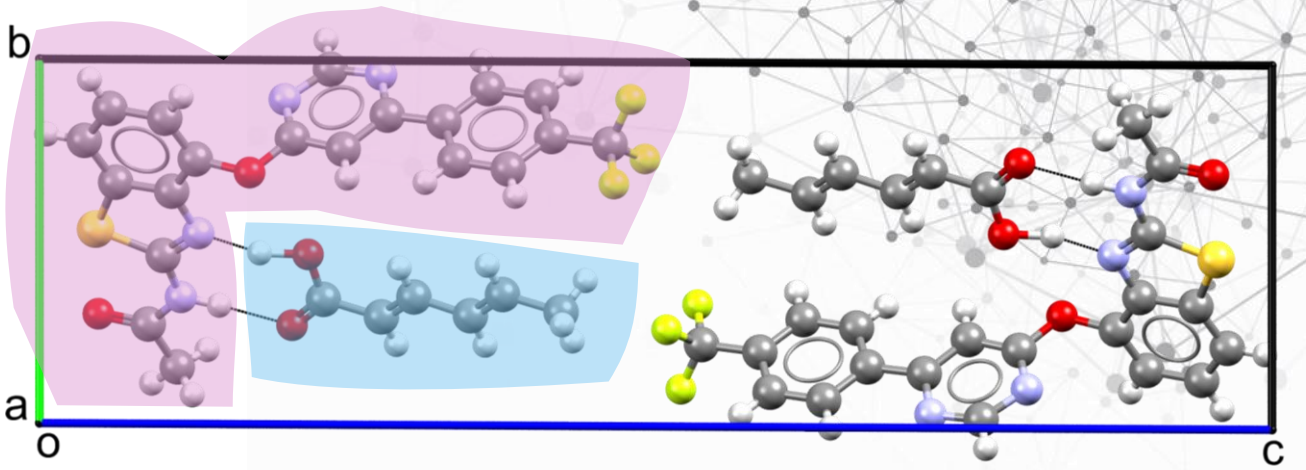
CSD Refcode: JURZOO





Molecular complementarity co-former screening

- Rapidly screen a set of co-formers against an Active Ingredient
- Assess how molecular descriptors vary across a set of molecules
- Submit multiple conformations of one or more molecules and compare them
- Designed to filter out molecules **unlikely** to form co-crystals

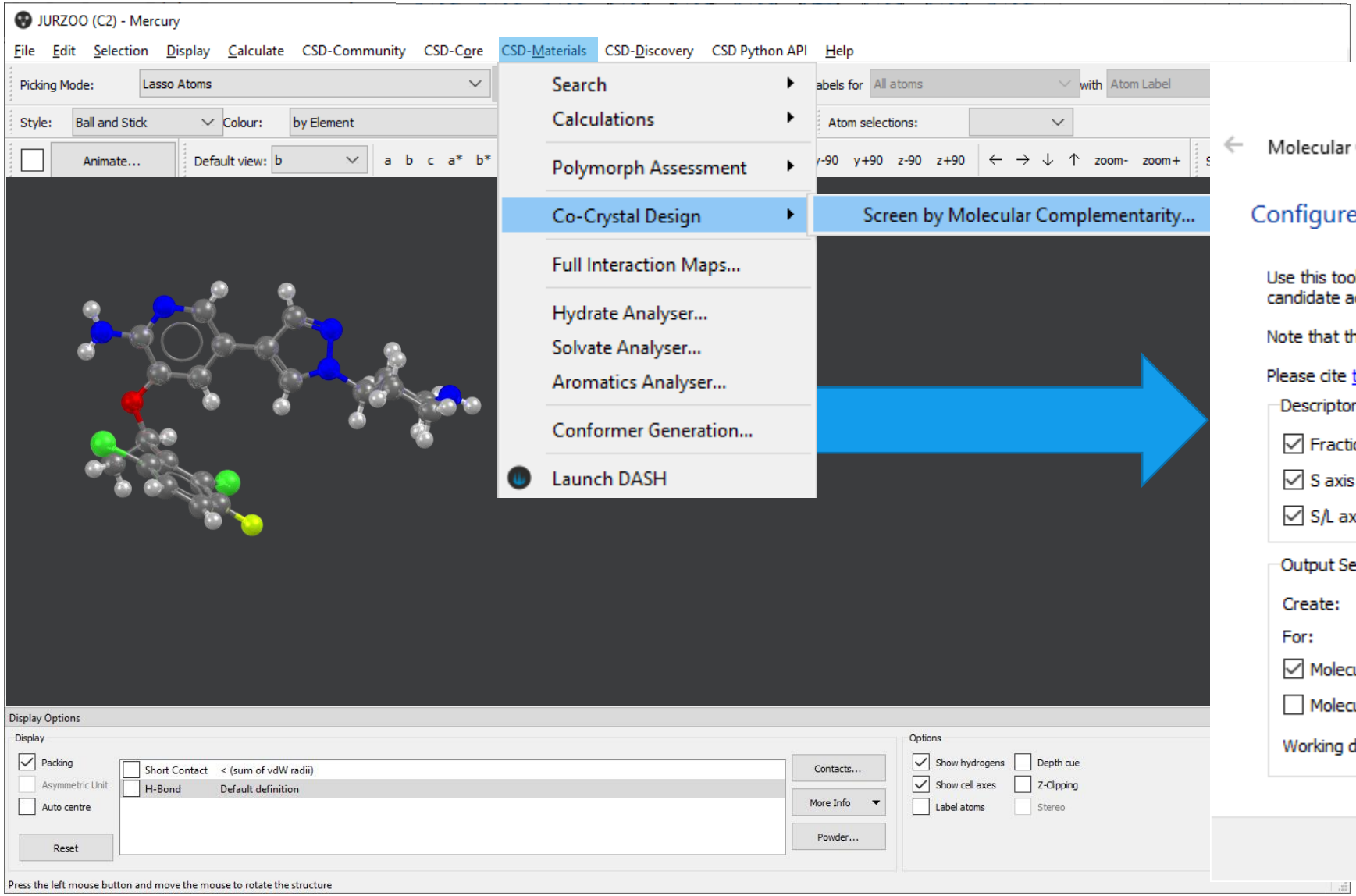


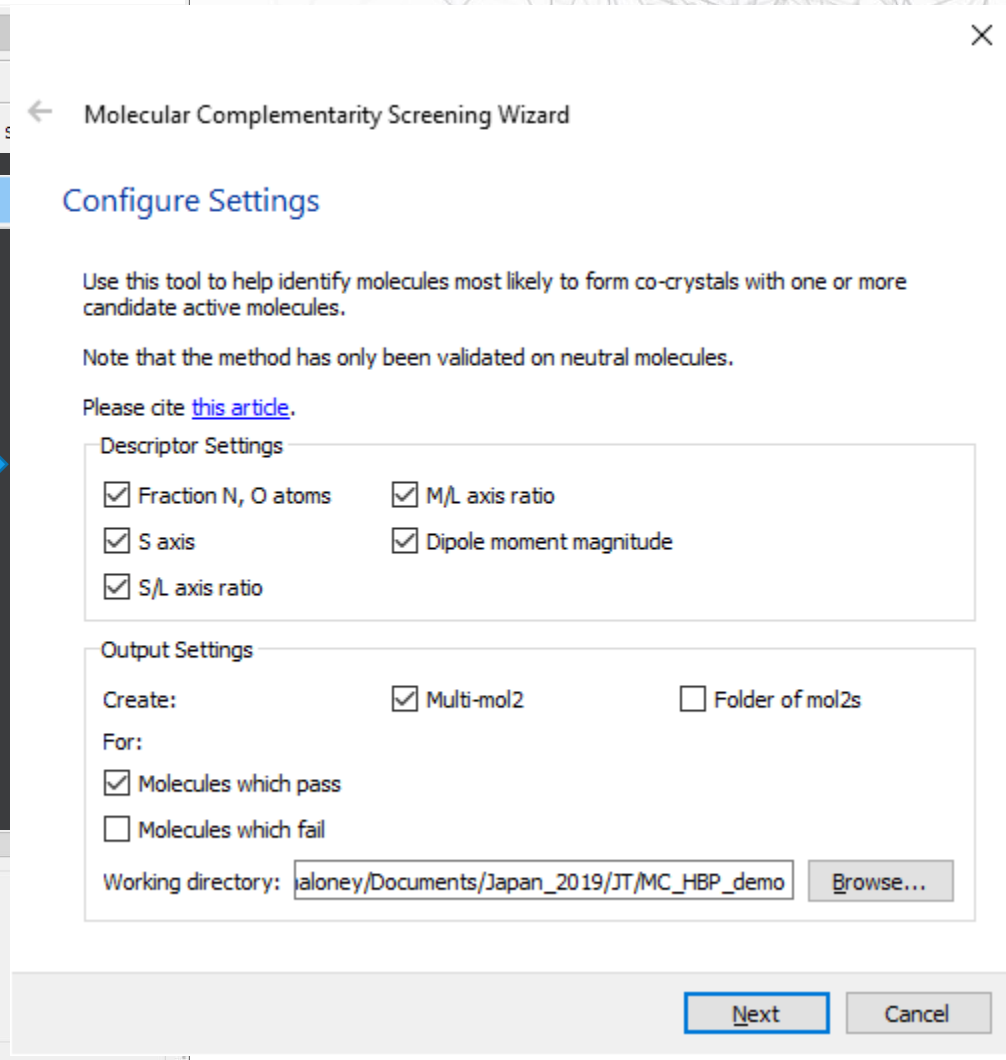
	ML axis ratio	S axis (/Å)	SL axis ratio	Dipole Moment (/Debye)	Fraction of NO
Sorbic Acid	0.50	4.17	0.38	1.03	0.25
AMG 517	0.66	6.46	0.35	1.65	0.20
AMG 517:Sorbic Acid |Delta|	0.16	2.29	0.03	0.62	0.05
|Delta|Pass Criteria	<0.32	<3.2	<0.28	<5.8	<0.29
AMG 517:Sorbic Acid |Delta| Meets Criteria	Yes	Yes	Yes	Yes	Yes

L. Fabian, Cryst. Growth Des., 2009, 9 1436-1443, DOI: 10.1021/cg800861m

CCDC

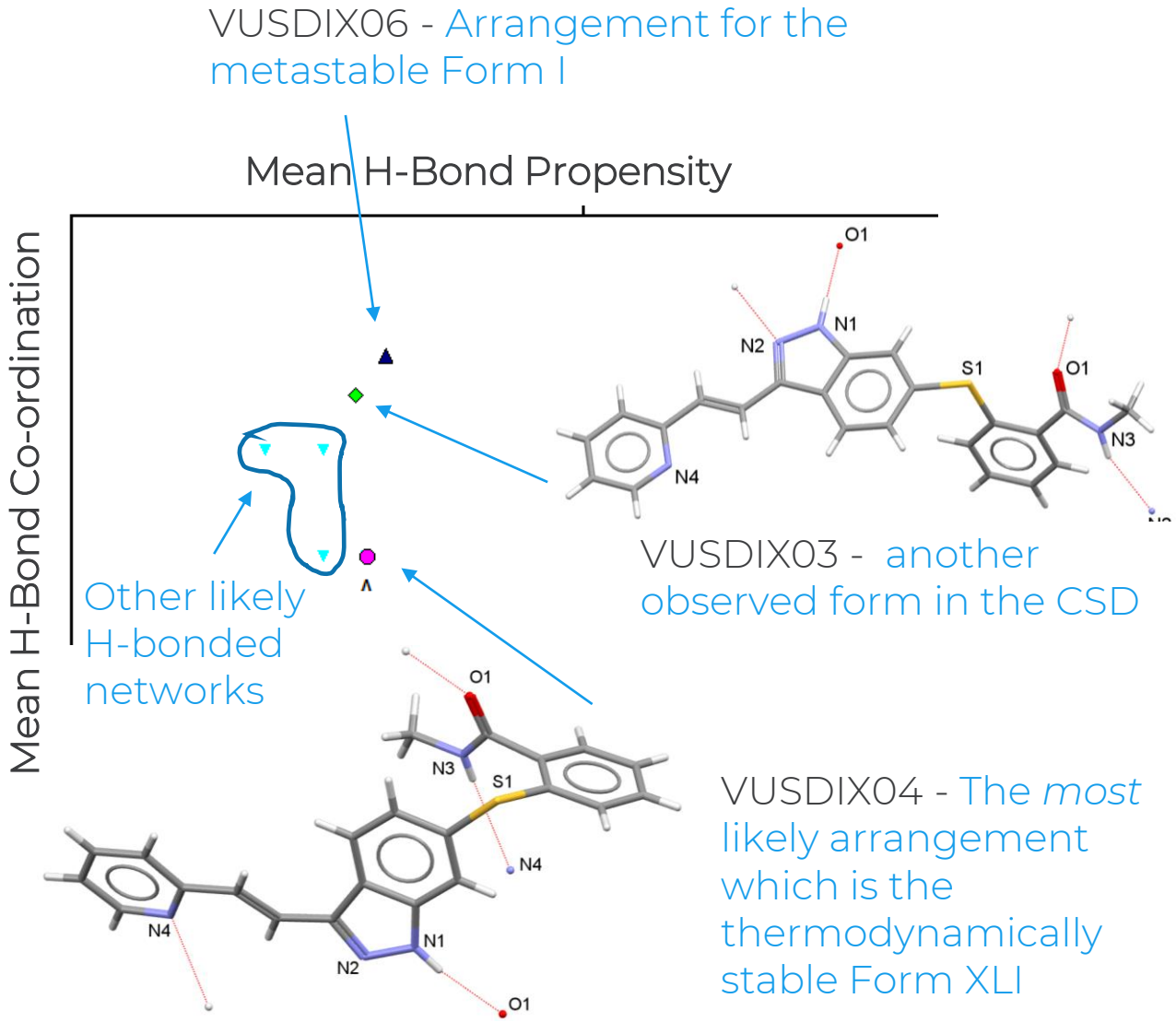
Performing molecular complementarity





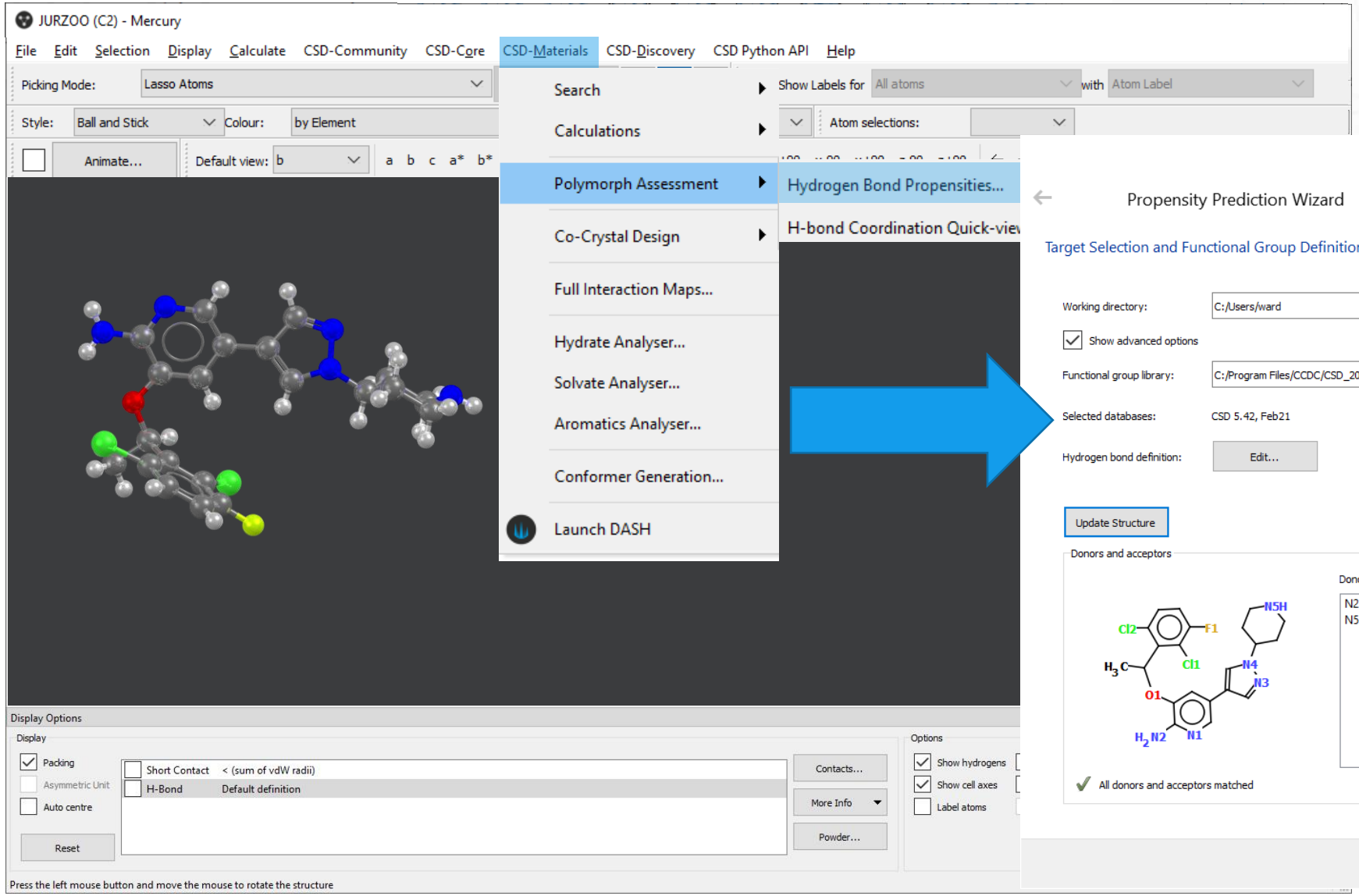
Hydrogen Bond Propensity

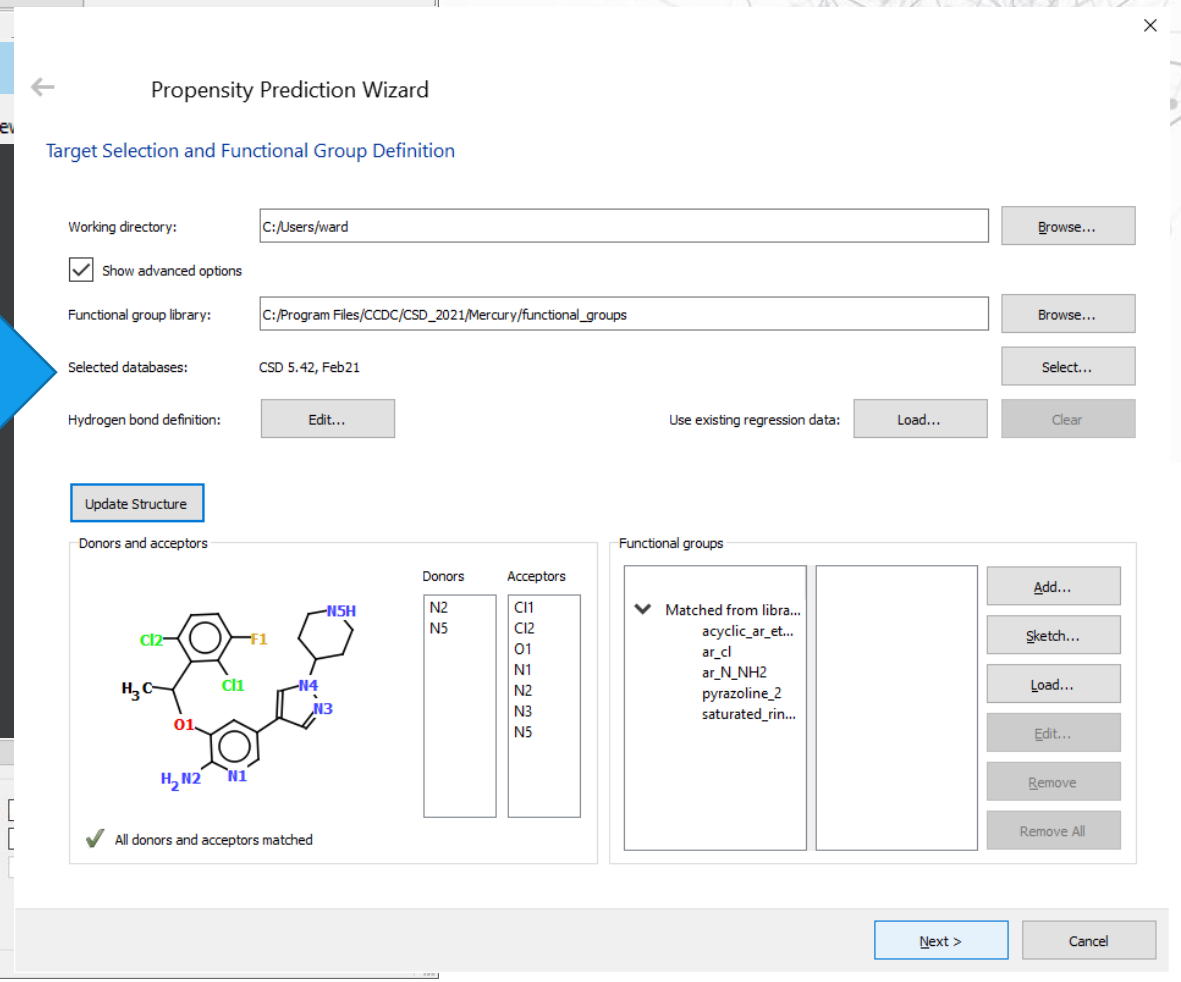
- Predict likely hydrogen bonds for a given molecule
- Assess crystal forms, e.g., by identifying sub-optimal hydrogen bonding
- Calculate hydrogen bond propensities for individual donor and acceptor groups
- Perform a comprehensive analysis of hydrogen bonding on a set of structures



Calculating a Hydrogen Bond Propensity

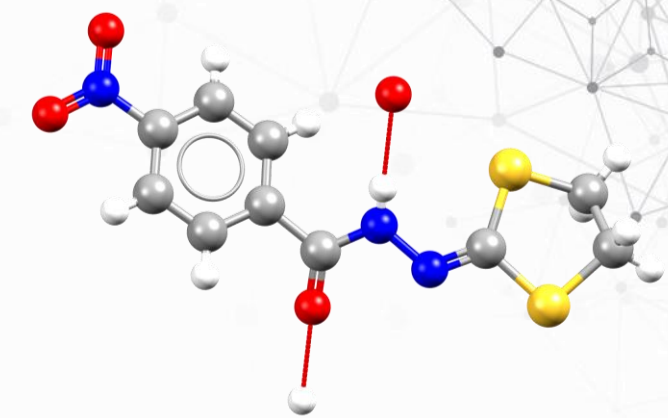
CSD Refcode: JURZOO

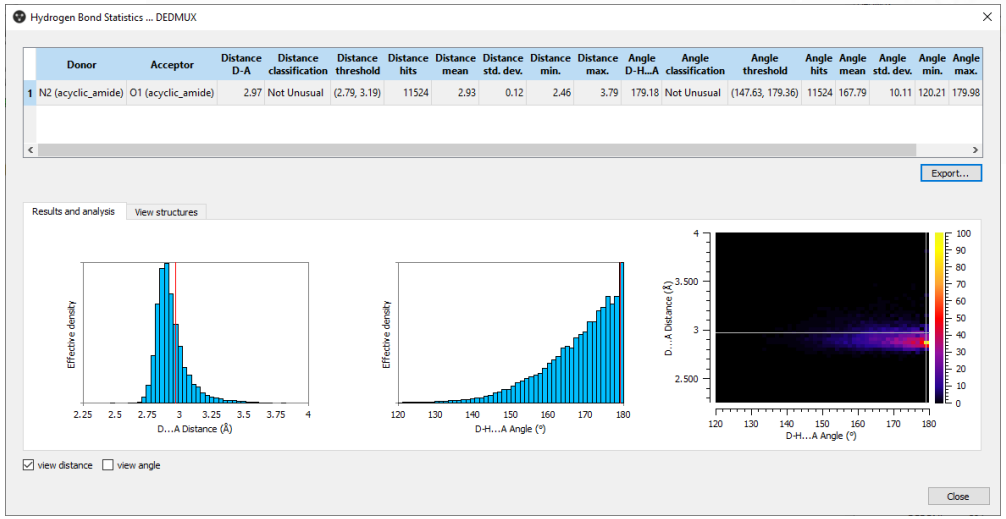




Hydrogen Bond Statistics

- Hydrogen Bond Statistics allows analysis of the usual- or unusual-ness of observed hydrogen bonds in a crystal structure.
- The interaction geometries are put in the context of the million+ crystal structures in the CSD.
- Contribute to assess the stability of a given crystal structure.

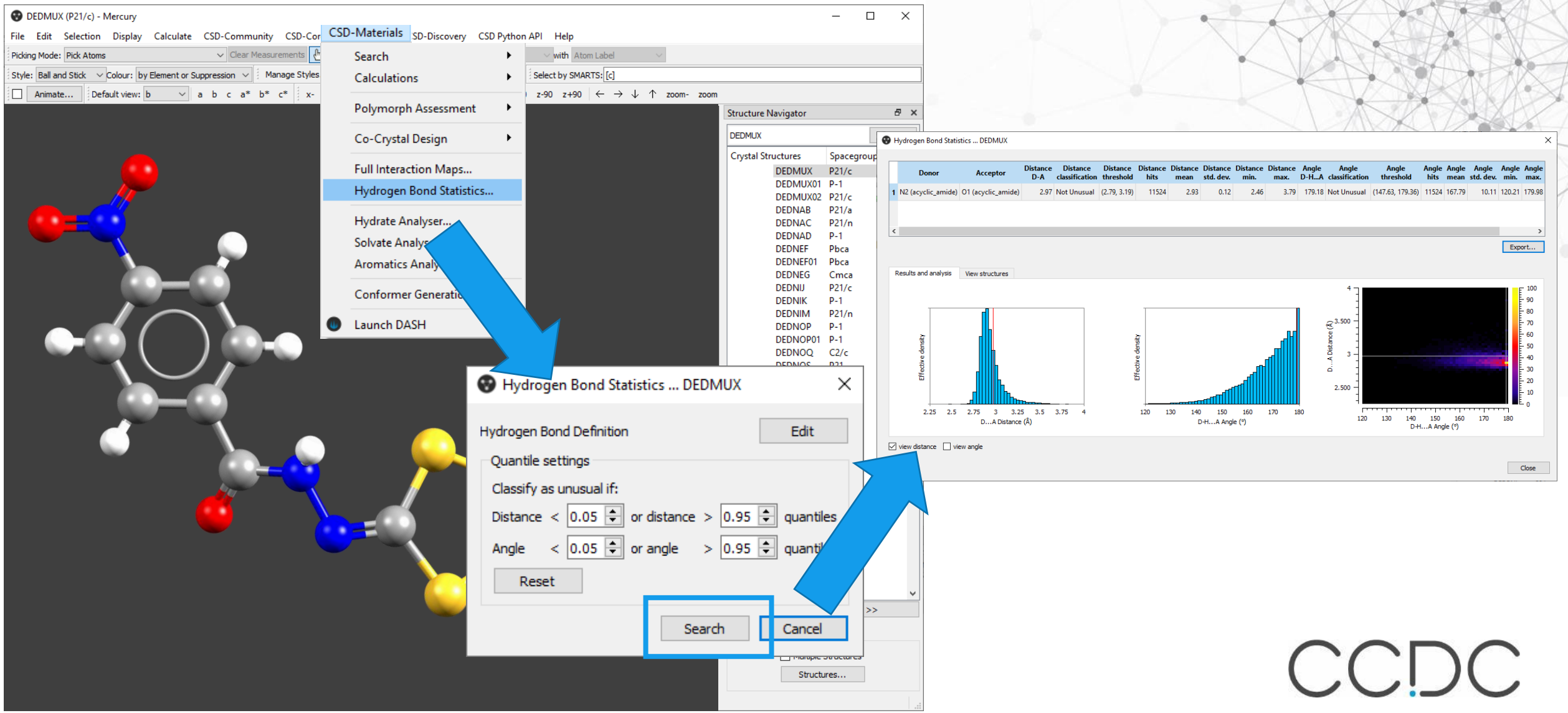




https://www.ccdc.cam.ac.uk/Community/blog/solid-form-stability-hydrogen-bond-statistics/

CCDC

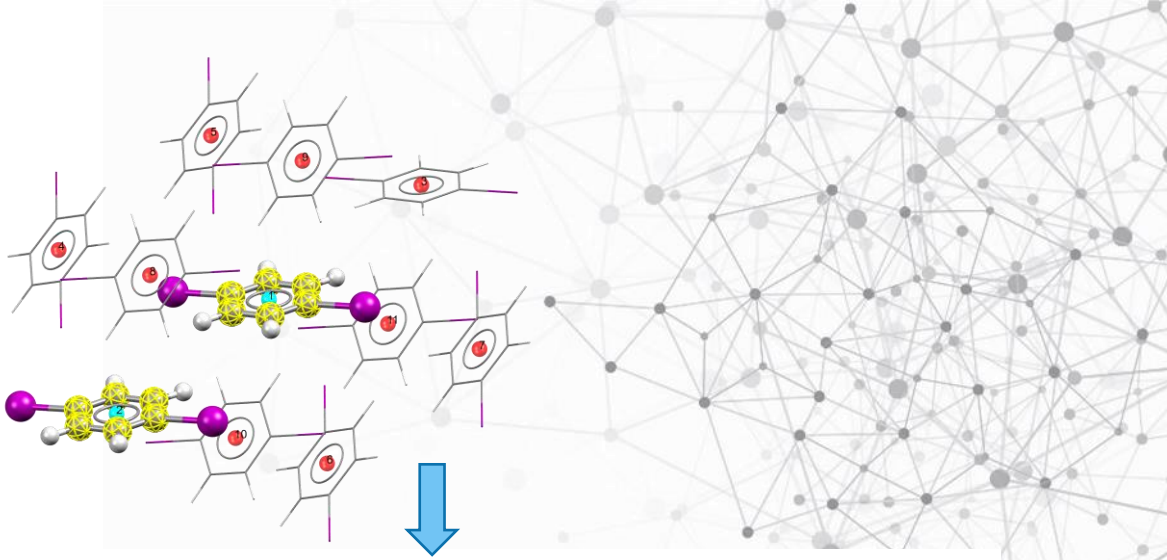
Calculating Hydrogen Bond Statistics



CCDC

Aromatics Analyser

- Intuitive visualisation & quantitative scoring of aromatic interactions
- Provides guidance on which geometries result in stabilising aromatic interactions
- Quantitative assessment provides score between 0 (no stabilising contribution) and 10 (an ideal aromatic interaction)
- Based on an Artificial Neural Network trained against DFT calculations



Aromatics Analyser... ZZZPRO03

Bond types may be edited using **Edit | Edit Structure...** from the main window

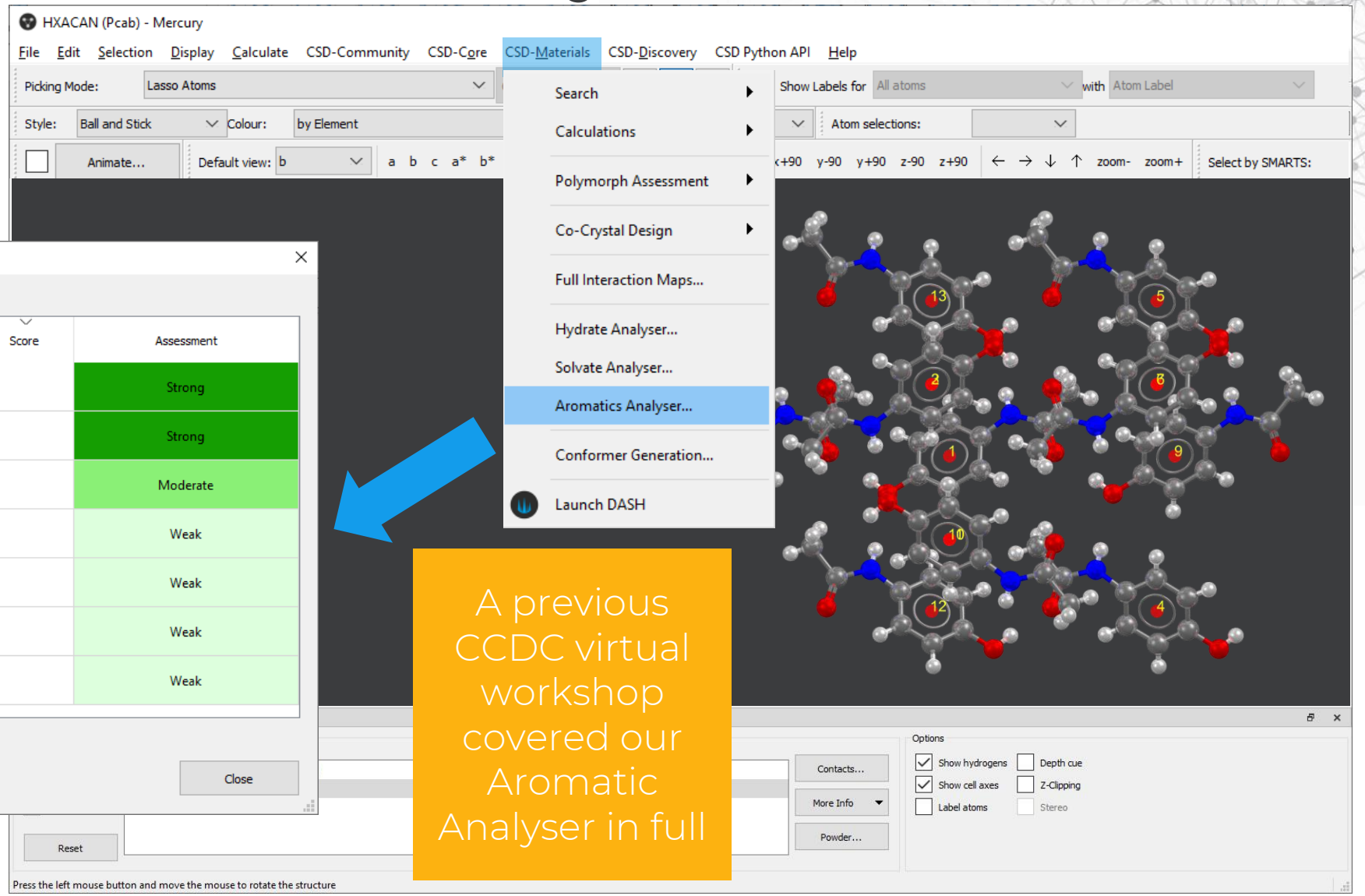
	Centroid1	Centroid2	Distance	Relative Orientation	Inter-olecula	Score	Assessment
1	1	8	4.79	46.23	Yes	8.3	Strong
2	1	9	4.79	46.23	Yes	8.3	Strong
3	1	10	4.79	46.23	Yes	8.3	Strong
4	1	11	4.79	46.23	Yes	8.3	Strong
5	1	2	6.17	0	Yes	5.2	Moderate
6	1	3	6.17	0	Yes	5.2	Moderate
7	1	4	9.04	43.74	Yes	0.5	Weak
8	1	5	9.04	43.74	Yes	0.5	Weak

☐ Include Intramolecular pairs ☐ Exclude symmetry equivalent interactions Calculate Export Atom info

Close

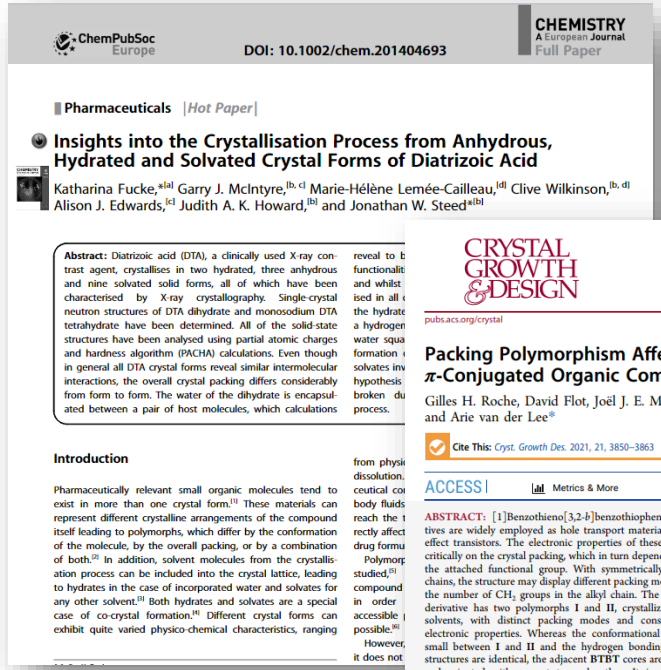
CCDC

Using the Aromatics Analyser



	Centroid1	Centroid2	Distance	Relative Orientation	Inter-molecular	Score	Assessment
1	1	2	4.65	58.43	Yes	8.9	Strong
2	1	10	4.87	50.79	Yes	8	Strong
3	1	12	5.94	26.95	Yes	5.9	Moderate
4	1	7	8.93	0	Yes	0.6	Weak
5	1	8	8.6	58.43	Yes	0.6	Weak
6	1	6	9.38	0	Yes	0.4	Weak
7	1	4	9.88	50.79	Yes	0.2	Weak

Scientific research using CSD-Materials



ChemPubSoc Europe — DOI: 10.1002/chem.201404693 — CHEMISTRY A European Journal Full Paper

Pharmaceuticals |Hot Paper|

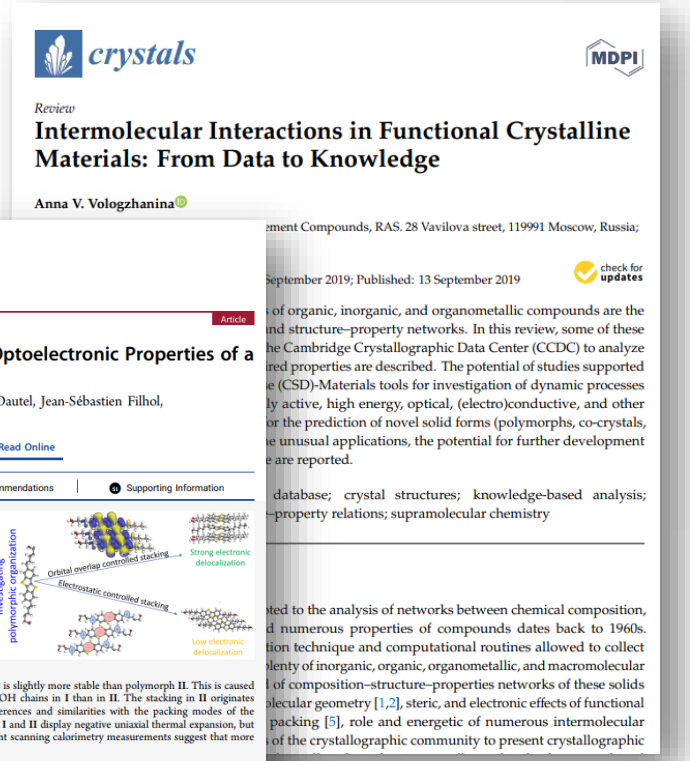
Insights into the Crystallisation Process from Anhydrous, Hydrated and Solvated Crystal Forms of Diatrizoic Acid

Katharina Fucke,*[a] Garry J. McIntyre,[b, c] Marie-Hélène Lemée-Cailleau,[d] Clive Wilkinson,[b, d] Alison J. Edwards,[c] Judith A. K. Howard,[b] and Jonathan W. Steed*[b]

Abstract: Diatrizoic acid (DTA), a clinically used X-ray contrast agent, crystallises in two hydrated, three anhydrous and nine solvated solid forms, all of which have been characterised by X-ray crystallography. Single-crystal neutron structures of DTA dihydrate and monosodium DTA tetrahydrate have been determined. All of the solid-state structures have been analysed using partial atomic charges and hardness algorithm (PACHA) calculations. Even though in general all DTA crystal forms reveal similar intermolecular interactions, the overall crystal packing differs considerably from form to form. The water of the dihydrate is encapsulated between a pair of host molecules, which calculations

Introduction

Pharmaceutically relevant small organic molecules tend to exist in more than one crystal form.[1] These materials can represent different crystalline arrangements of the compound itself leading to polymorphs, which differ by the conformation of the molecule, by the overall packing, or by a combination of both.[2] In addition, solvent molecules from the crystallisation process can be included into the crystal lattice, leading to hydrates in the case of incorporated water and solvates for any other solvent.[3] Both hydrates and solvates are a special case of co-crystal formation.[4] Different crystal forms can exhibit quite varied physico-chemical characteristics, ranging



crystals — MDPI

Review

Intermolecular Interactions in Functional Crystalline Materials: From Data to Knowledge

Anna V. Vologzhanina

ement Compounds, RAS. 28 Vavilova street, 119991 Moscow, Russia;

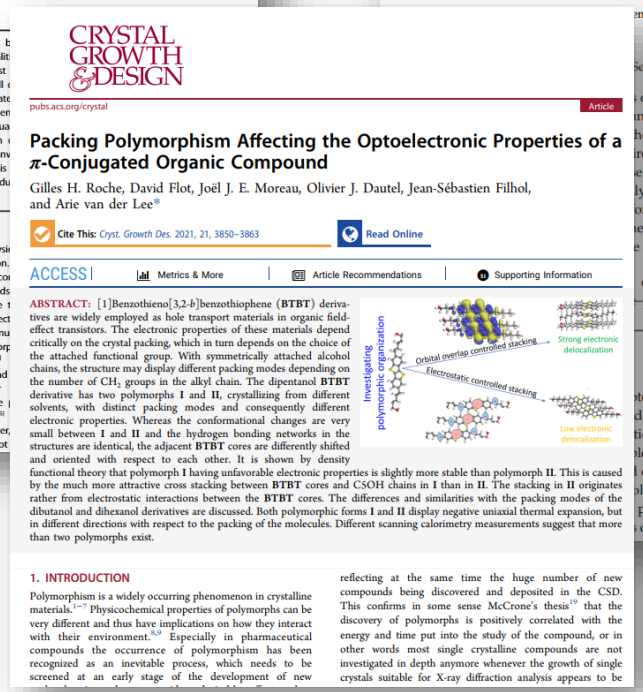
September 2019; Published: 13 September 2019

check for updates

of organic, inorganic, and organometallic compounds are the nd structure–property networks. In this review, some of these he Cambridge Crystallographic Data Center (CCDC) to analyze red properties are described. The potential of studies supported e (CSD)-Materials tools for investigation of dynamic processes ly active, high energy, optical, (electro)conductive, and other or the prediction of novel solid forms (polymorphs, co-crystals, e unusual applications, the potential for further development e are reported.

database; crystal structures; knowledge-based analysis; property relations; supramolecular chemistry

ted to the analysis of networks between chemical composition, d numerous properties of compounds dates back to 1960s. tion technique and computational routines allowed to collect lenty of inorganic, organic, organometallic, and macromolecular l of composition–structure–properties networks of these solids lecular geometry [1,2], steric, and electronic effects of functional packing [5], role and energetic of numerous intermolecular of the crystallographic community to present crystallographic



CRYSTAL GROWTH & DESIGN

pubs.acs.org/crystal — Article

Packing Polymorphism Affecting the Optoelectronic Properties of a π-Conjugated Organic Compound

Gilles H. Roche, David Flot, Joël J. E. Moreau, Olivier J. Dautel, Jean-Sébastien Filhol, and Arie van der Lee*

Cite This: *Cryst. Growth Des.* 2021, 21, 3850–3863 — Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: [1]Benzothieno[3,2-*b*]benzothiophene (**BTBT**) derivatives are widely employed as hole transport materials in organic field-effect transistors. The electronic properties of these materials depend critically on the crystal packing, which in turn depends on the choice of the attached functional group. With symmetrically attached alcohol chains, the structure may display different packing modes depending on the number of CH_2 groups in the alkyl chain. The dipentanol **BTBT** derivative has two polymorphs **I** and **II**, crystallizing from different solvents, with distinct packing modes and consequently different electronic properties. Whereas the conformational changes are very small between **I** and **II** and the hydrogen bonding networks in the structures are identical, the adjacent **BTBT** cores are differently shifted and oriented with respect to each other. It is shown by density functional theory that polymorph **I** having unfavorable electronic properties is slightly more stable than polymorph **II**. This is caused by the much more attractive cross stacking between **BTBT** cores and CSOH chains in **I** than in **II**. The stacking in **II** originates rather from electrostatic interactions between the **BTBT** cores. The differences and similarities with the packing modes of the dibutanol and dihexanol derivatives are discussed. Both polymorphic forms **I** and **II** display negative uniaxial thermal expansion, but in different directions with respect to the packing of the molecules. Different scanning calorimetry measurements suggest that more than two polymorphs exist.

Investigating polymorphic organization — Orbital overlap controlled stacking — Strong electronic delocalization — Electrostatic controlled stacking — Low electronic delocalization

1. INTRODUCTION

Polymorphism is a widely occurring phenomenon in crystalline materials.[1−7] Physicochemical properties of polymorphs can be very different and thus have implications on how they interact with their environment.[8,9] Especially in pharmaceutical compounds the occurrence of polymorphism has been recognized as an inevitable process, which needs to be screened at an early stage of the development of new

reflecting at the same time the huge number of new compounds being discovered and deposited in the CSD. This confirms in some sense McCrone's thesis[19] that the discovery of polymorphs is positively correlated with the energy and time put into the study of the compound, or in other words most single crystalline compounds are not investigated in depth anymore whenever the growth of single crystals suitable for X-ray diffraction analysis appears to be

Figure 2. Flowchart for the *CSD Motif Search*.

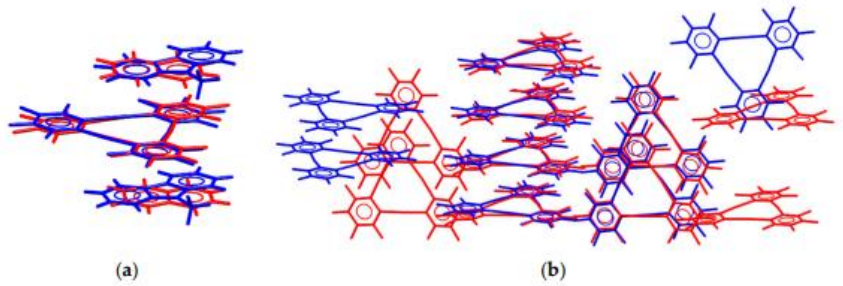


Figure 3. Examples of the (**a**) *Packing Feature* and (**b**) *Packing Similarity* search hits. A sandwich of a TPPM with two aromatic C_6 rings was constructed as the *Packing Feature* from a reference (red) {MOXMIV}[a] molecule and was also found in (blue) {QATTAH}. The *Packing Similarity* comparison of an orthorhombic polymorph of TPPM (blue) {MOXMAN02} with its' three monoclinic polymorphs gives the best similarity with (red) {MOXMAN03}. [a] Here and below a six-letter CSD-Refcode of a compound is given in Figure braces.

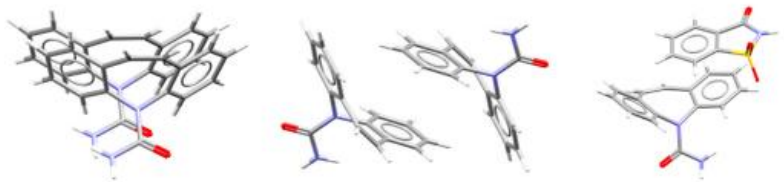


Figure 4. Carbamazepin molecules packed in (**a**) "translation stacks" in {CBMZPN11}, (**b**) "inversion cups" in {CBMZPN01} and (**c**) "co-former pairs" in {UNEZAO}.

https://pubs.acs.org/doi/abs/10.1021/acs.cgd.1c00177
https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/chem.201404693
https://www.mdpi.com/2073-4352/9/9/478

CCDC

What have we learnt?

- How informatics and data-driven approaches can be used to understand solid form design and risk assessment.
- How to use Motif searching to investigate motifs frequency in the CSD.
  - Using a pre-defined motif or a bespoke one.
- How to set up and run crystal packing feature searches.
  - How to analyse the results generating plots.
- How to perform crystal packing similarity searches.
  - On a pair of structures (e.g., anhydrous-hydrate).
  - Within a refcode family to identify polymorphs.
- What other tools available in the CSD-Materials suite can be used to do.

CCDC

https://www.ccdc.cam.ac.uk/Community/educationalresources/

Want to explore more?

Training and Educational Resources

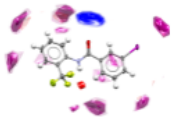
The wealth of information contained within the Cambridge Structural Database (CSD) extends far beyond a collection of crystal structures. Knowledge derived from these materials informs much of chemistry, biochemistry, and biology. Chemical and structural concepts are often difficult to grasp without real world, interactive examples for students to explore.

The CCDC and our colleagues continually produce educational materials for use in classroom and computer lab settings, or as independent study modules. Many of these materials make use of the Teaching Subset - a freely available set of over 750 structures that can be investigated with the free version of our Mercury visualisation and analysis program. Of course, our database of over one million entries are available for free through our Access Structures portal.

If you are an educator looking for supplementary teaching materials, find out more about the Teaching Database **here**. If you have developed your own modules using the CSD and would like to share them with the broader community, please contact us at education@ccdc.cam.ac.uk.

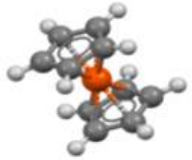
To keep up to date with the latest news from education and outreach at the CCDC, sign up for the Education and Outreach Newsletter **here**.

Register for E&O newsletter

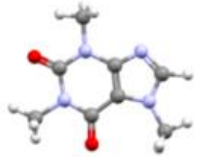


CSD-Materials

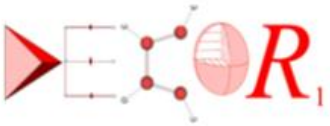
Tools to help you to understand your material's behaviours and refine its properties.



Information on the Teaching Subset



Access a series of teaching modules for use in the classroom



DECOR: Educational Resources for Teaching Crystallography

Self-guided workshops

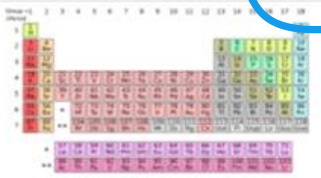
Download a series of self-guided workshop materials for CCDC tools and features

Watch software training and support videos

CSDU modules

On-demand modules with completion certificate

Access fun science activities for kids through the CCDC Home learning page



Explore the Periodic Table through Crystal Structures

YouTube and LabTube channels

CCDC