Advanced Structural Analysis of Porous Materials Using Mercury

CCDC Virtual Workshop

7th October 2025

REC

CCDC

advancing structural science

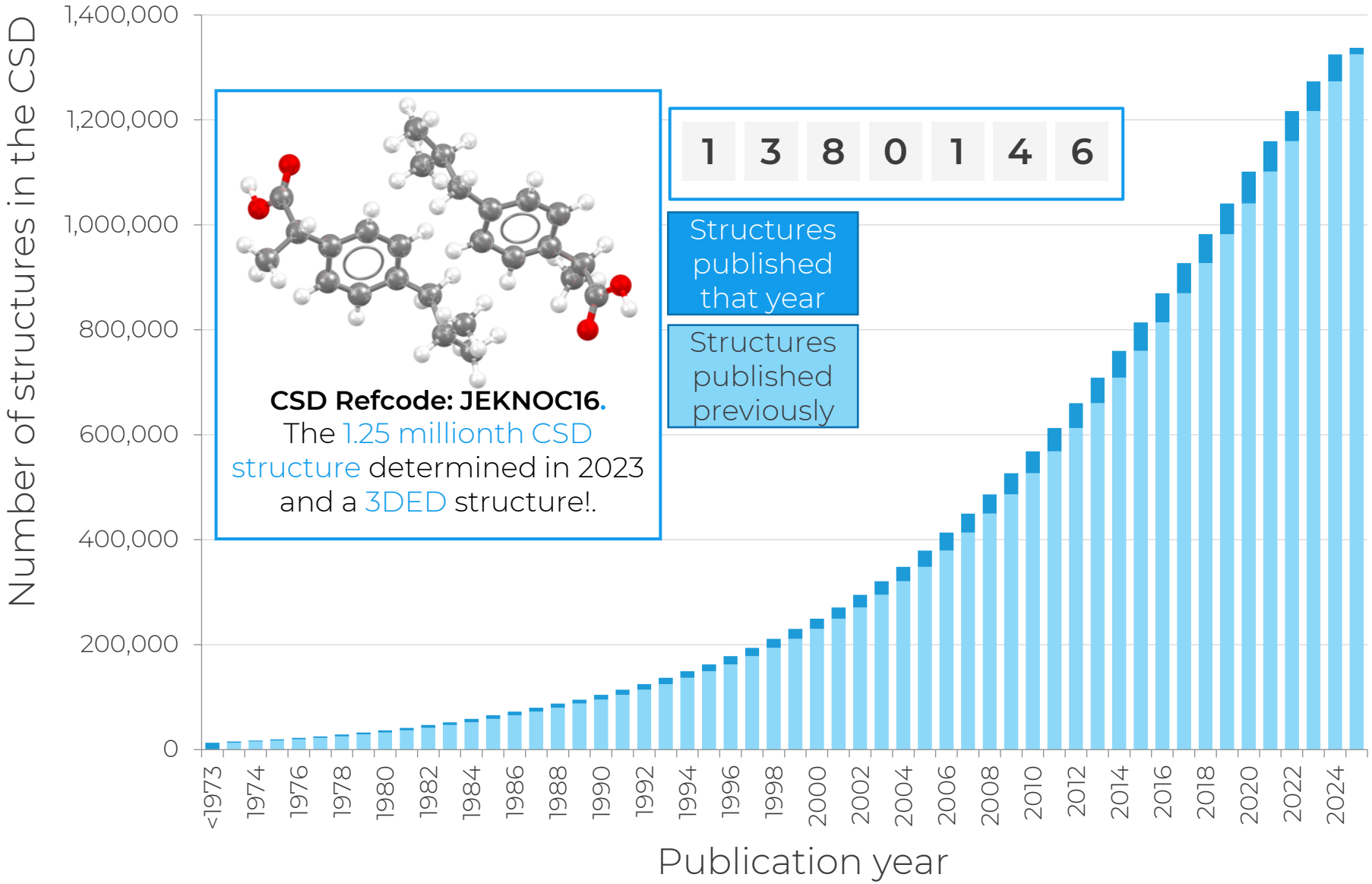
Learning outcomes

- Become familiar with the **CSD**.
- Which **subsets** of the CSD are available, how to access them in ConQuest, Mercury and the CSD Python API, and how they can help you in your research.
- How to **visualise** porous materials effectively in **Mercury**.
- How to **analyse** pore space in your structures using **Pore Analyser**.
- About applications of the **disorder visualization feature**, including recent advances and integrations with other functionality available in Mercury.
- And more Mercury **tips and tricks**!

We will make the recording available to you in the next few days.

CCDC

The Cambridge Structural Database

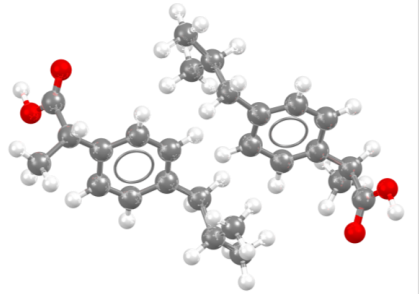


- Every published structure
  - Inc. ASAP & early view
  - *CSD Communications*
  - Patents
  - University repositories
  - Thesis
- Every entry enriched and annotated by experts
- Discoverability of data and knowledge
- Sustainable for 60 years
- A trusted CoreTrustSeal repository

Certified as Trustworthy by CoreTrustSeal

CCDC

CSD Refcodes



CSD Refcode - JEKNOC16

What is JEKNOC16?

- 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

Porous materials refcode families

- Different solvents or adsorbates would be assigned to *different* refcode families, even thoug the framework might be the same.

Some interesting refcodes are: KITTEN, BATMAN, DISNEY, GAUTAM, GLYCIN

CCDC

Curation

A combination of manual and automated curation

Identifier assigned
CCDC number, CSD Refcode
CCDC 236078, CSD-ETEXIK

Information extracted from CIF

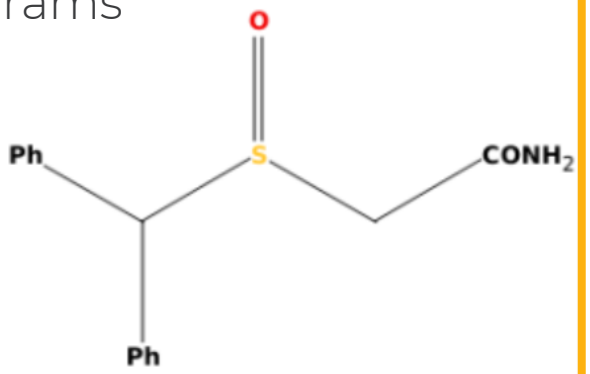
Spacegroup: $P2_1/n$
Cell: *a* 20.9610(10)Å *b* 9.7061(5)Å *c* 20.8940(10)Å, α 90° β 139.3360(10)° γ 90°
R factor: 4.2%

Structural: Atom coordinates, ADPs

Experimental information: Temperature, recrystallisation solvent, crystal colour, crystal habit

Naming: 2-((Diphenylmethyl)sulfi
Synonym: 2-(Benzhydrylsulfinyl)a

Chemically meaningful 2D Diagrams



Linking: DrugBank: DB00745
For Stereoisomer see KAHVUL

New for 2025

- Publication titles
- Enhanced refinement properties (including wR_2, GooF, Max Shift, Residual density max/min)
- Predicted properties (including HOMO-LUMO gap, Singlet and Triplet state energies)
- Calculated properties (including total surface area, geometric volume, pore limiting diameter, max pore diameter)

Information added
- Melting points
- Bioactivity details
- Natural source data
- Oxidation states
- Polymorph families

https://www.ccdc.cam.ac.uk/discover/blog/how-your-structure-is-curated-in-the-csd/

Inside the Cambridge Structural Database

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

Organic
45%

Metal-Organic
55%

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

Not Polymeric
88%

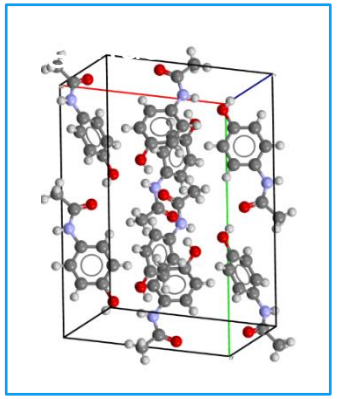
Polymeric: 12%

Single Component
58%

Multi Component
42%

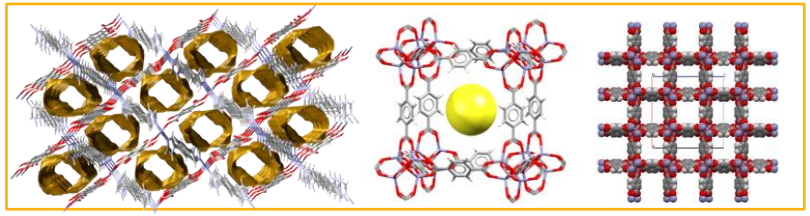
Organic

- Drugs
- Agrochemicals
- Pigments
- Explosives
- Protein ligands



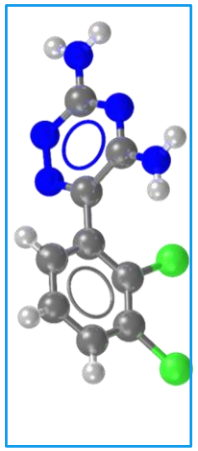
Metal-Organic

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



Additional data

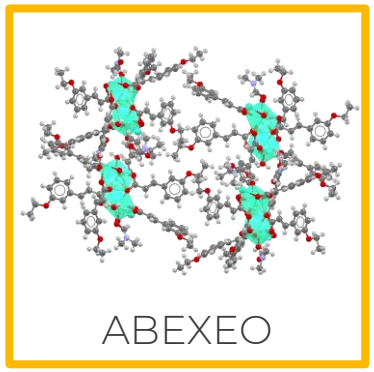
- 14,019 polymorph families
- 176,704 melting points
- 1,127,300 crystal colours
- 1,005,205 crystal shapes
- 32,100 bioactivity details
- 14,344 natural source data
- > 400,000 oxidation states

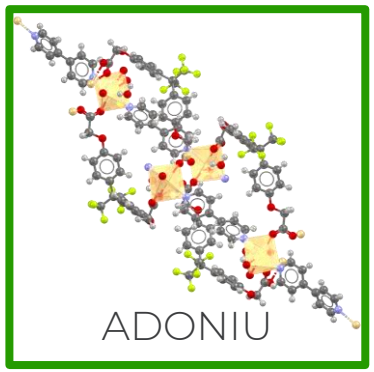


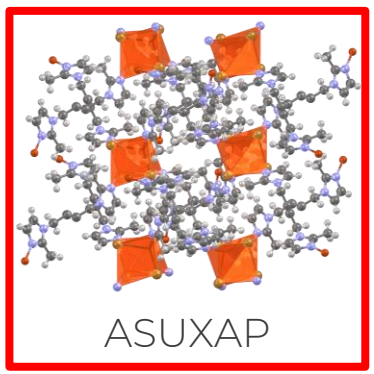
Links and subsets

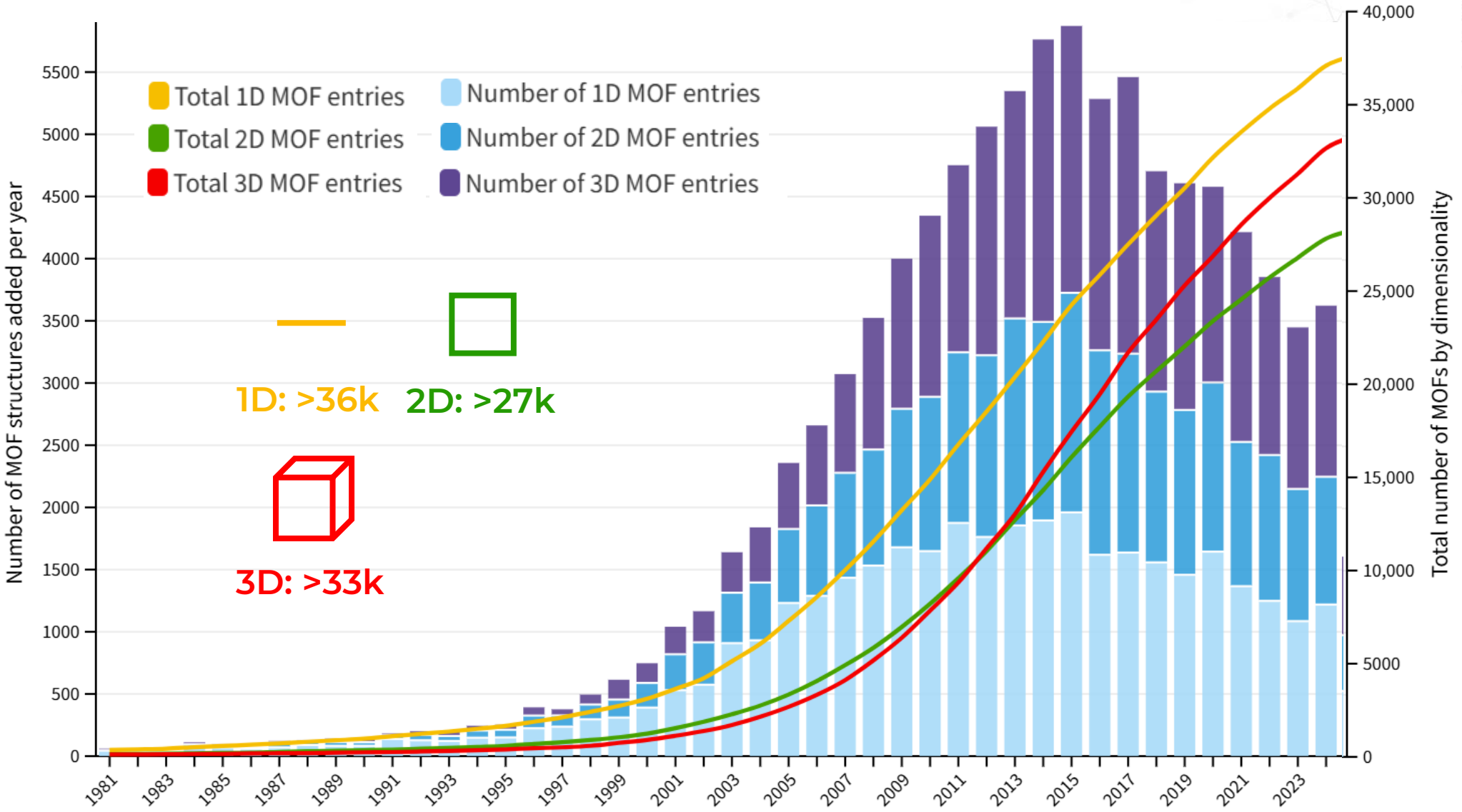
- DrugBank
- Druglike
- MOFs
- PDB ligands
- PubChem
- ChemSpider
- Pesticide PDB

Inside the CSD

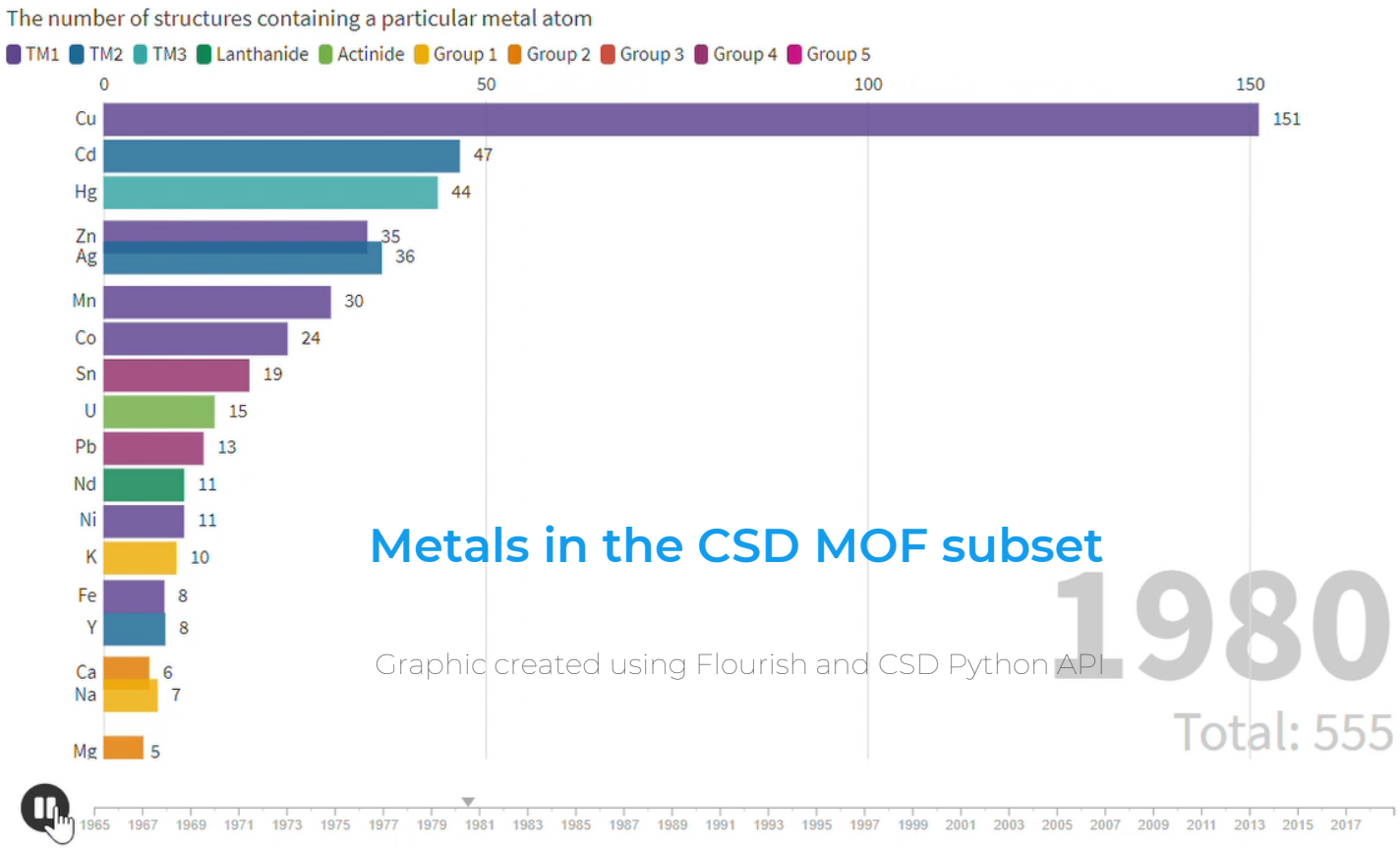

ABEXEO


ADONIU


ASUXAP



Exploring coordination polymers in the CSD



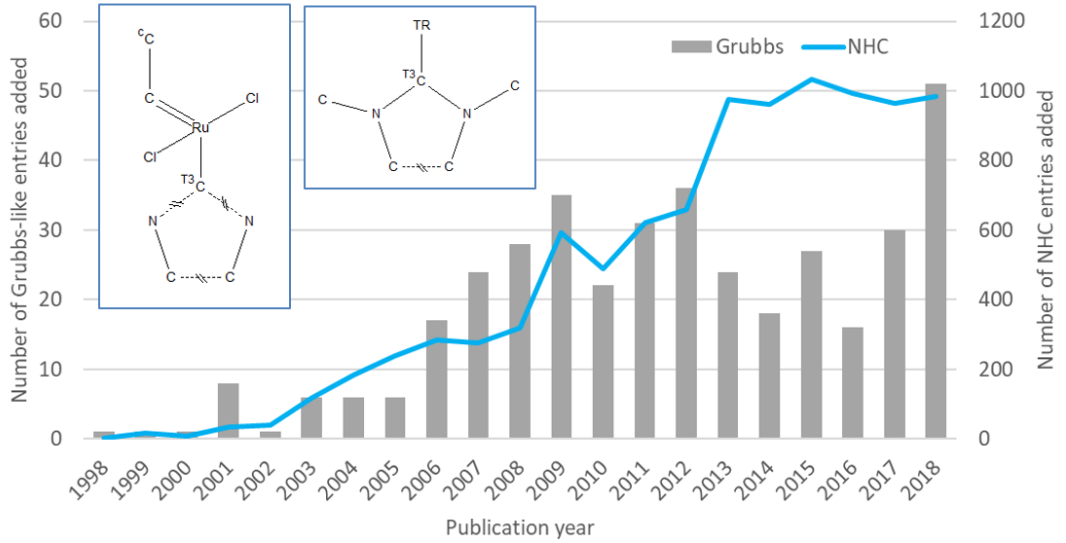
- Over **100k** have been synthesized.
- Growing **potential applications**, each with unique property requirements.
- Materials informatics using data in the CSD can help to **focus research efforts** to find the right MOFs for each application.

CCDC

Industrial applications

- Energetic materials
- Ferroelectricity
- Catalysis
- Gas storage
- Gas separation
- Detection and sensing
- Batteries
- Drugs
- Semiconductors
- Filtration
- Carbon capture

Catalysts in the CSD



Using materials informatics to support materials development

Focus experimental efforts

Predict future successes and risks

Identify **patterns** to spark further investigation

Systematically **search** the existing landscape

Data is the fuel

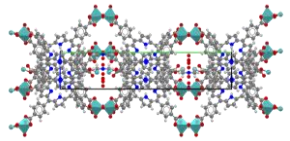
CCDC

The CSD Portfolio today

CSDCore.

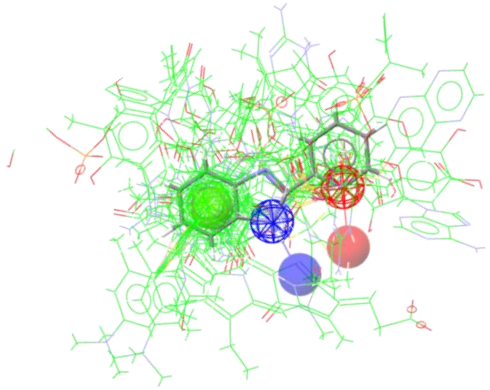
Search, visualise, analyse and communicate structural data
Insights into molecular and crystal shape and interactions





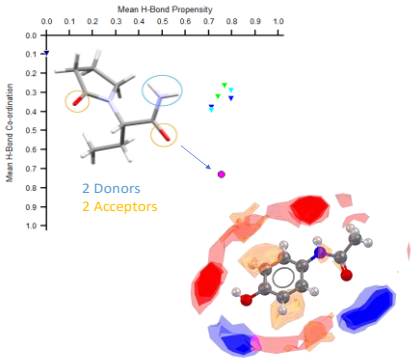
CSDDiscovery.

Design of new molecules



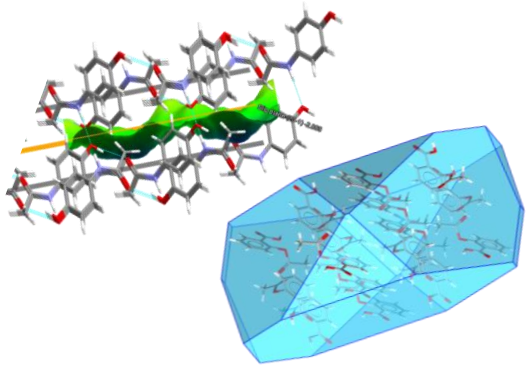
CSDMaterials.

Assessment of solid form stability and properties



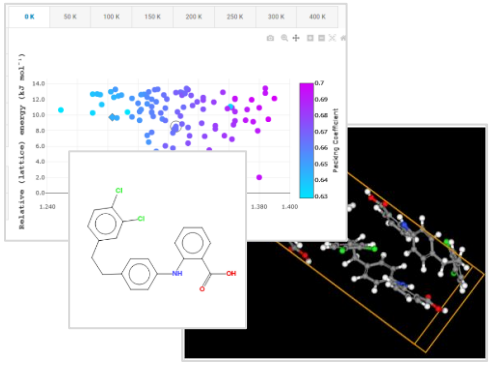
CSDParticle.

Anticipate particle properties and behaviour



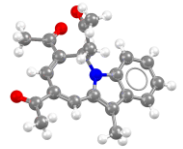
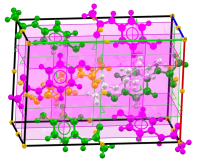
CSDTheory.

Insights from predicted structure landscapes



CSDCommunity.

Deposit, publish, access and visualise structural data
Free functionality to share and learn from structures

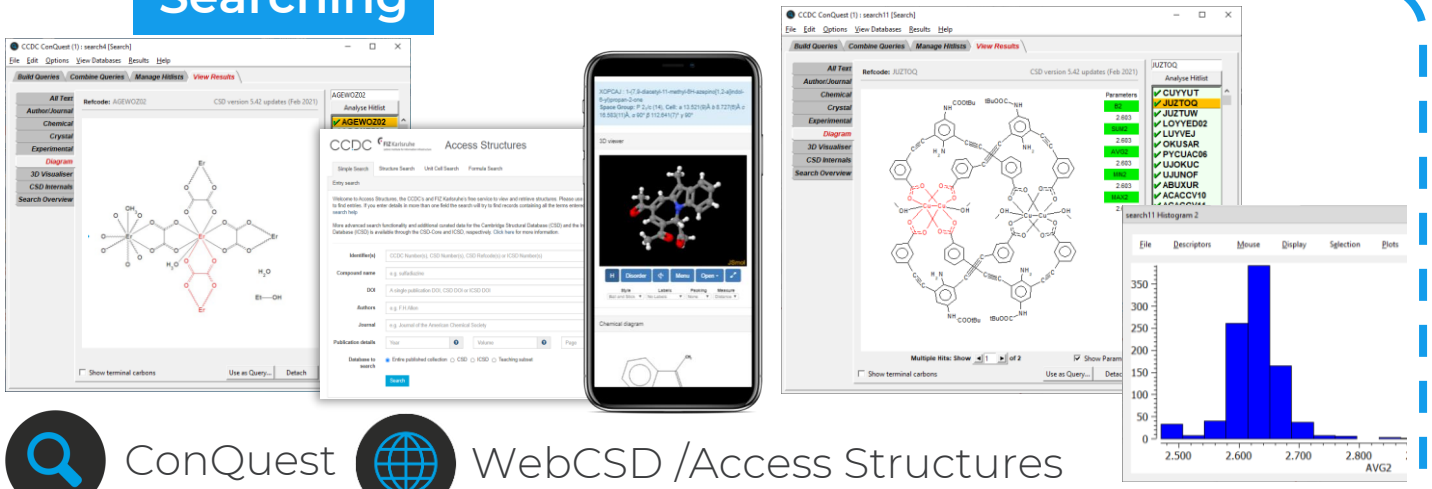



Medicinal & Computational Chemists ♦ Crystallographers & Structural Biologists ♦ Solid Form & Crystallisation Scientists ♦ Functional Materials Scientists ♦ Educators ♦ Industry and Academia

CCDC

Using the CSD for porous materials

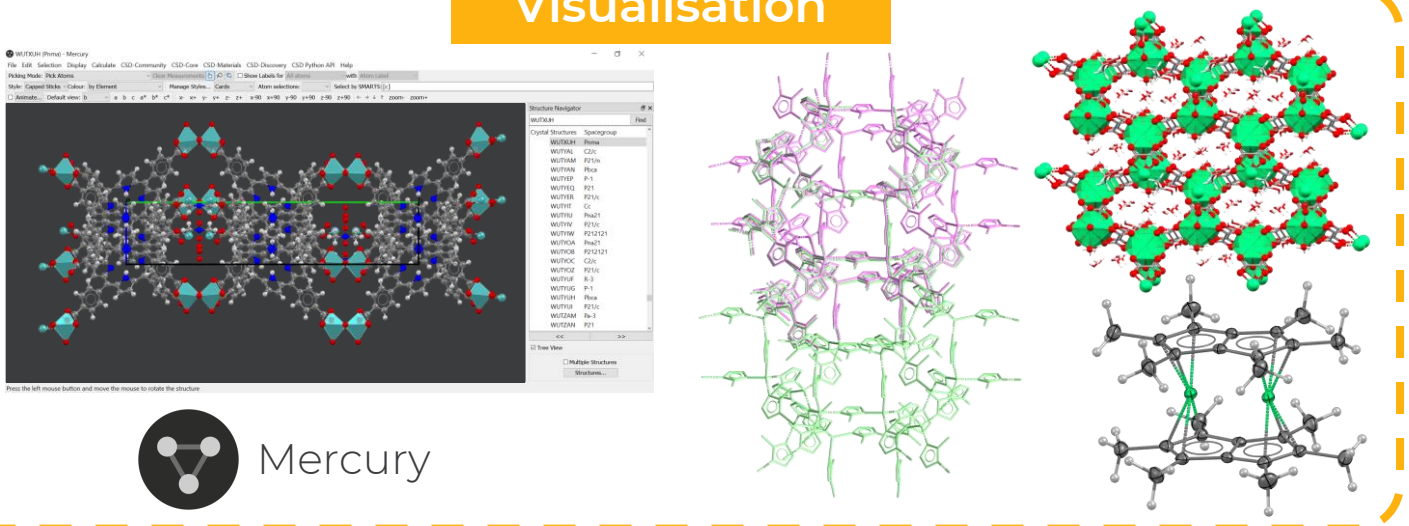
Searching



ConQuest

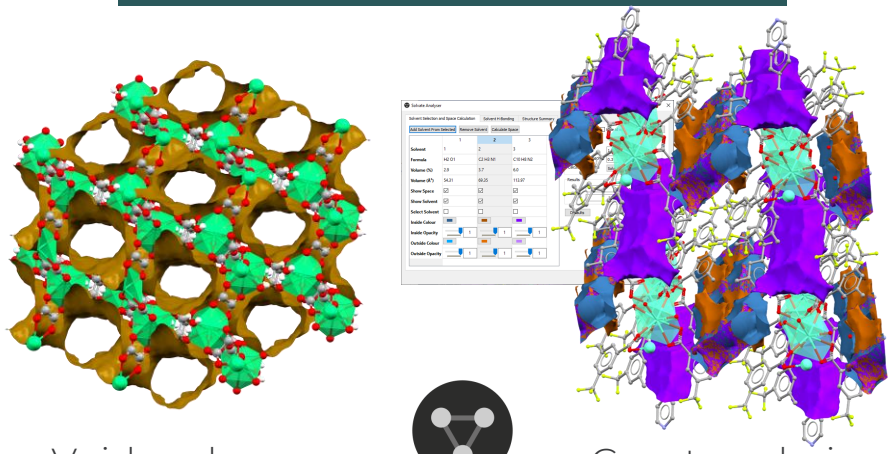
WebCSD /Access Structures

Visualisation



Mercury

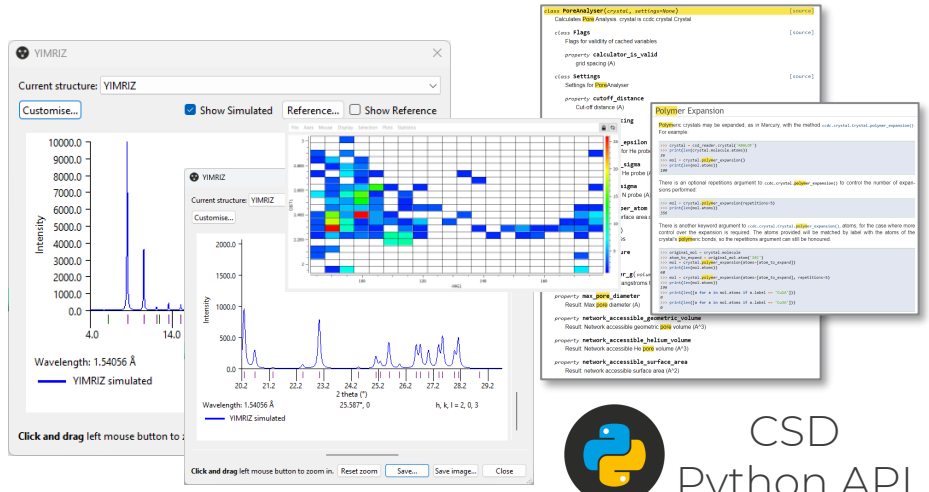
Structural Analysis



Void and pore analysis

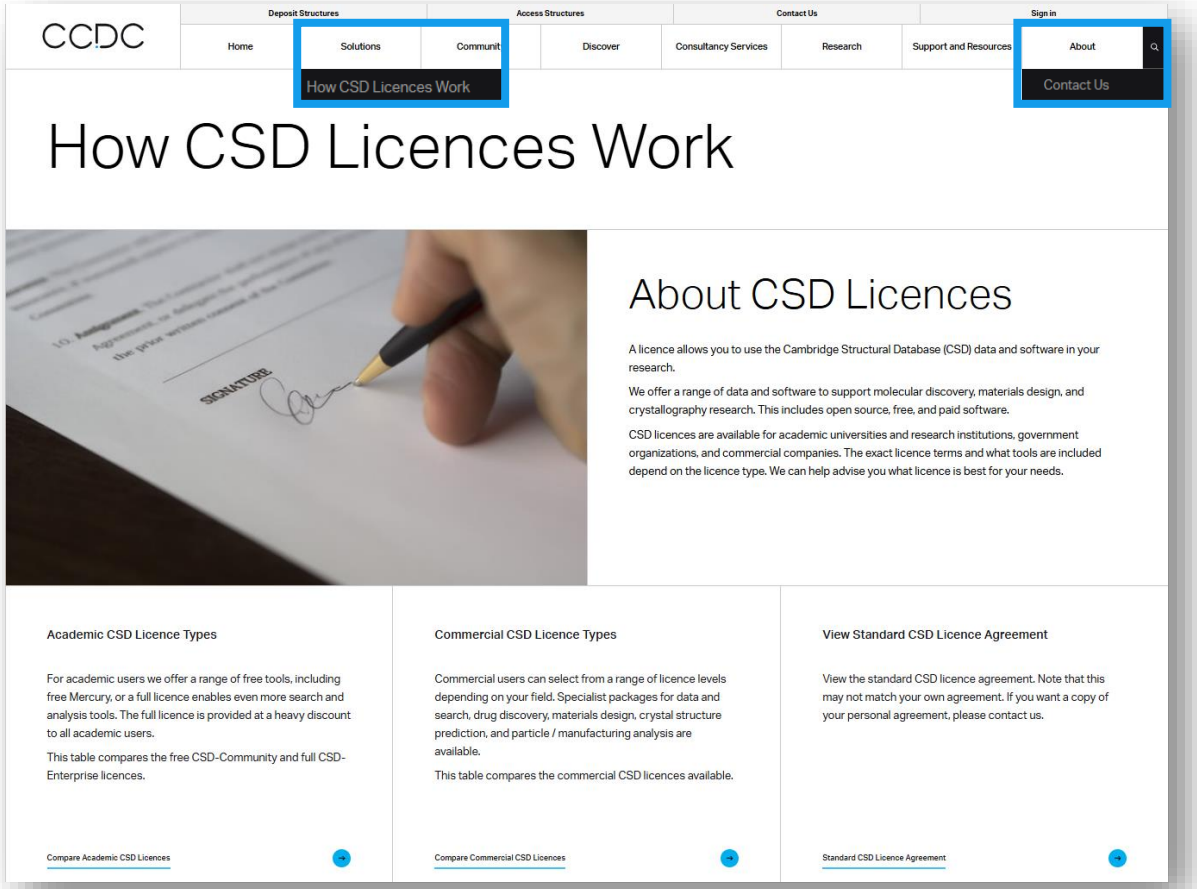
Mercury

Guest analysis



CSD Python API

What can you do with a licence?



- **Supporting research**: The CSD Python API enables bulk access to the CSD and large-scale structural surveys to advance your research.
- **Responsible data use:** We encourage and support use for research and analysis, but redistribution and large-scale publication are restricted to protect the CSD's sustainability.
- **Publication guidelines:** Small extracts and visualizations can be shared, but full datasets and CSD derived models require permission.
- **Supporting you:** More information is available on our website, but if in doubt, contact us—we're happy to help!

https://www.ccdc.cam.ac.uk/solutions/csd-licence/

support@ccdc.cam.ac.uk

CCDC

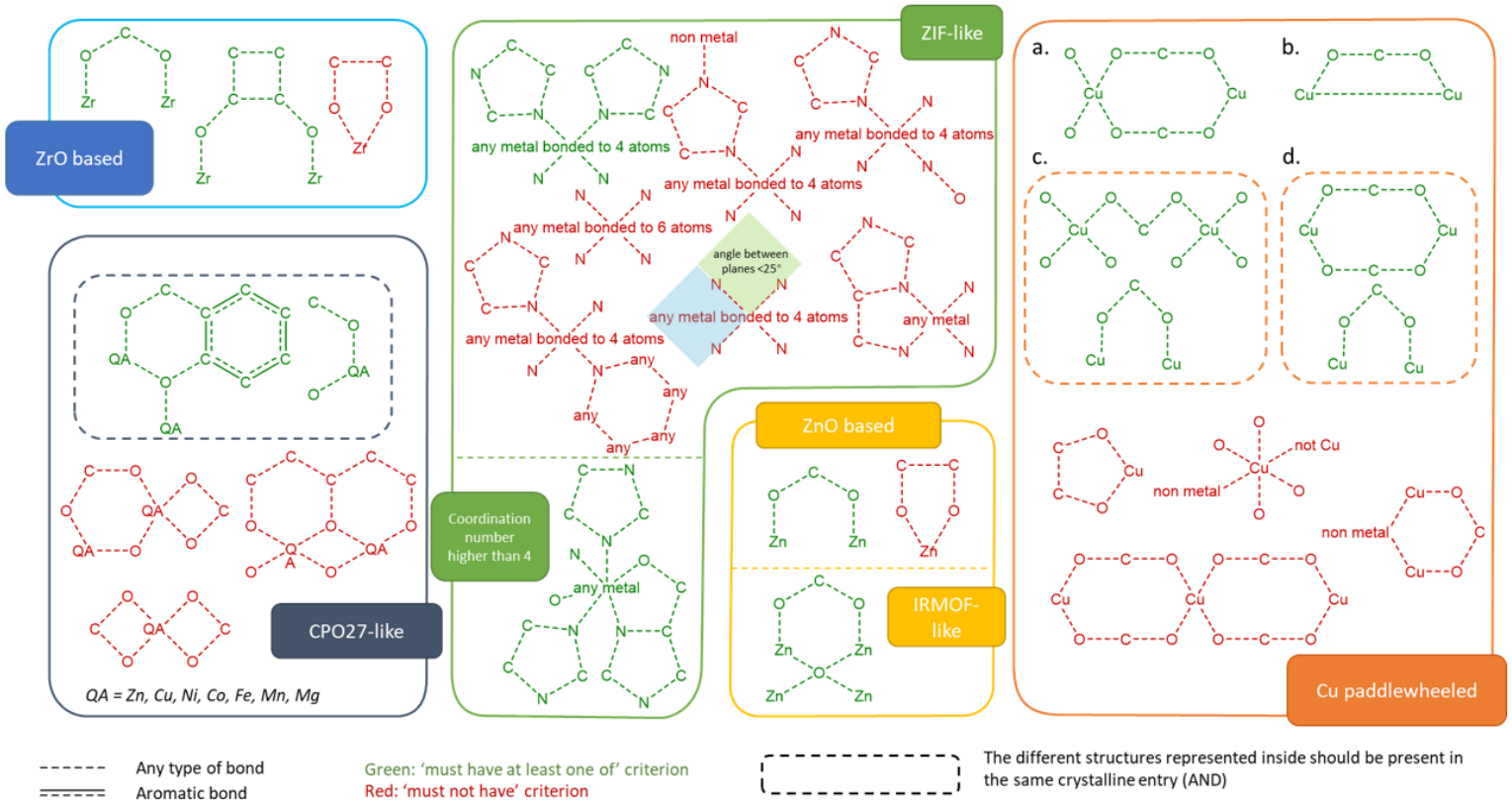
Agenda

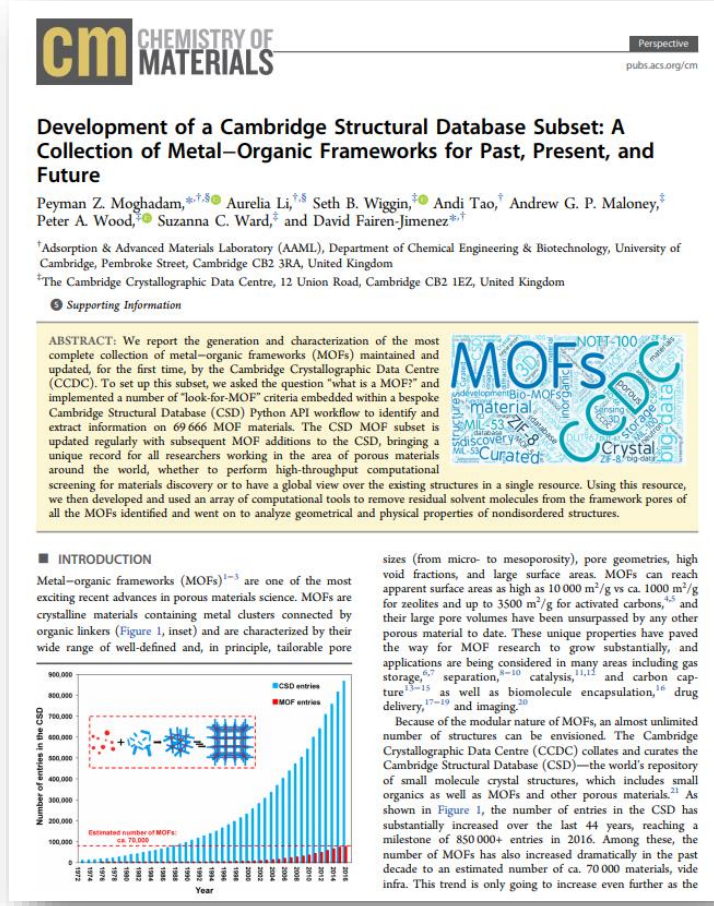
We are recording today's session and will make the recording available to you in the next few days

- *Show One:* Introduction to the CSD and porous materials in the database
- *Show One:* Case studies from Prof. Maria Celeste Bernini
- *Show One:* Mercury functionality for porous materials
- *Try One:* Hands-on exercises
- *Explore More: M*ore tips and tricks for porous materials and case studies
- *Explore More:* Quiz and summary
- *Extra time:* More time for hands on and Q&A

CCDC

Complex structures and searches





cm CHEMISTRY OF MATERIALS

Perspective

pubs.acs.org/cm

Development of a Cambridge Structural Database Subset: A Collection of Metal–Organic Frameworks for Past, Present, and Future

Peyman Z. Moghadam,*[†,§] Aurelia Li,[†,§] Seth B. Wiggin,[‡] Andi Tao,[†] Andrew G. P. Maloney,[‡] Peter A. Wood,[‡] Suzanna C. Ward,[‡] and David Fairen-Jimenez*[,†]

[†]Adsorption & Advanced Materials Laboratory (AAML), Department of Chemical Engineering & Biotechnology, University of Cambridge, Pembroke Street, Cambridge CB2 3RA, United Kingdom

[‡]The Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, United Kingdom

Supporting Information

ABSTRACT: We report the generation and characterization of the most complete collection of metal–organic frameworks (MOFs) maintained and updated, for the first time, by the Cambridge Crystallographic Data Centre (CCDC). To set up this subset, we asked the question "what is a MOF?" and implemented a number of "look-for-MOF" criteria embedded within a bespoke Cambridge Structural Database (CSD) Python API workflow to identify and extract information on 69 666 MOF materials. The CSD MOF subset is updated regularly with subsequent MOF additions to the CSD, bringing a unique record for all researchers working in the area of porous materials around the world, whether to perform high-throughput computational screening for materials discovery or to have a global view over the existing structures in a single resource. Using this resource, we then developed and used an array of computational tools to remove residual solvent molecules from the framework pores of all the MOFs identified and went on to analyze geometrical and physical properties of nondisordered structures.

INTRODUCTION

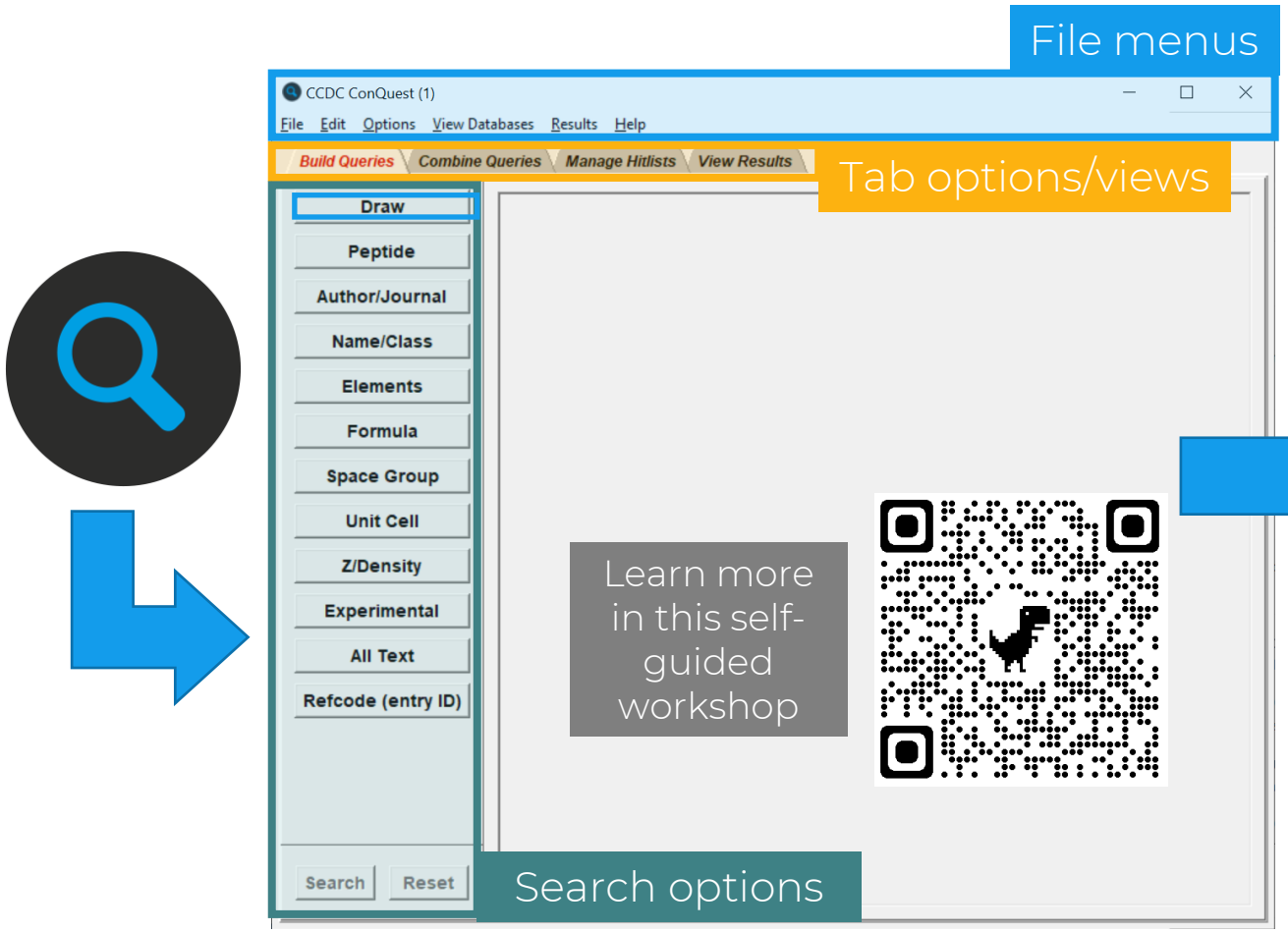
Metal–organic frameworks (MOFs)[1−5] are one of the most exciting recent advances in porous materials science. MOFs are crystalline materials containing metal clusters connected by organic linkers (Figure 1, inset) and are characterized by their wide range of well-defined and, in principle, tailorable pore sizes (from micro- to mesoporosity), pore geometries, high void fractions, and large surface areas. MOFs can reach apparent surface areas as high as 10 000 m^2/g vs ca. 1000 m^2/g for zeolites and up to 3500 m^2/g for activated carbons,[6,7] and their large pore volumes have been unsurpassed by any other porous material to date. These unique properties have paved the way for MOF research to grow substantially, and applications are being considered in many areas including gas storage,[6,7] separation,[8−10] catalysis,[11,12] and carbon capture[13−15] as well as biomolecule encapsulation,[16] drug delivery,[17−19] and imaging.[20]

Because of the modular nature of MOFs, an almost unlimited number of structures can be envisioned. The Cambridge Crystallographic Data Centre (CCDC) collates and curates the Cambridge Structural Database (CSD)—the world's repository of small molecule crystal structures, which includes small organics as well as MOFs and other porous materials.[21] As shown in Figure 1, the number of entries in the CSD has substantially increased over the last 44 years, reaching a milestone of 850 000+ entries in 2016. Among these, the number of MOFs has also increased dramatically in the past decade to an estimated number of ca. 70 000 materials, vide infra. This trend is only going to increase even further as the

P. Z. Moghadam, A. Li, S. B. Wiggin, A. Tao, A. G. P. Maloney, P. A. Wood, S. C. Ward, D. Fairen-Jimenez, *Chem. Mater.*, 2017, 29 2618-2625, DOI: 10.1021/acs.chemmater.7b00441

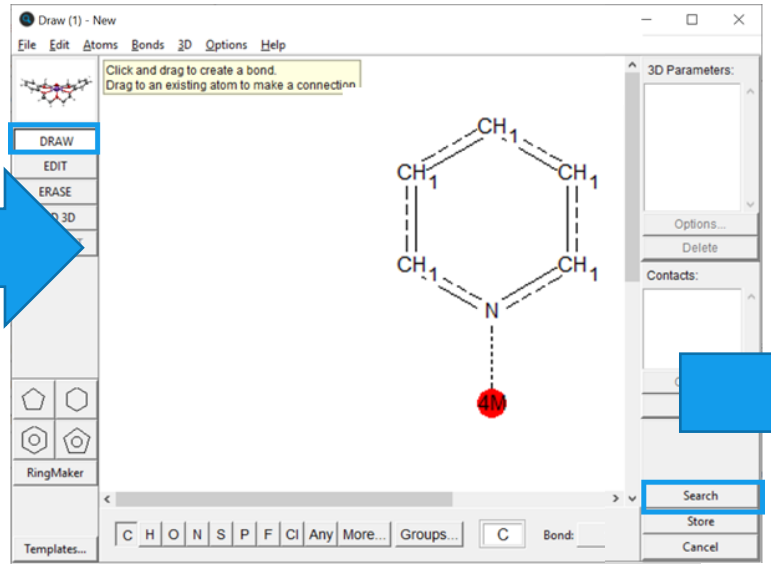
CCDC

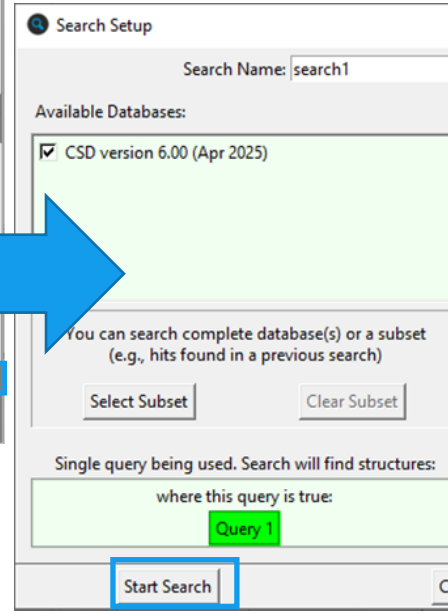
How to search in ConQuest



1. Build a Query: what do you want to find?

2. Click search filters: do you restrict your s





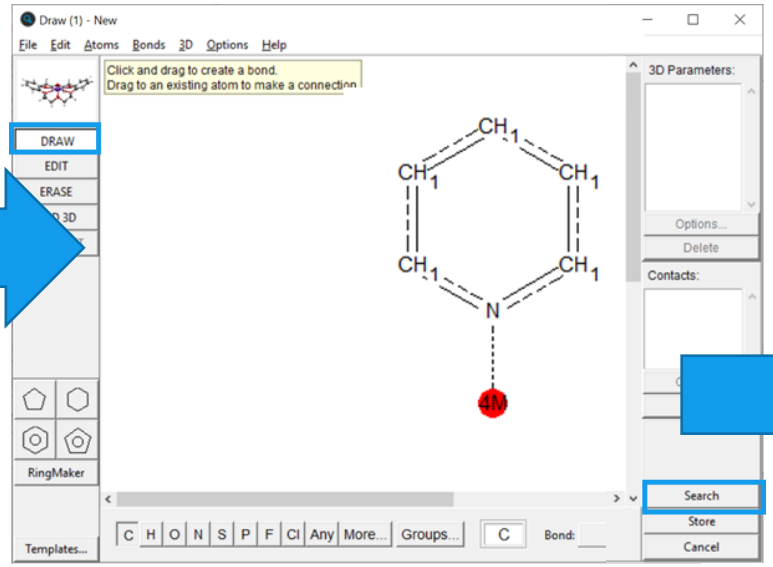
CCDC

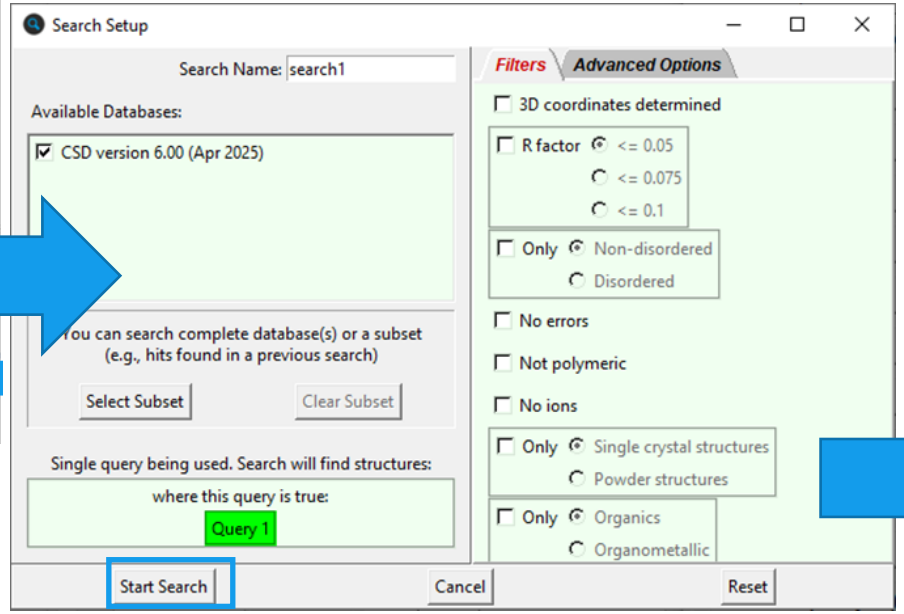
How to search in ConQuest

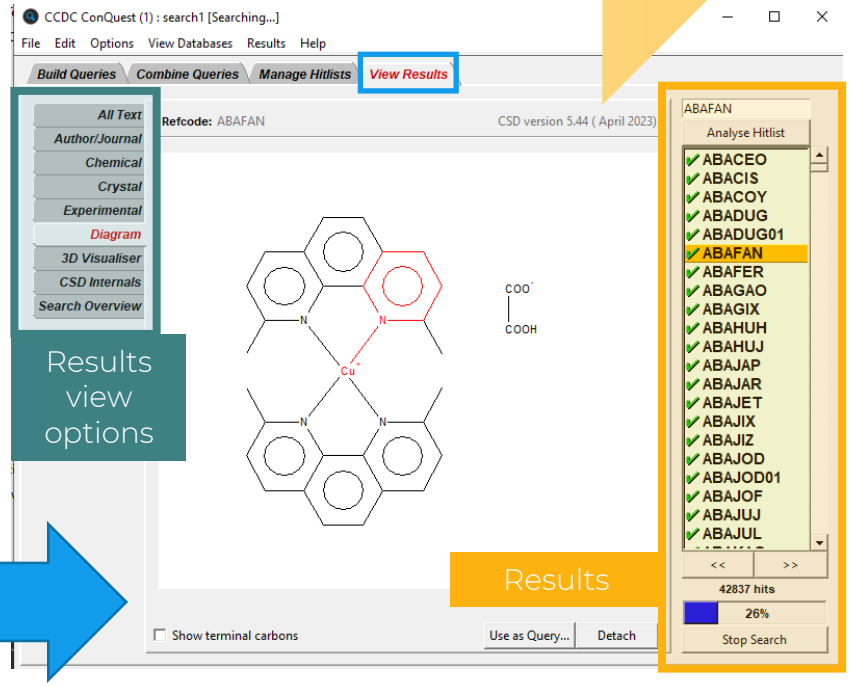
1. Build a Query: what do you want to find?

2. Click search and select filters: do you need to restrict your search?

3. Visualise and analyse results: what can you learn from this data?

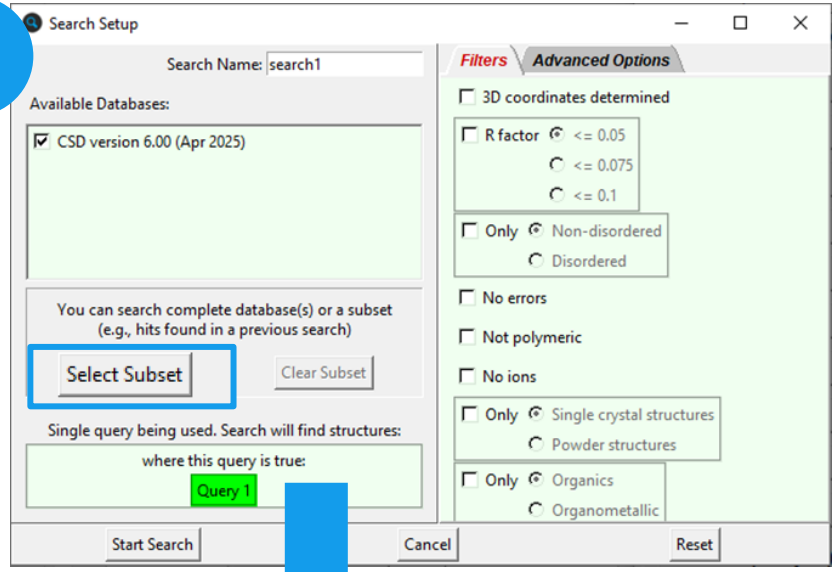




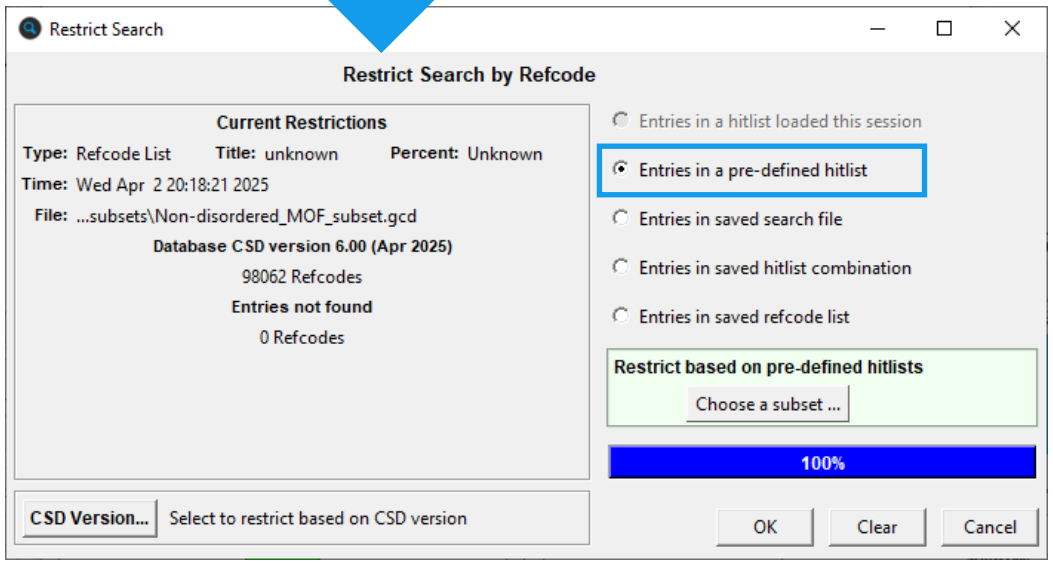


Using the subsets in ConQuest

1

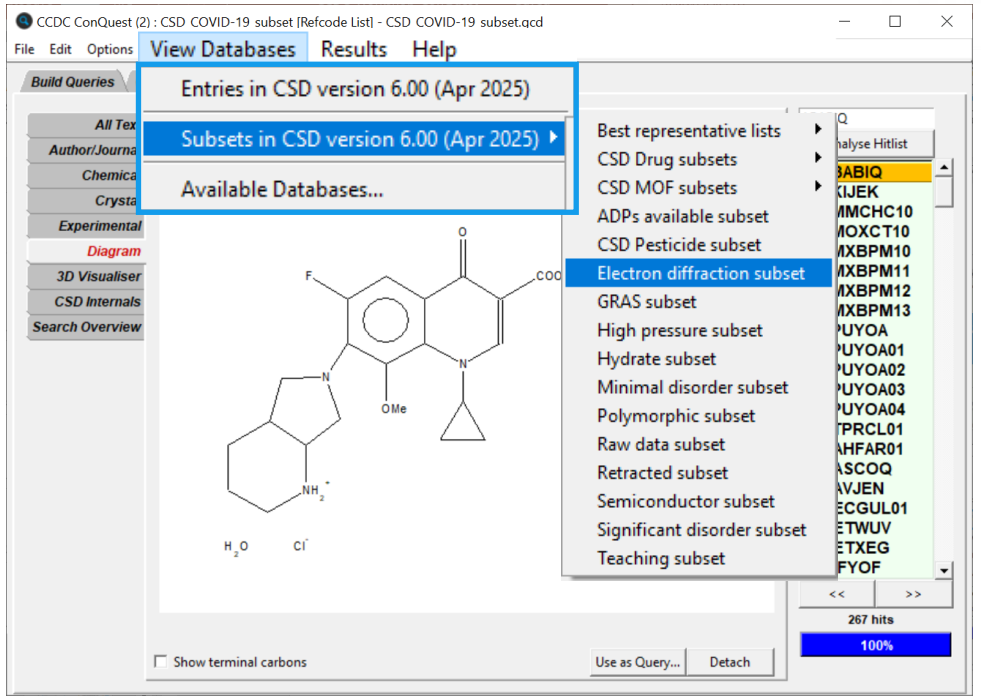


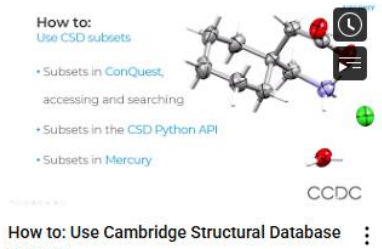
Restrict a search to a subset



2

Load a subset in the hitlists





NEW!
Watch on YouTube

CCDC

Using the subsets in the CSD Python API

Subsets

class ccdc.io.**Subsets** [source]

This class provides a simple way to access pre-defined CSD subsets.

Example:

```
>>> mof_reader = EntryReader(subset=Subsets.MOF)
```

The returned *reader* object is the same as if the *Reader* class has been initialized with the associated GCD file directly.

Subsets available:

- ADP
- BEST_HYDROGENS
- BEST_LOW_TEMP
- BEST_RFACTOR
- BEST_ROOM_TEMP
- COVID19
- DRUG
- DRUG_SINGLE_COMPONENT
- ELECTRON
- HIGH_PRESSURE
- HYDRATE
- MOF
- MOF_NO_DISORDER
- MOF_1D
- MOF_2D
- MOF_3D
- PESTICIDE
- POLYMORPHIC
- TEACHING
- MINIMAL_DISORDER
- SIGNIFICANT_DISORDER

CSD Python API documentation and cookbook

Entry examples

Create indexes of useful information for subsets of CSD entries

Note that this script makes use of functionality from the cookbook utility module.

```
#!/usr/bin/env python
#
# This script can be used for any purpose without limitation subject to the
# conditions at http://www.ccdc.cam.ac.uk/Community/Pages/Licences/v2.aspx
#
# This permission notice and the following statement of attribution must be
# included in all copies or substantial portions of this script.
#
# 2015-06-17: created by the Cambridge Crystallographic Data Centre
#

'''
Provide information on a set of structures in the CSD.

This script takes as input a gcd file (a text file with CSD refcodes) and
writes out the identifier, author(s), literature reference, formula, compound
name and compound synonym(s). The output can be formatted as csv or html.

'''
from __future__ import division, absolute_import, print_function
import six
import sys
import os
import csv
import html
import argparse
import codecs

from ccdc.io import EntryReader


class Writer(object):
    def __init__(self, infile, out, format='csv'):
        try:
            self.rdr = EntryReader(infile, format='identifiers')
        except RuntimeError:
            print('Failed to read input file %s!' % infile)
            exit(1)

        self.out = out
        getattr(self, format + '_header')()
        for e in self.rdr:
            getattr(self, format + '_line')(e)
        getattr(self, format + '_footer')()

    def csv_header(self):
        data = ','.join([
            'Identifier',
            'Author',
            'Literature Ref',
            'Formula',
            'Compound Name',
            'Synonym'
```

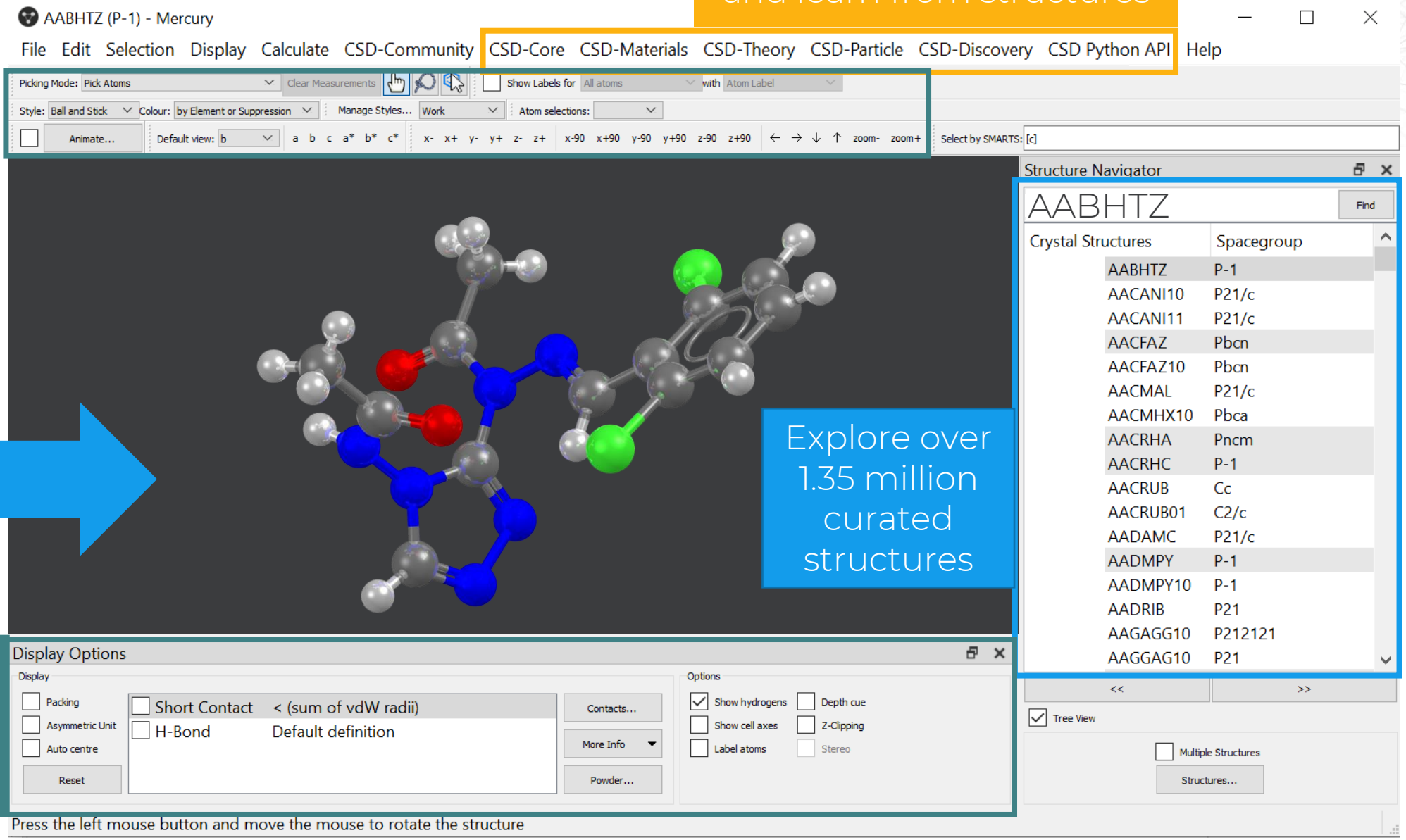
Learn about searching with the CSD Python API in this free CSDU module

CCDC

Mercury Overview

More advanced functionality to analyse and learn from structures

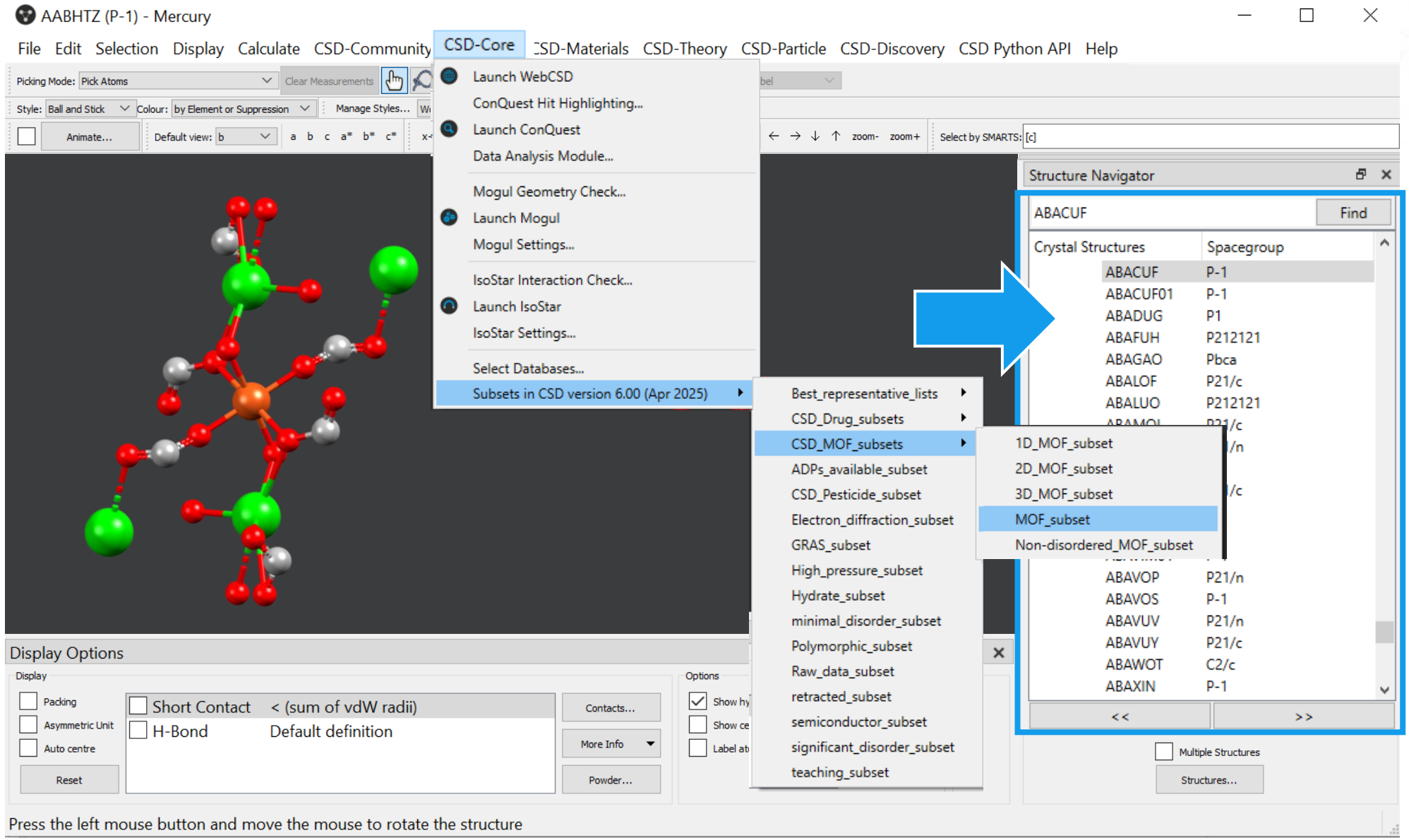
Learn on demand with the CSDU module!



Explore over 1.35 million curated structures

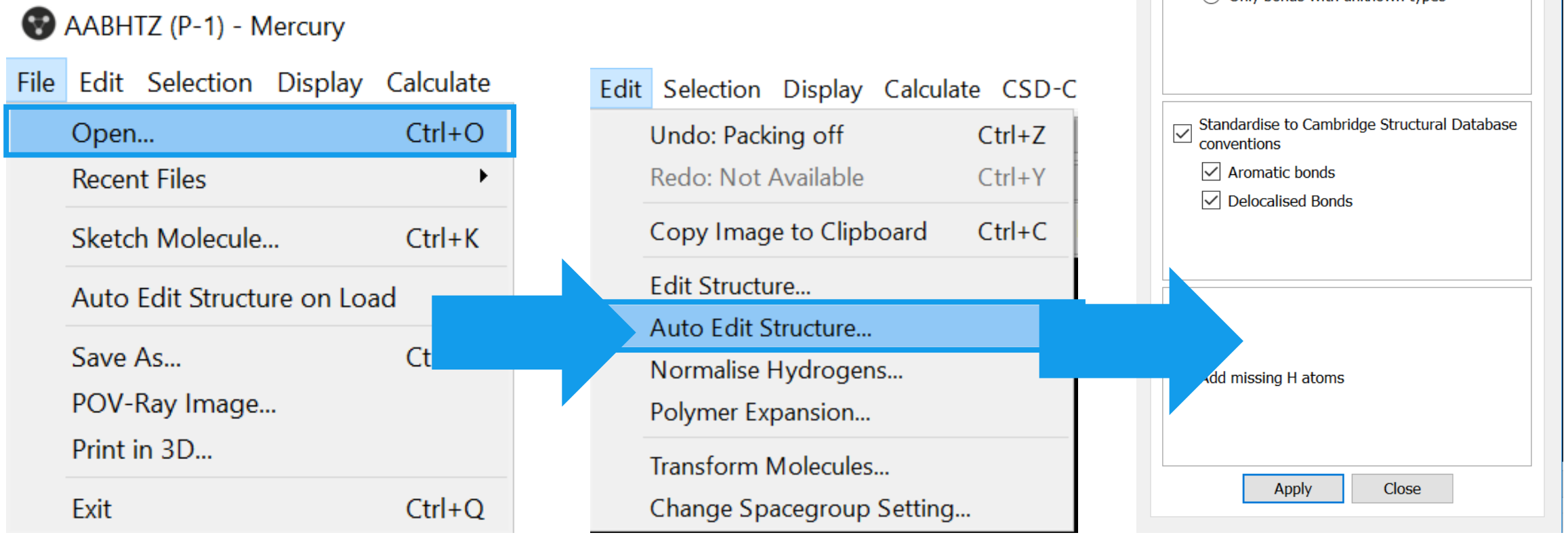
Display options to visualise and navigate structures

Using the subsets in Mercury

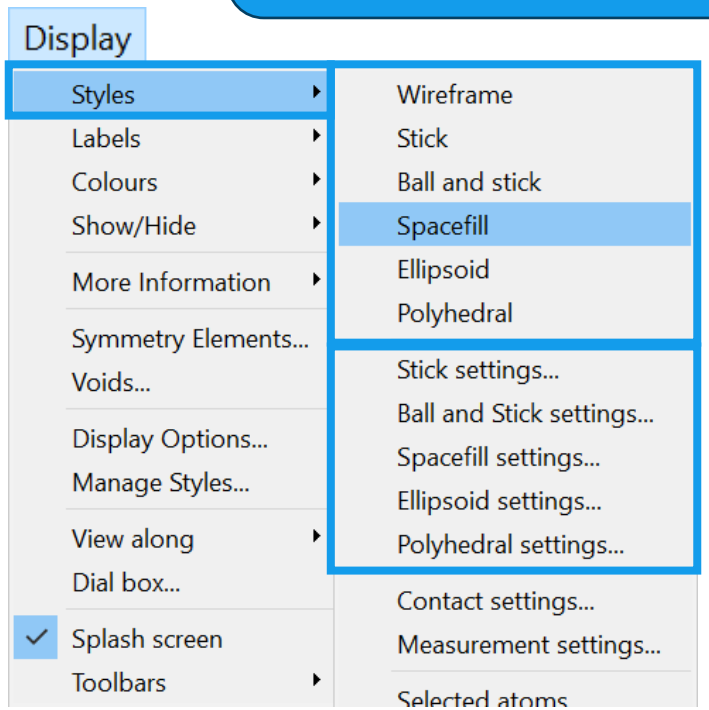


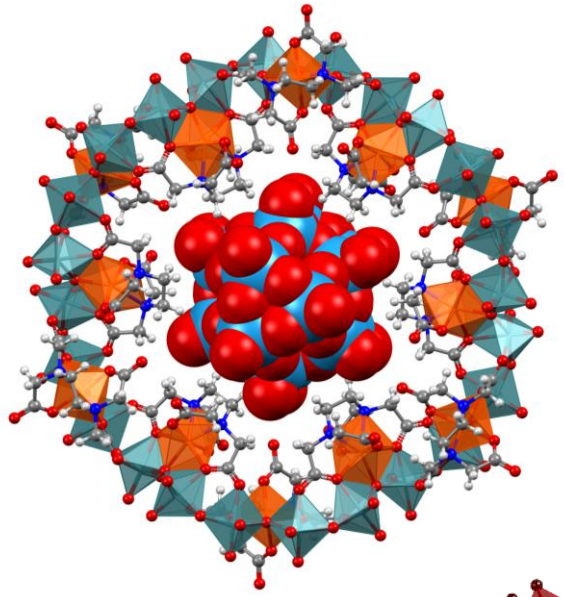
Using your own file

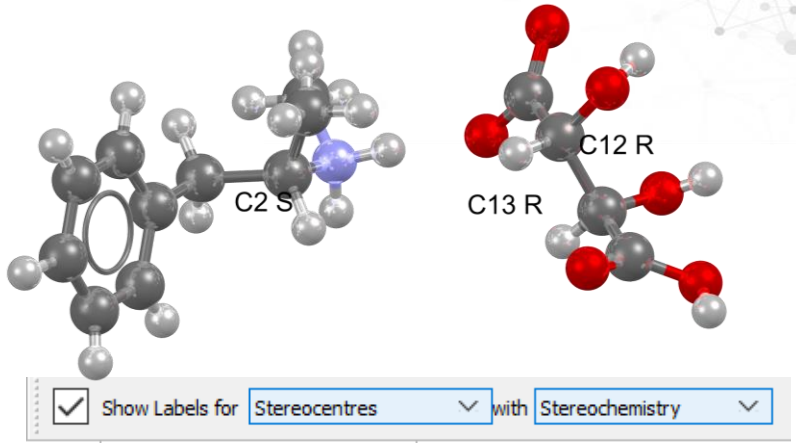
- File > Open – to open one of your own files for example a CIF
- Edit > Auto Edit Structure to assign bond types
- File > Auto Edit Structure on Load to automate

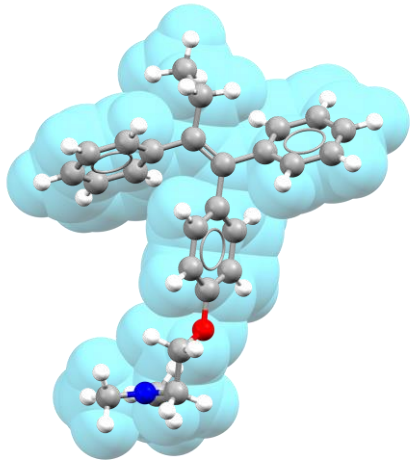


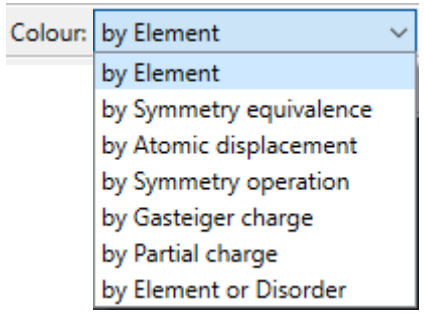
Lots of display and analysis options!

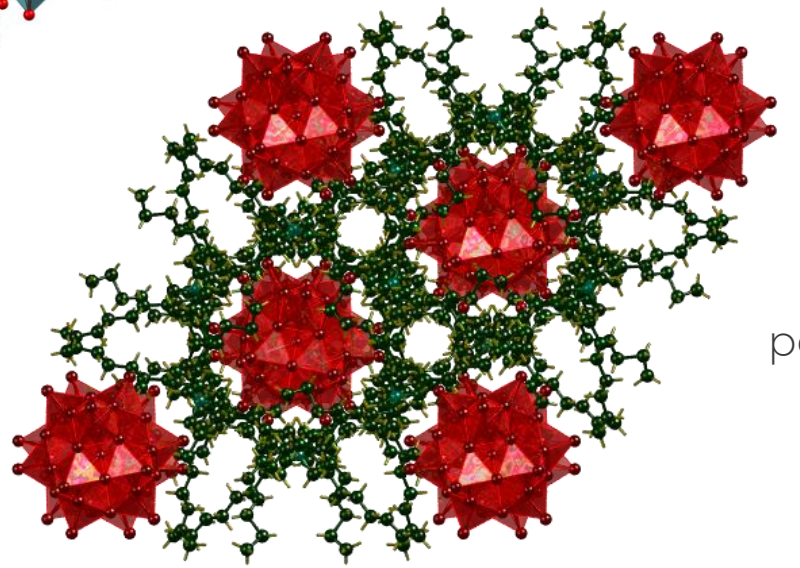


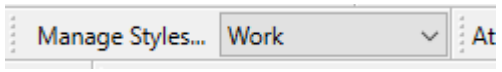








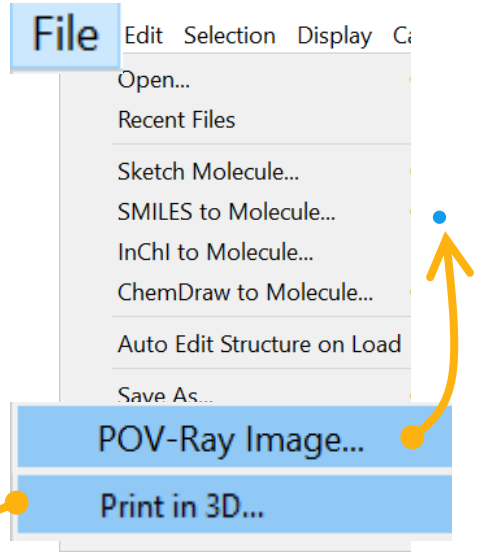




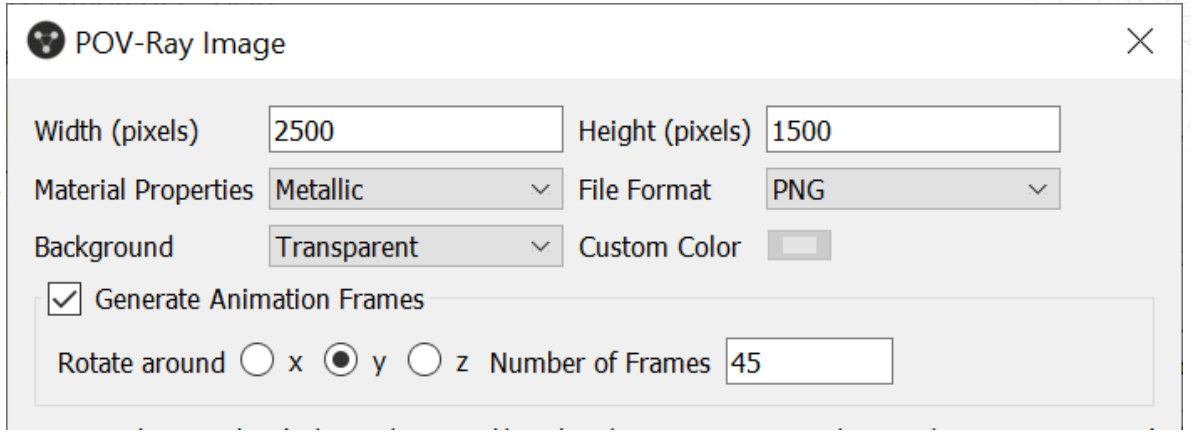
Watch the video on creating personalised styles

CCDC

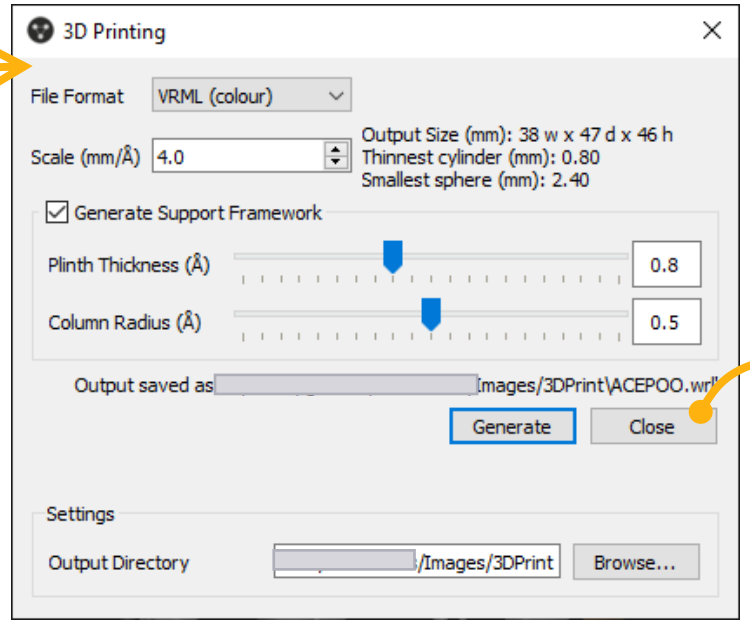
Features for effective visualization

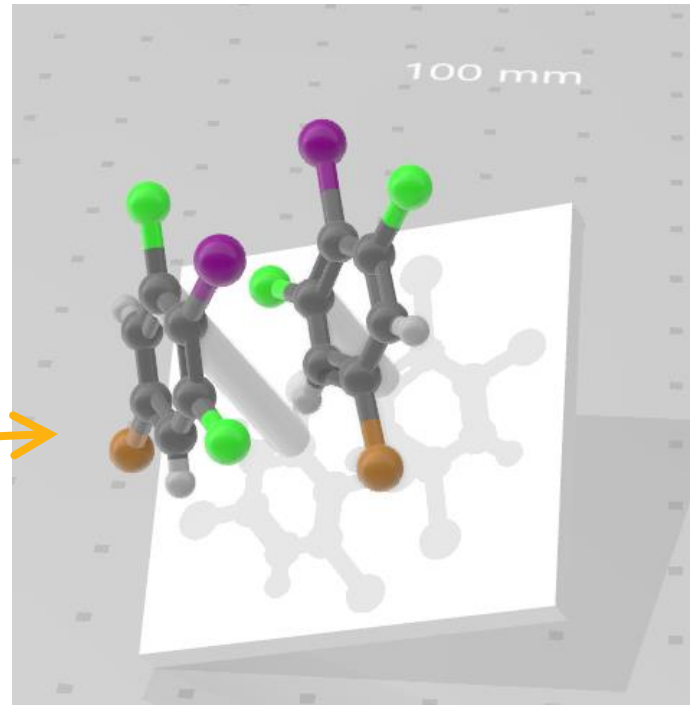


- Create high resolution images and frames for video or gifs

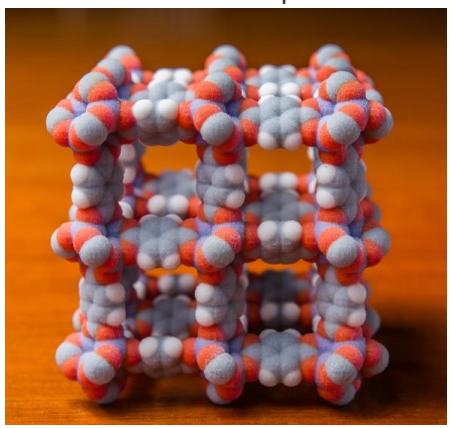


- Create files for 3D print

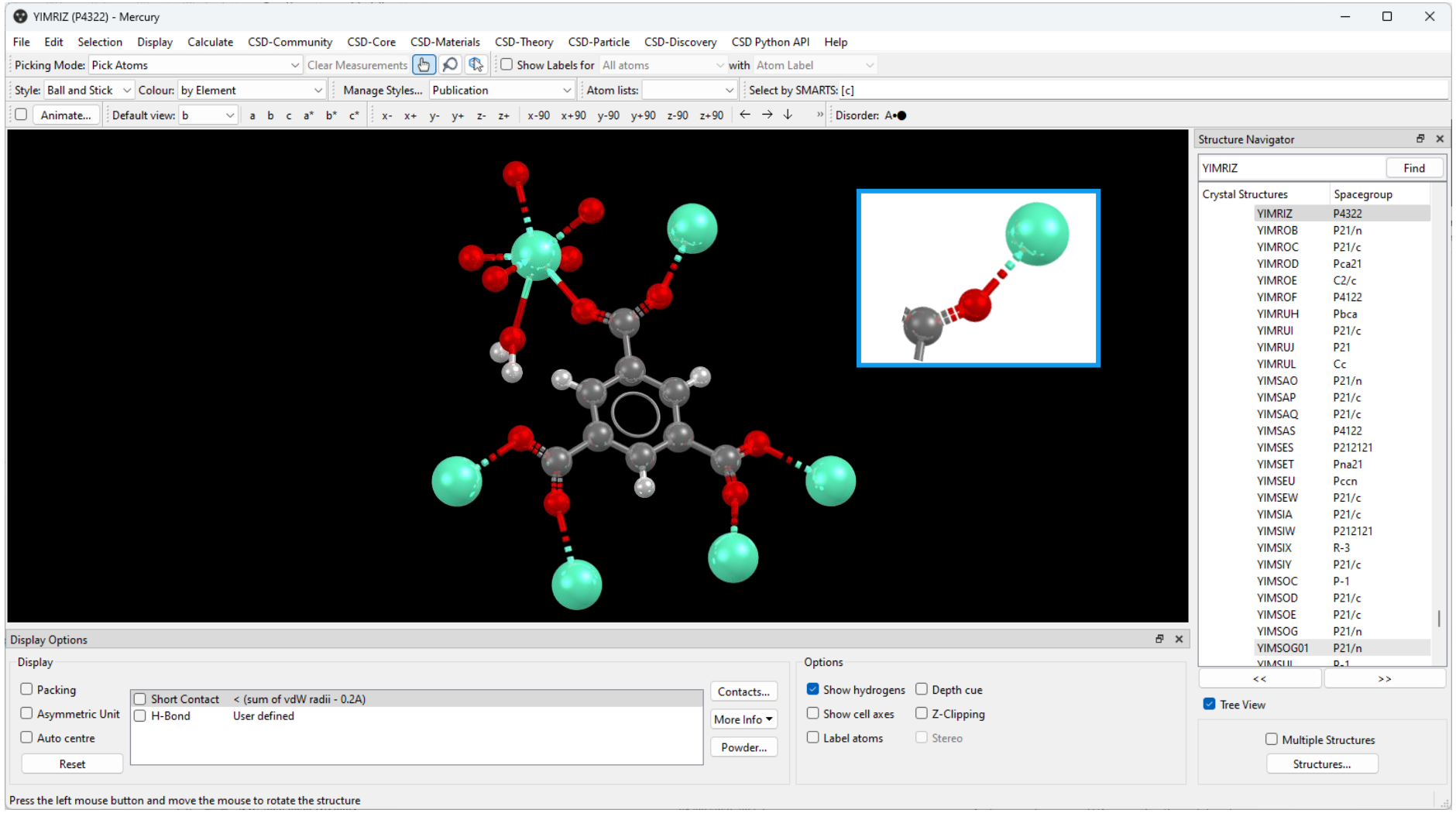




Example of 3D printed structure of CSD Refcode SAHYIK



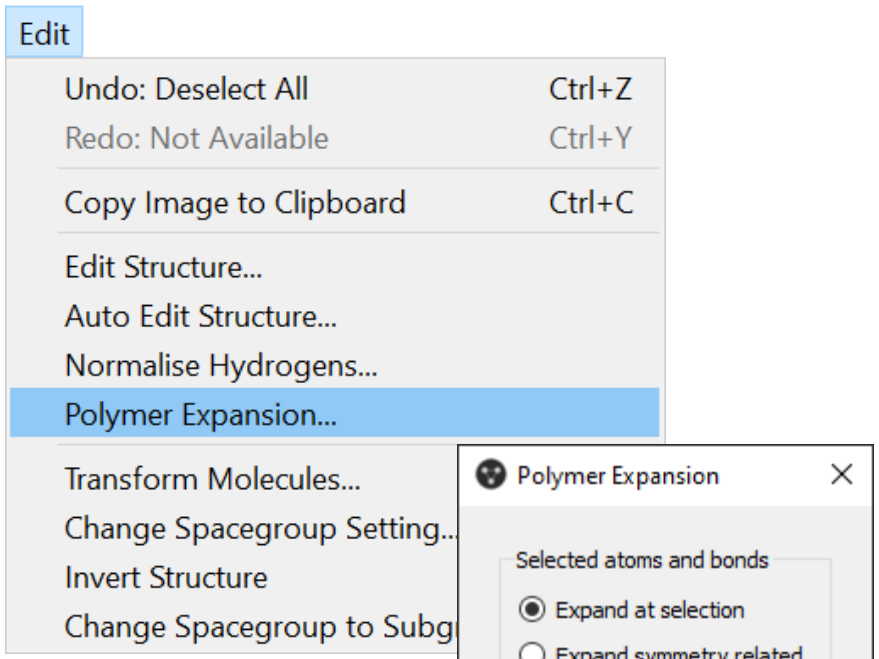
Polymeric bond representations

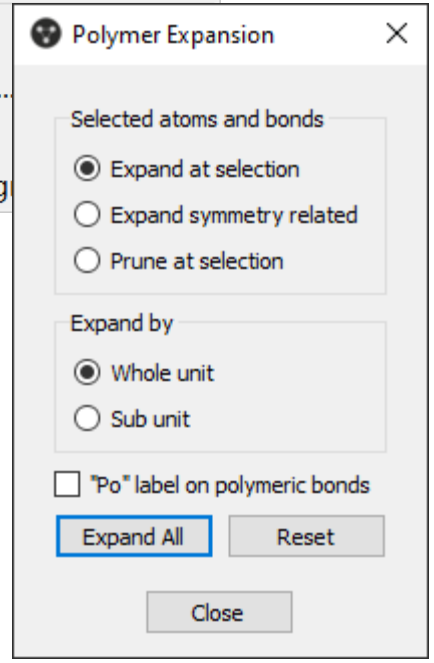


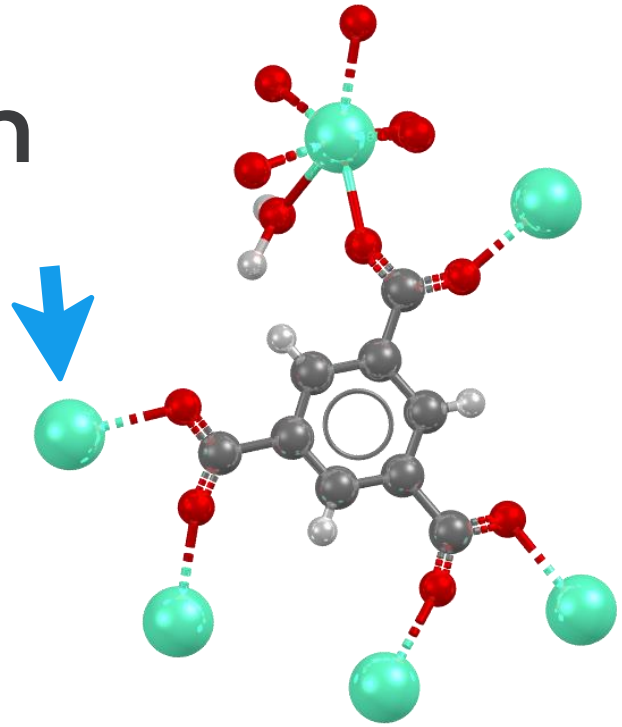
In Mercury, polymeric bonds are drawn alternating long and short lines.

CCDC

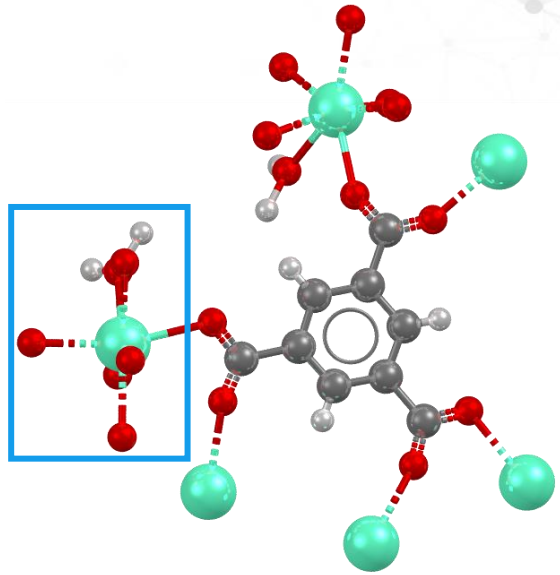
Polymer Expansion

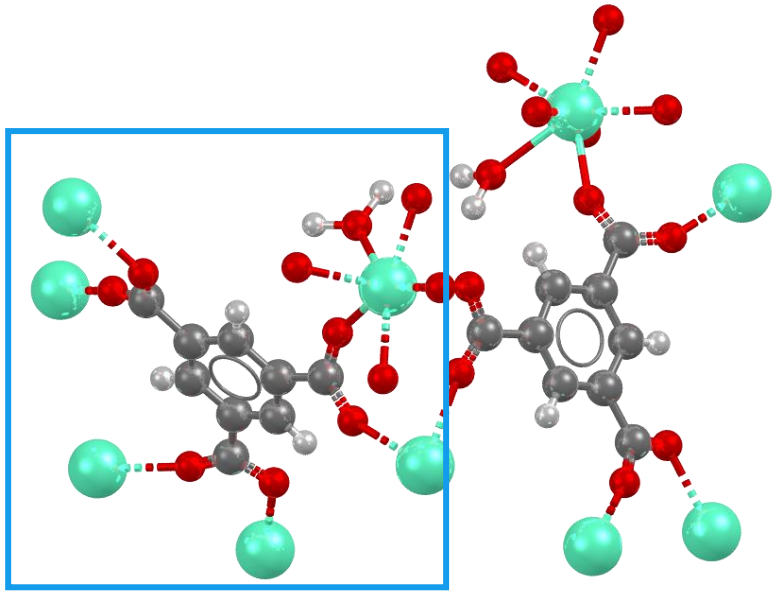






By **Sub unit**
it will add atoms and bonds up to the next polymeric atom or bond in the structure.

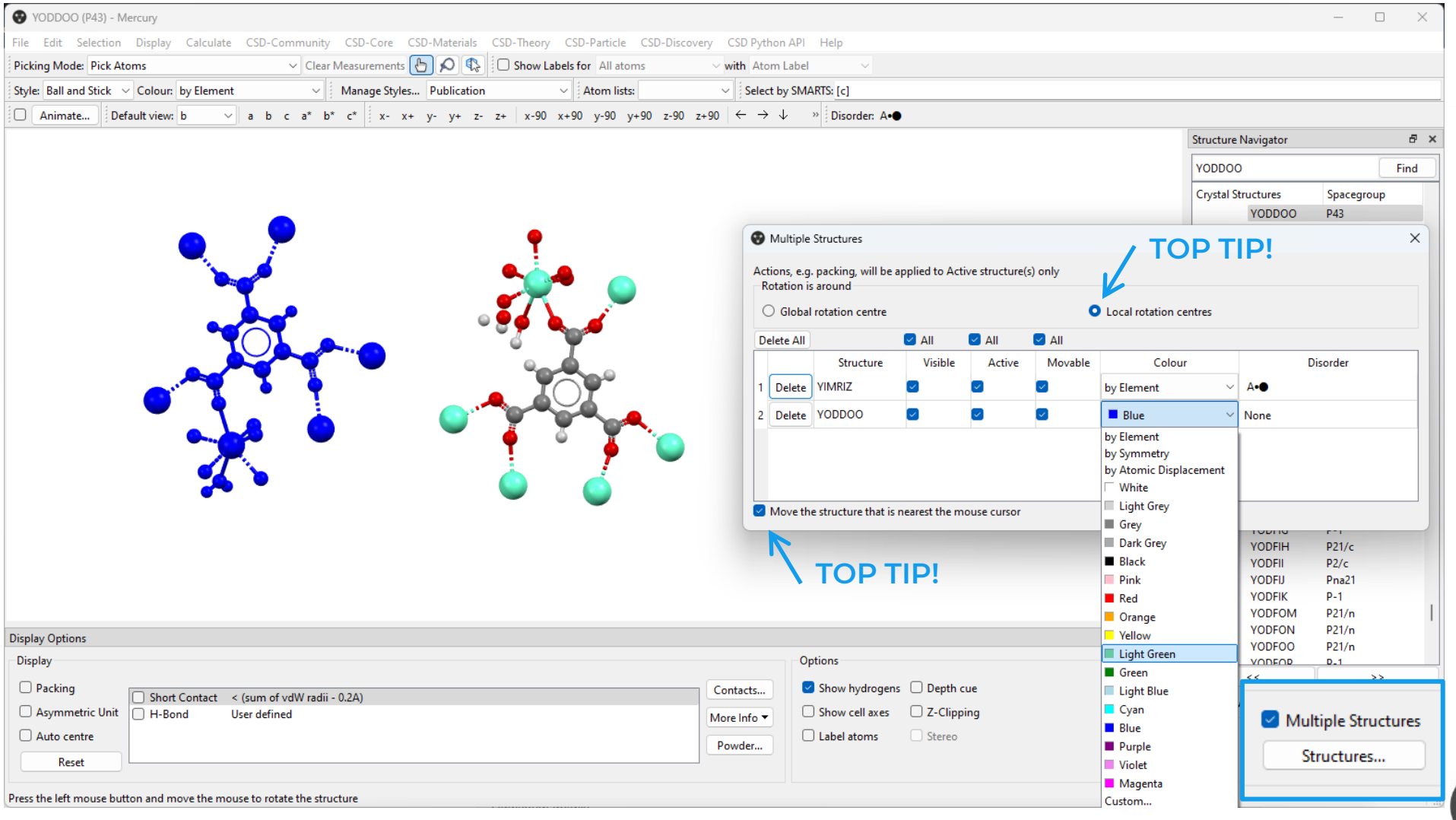




By **Whole unit**
It will add an additional repeat of the initial crystal chemical unit to the structure

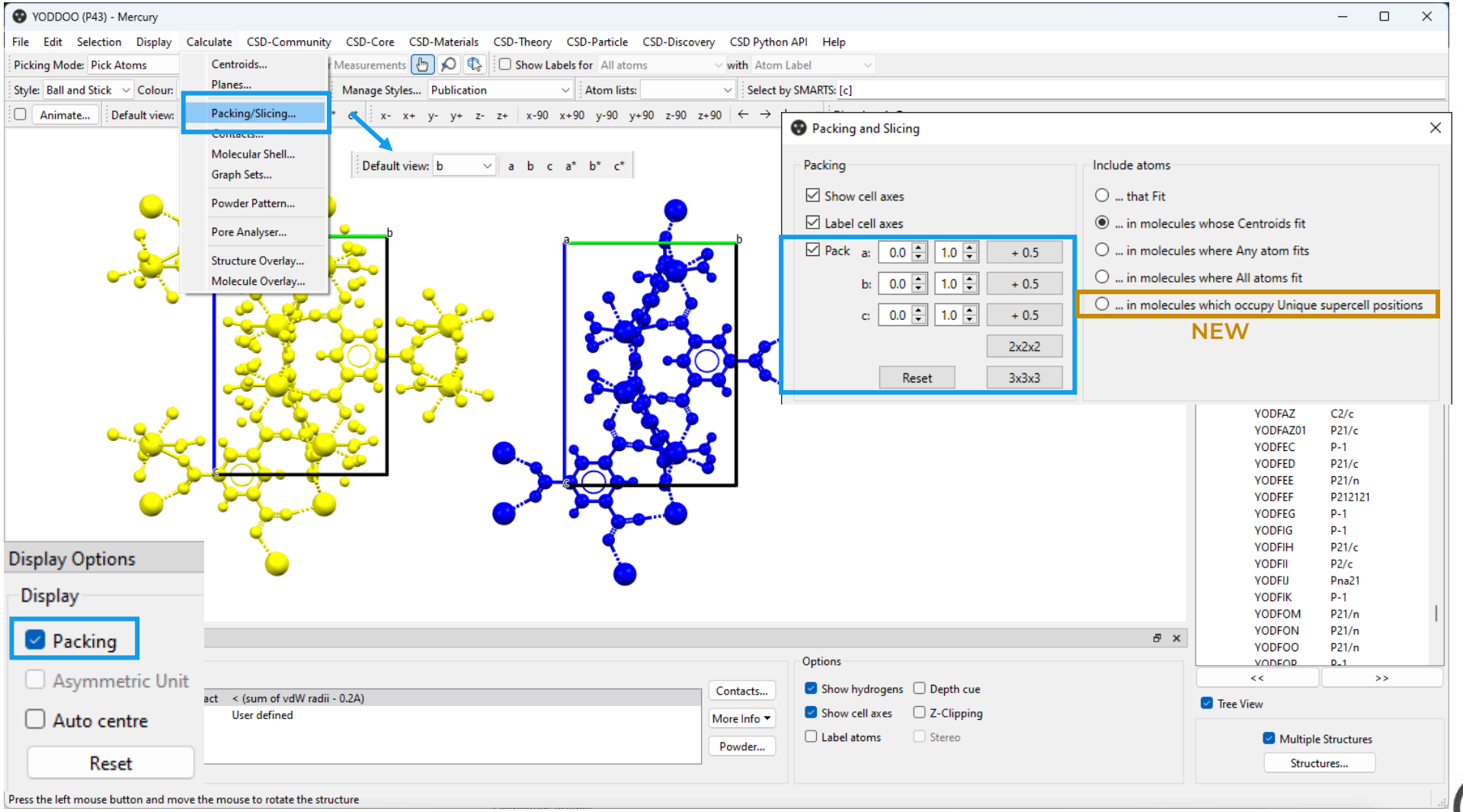
CCDC

Multiple Structures



Use Multiple Structures for structure comparisons. Structures can be in different colours – helpful when overlaying structures.

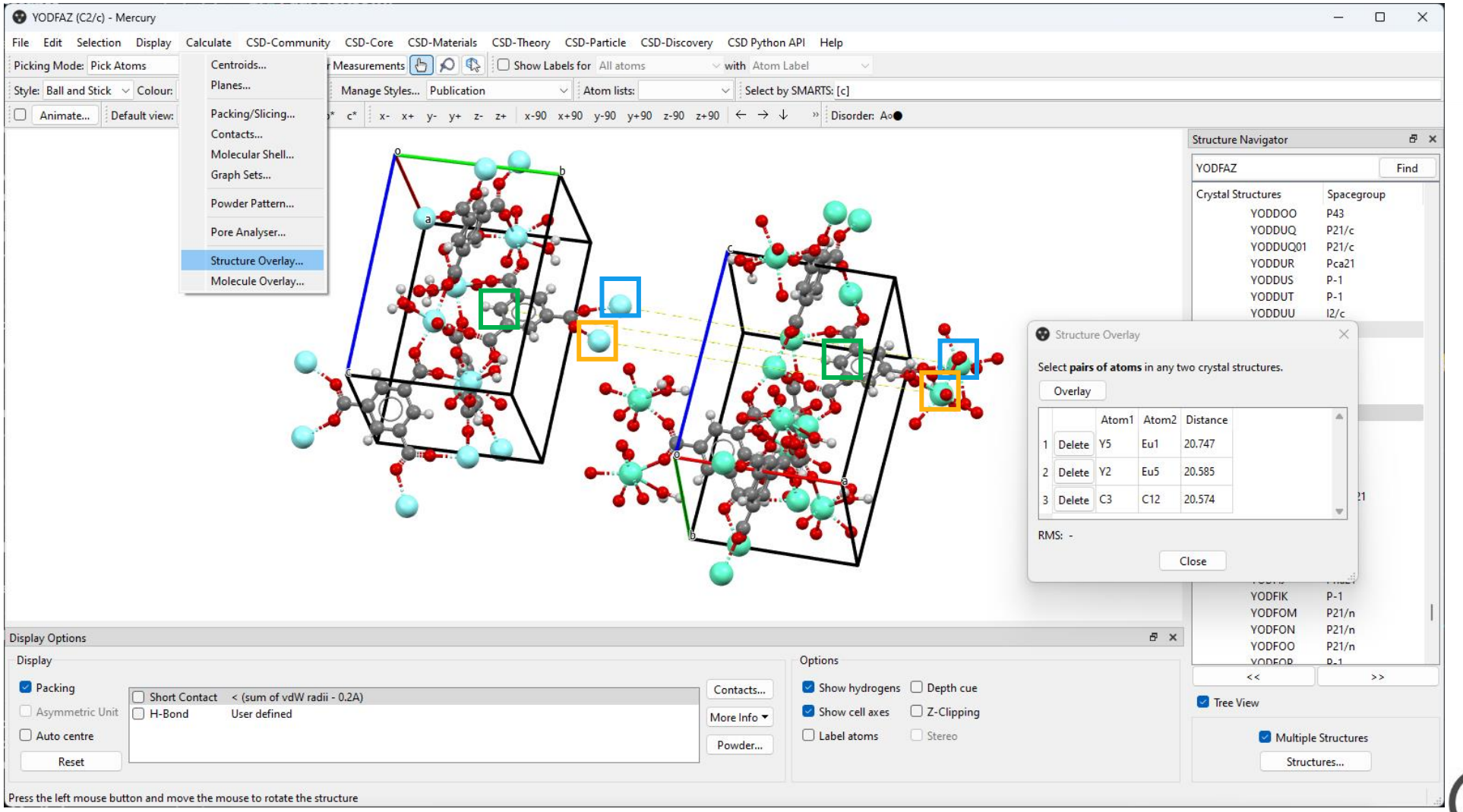
Crystal Packing



Use Packing to compare larger structural units – helpful if polymer repeat unit is not identical.

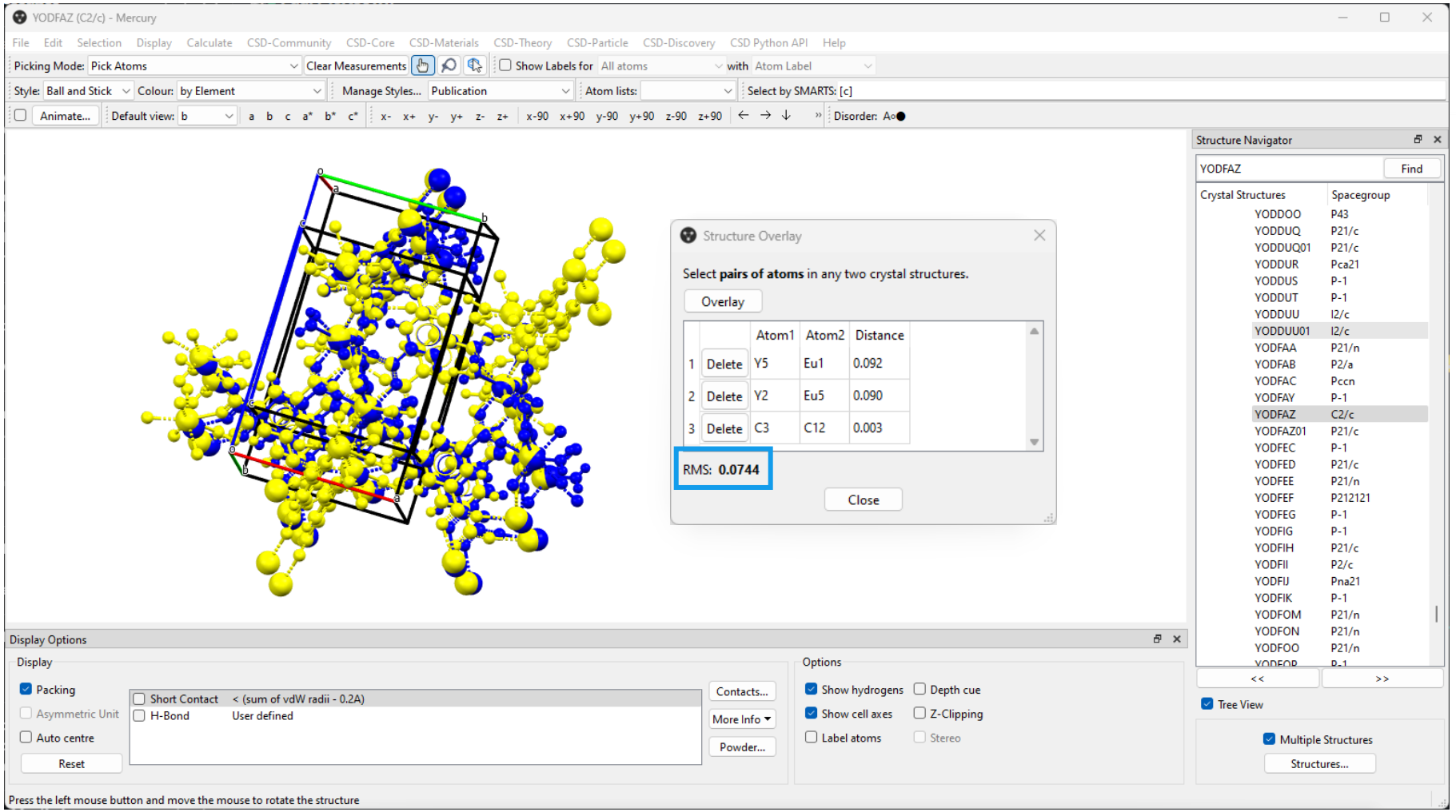
Use Default view for(initial) alignment.

Structure Overlay



Select pairs of atoms in the same structural environment. The more pairs you use, the more realistic the RMS value will be.

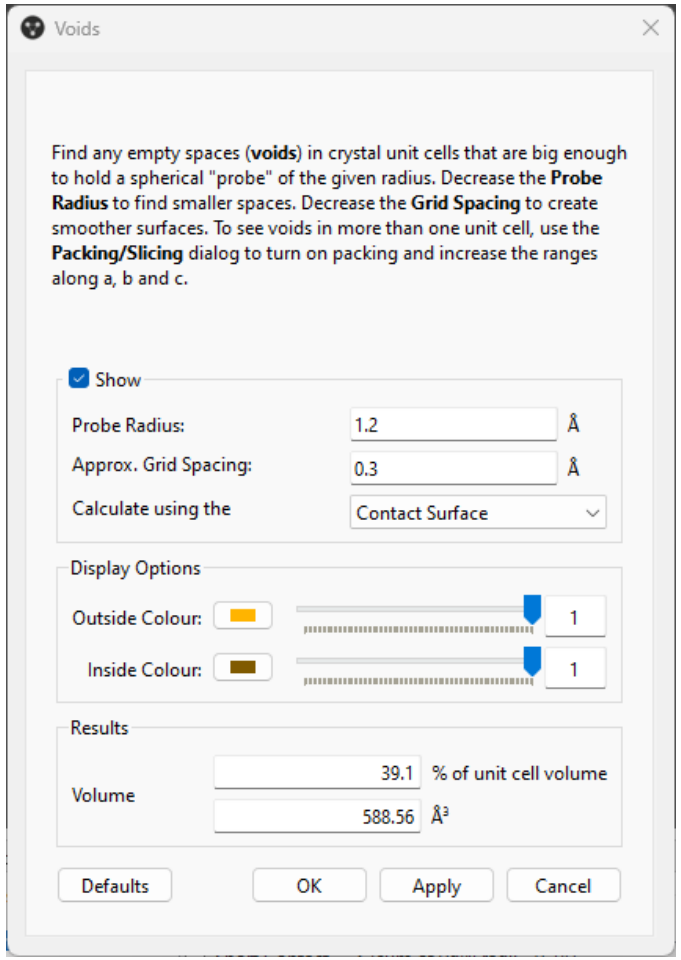
Structure Overlay

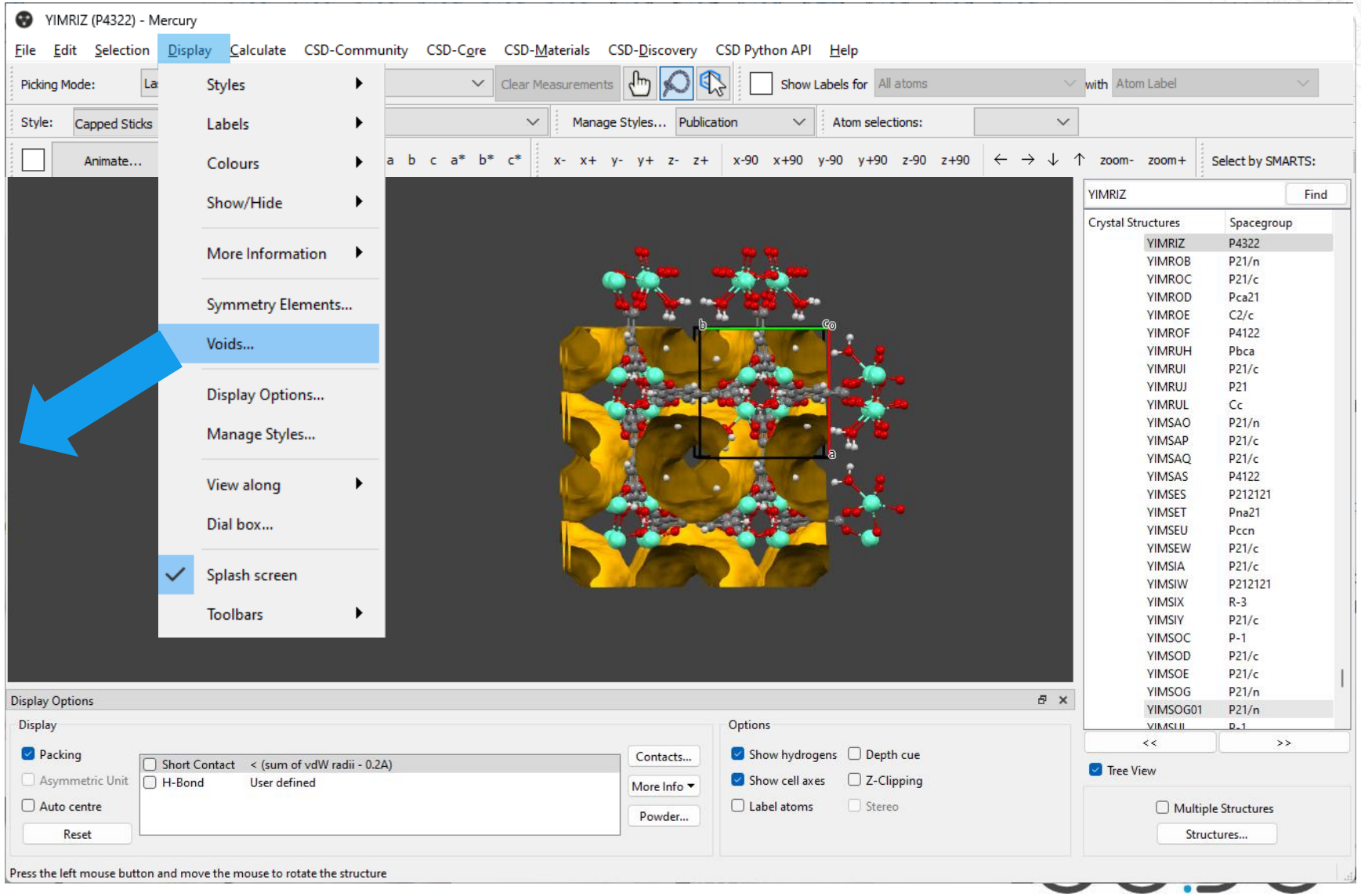


Select pairs of atoms in the same structural environment. The more pairs you use, the more realistic the RMS value will be.

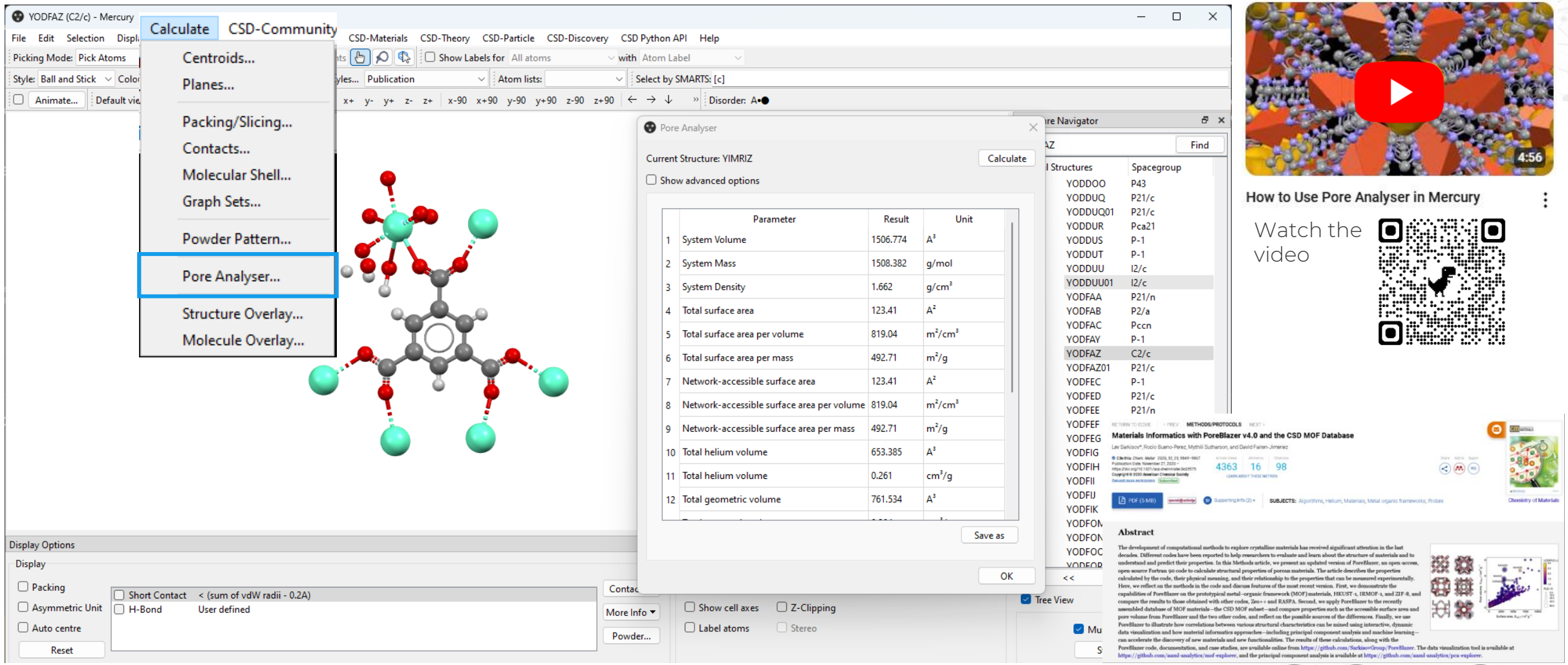
CCDC

Exploring voids





Pore Analyser



How to Use Pore Analyser in Mercury

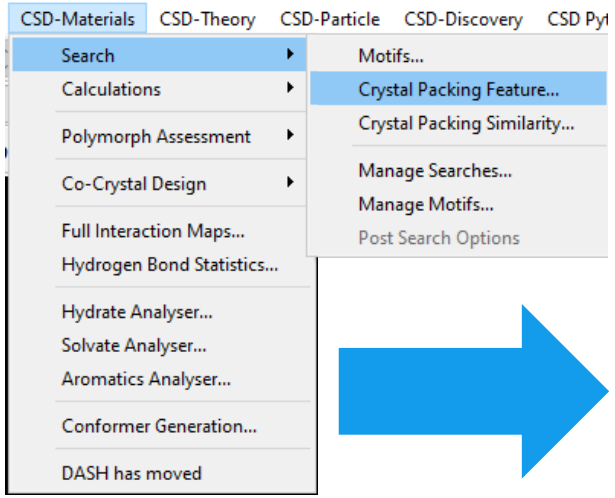
Watch the video

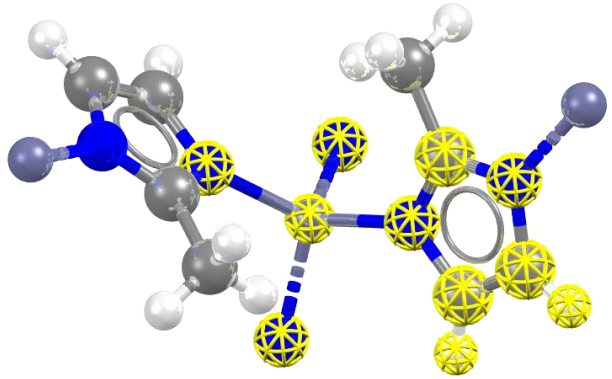
Lev Sarkisov*, Rocio Bueno-Perez, Mythili Sutharson, and David Fairen-Jimenez, Chem. Mater. 2020, 32, 23, 9849–9867

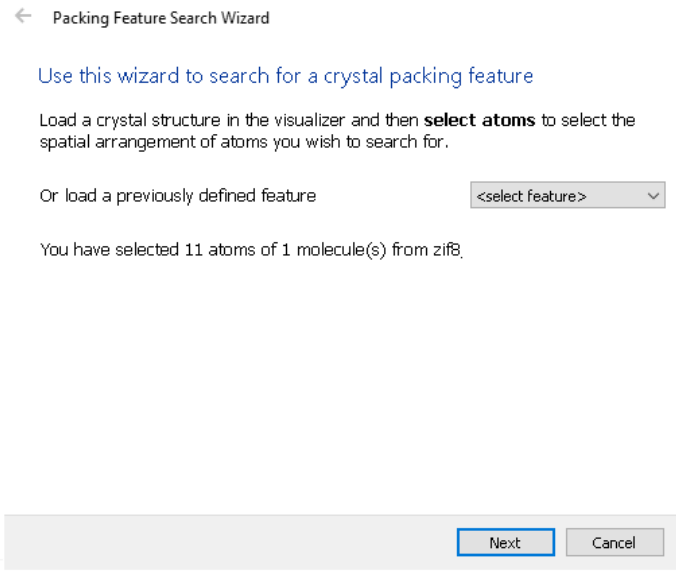
CCDC

Crystal Packing Feature Search

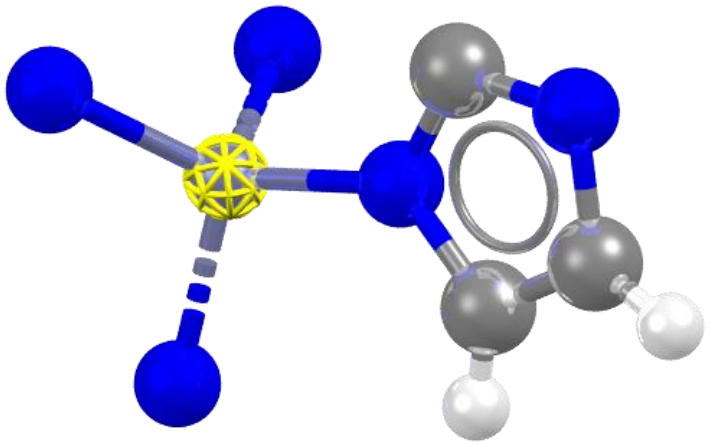
- Perform a substructure search.
- Investigate conformations of molecules or bonded fragments.
- Search for non-covalent interactions such as π-π or hydrogen bond interactions.
- Search for particular spatial arrangements of functional groups.
- Search for particular spatial arrangements of molecules.

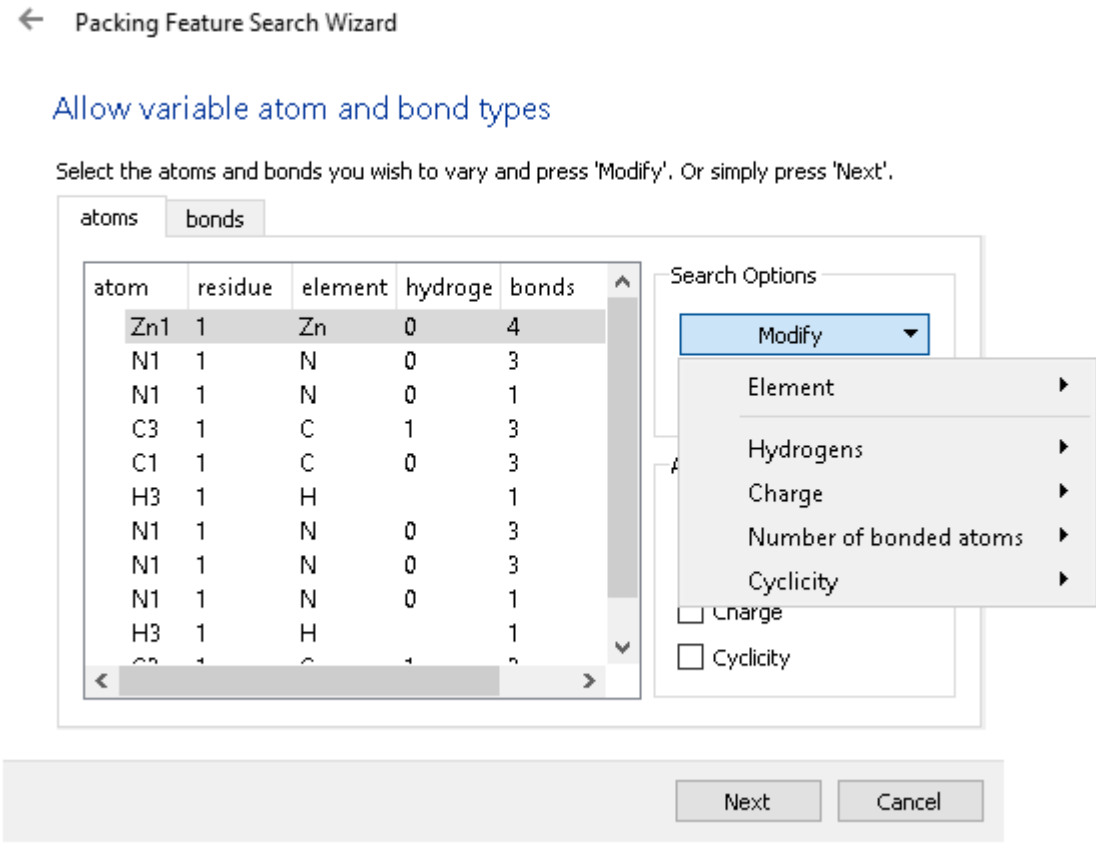




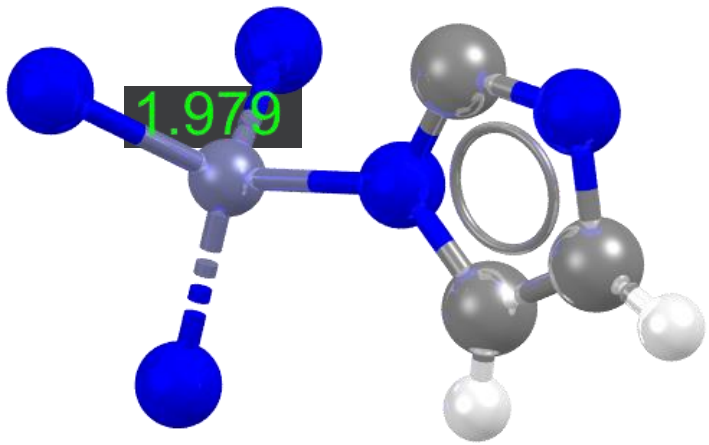


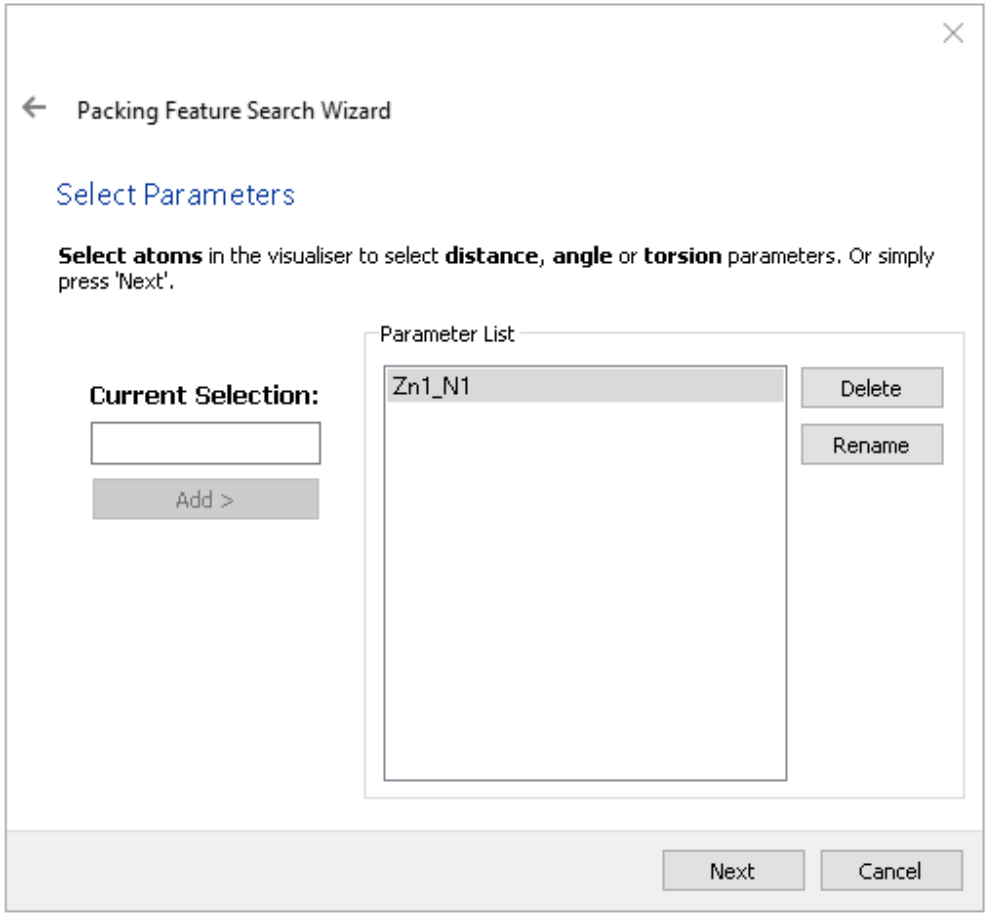
Additional search functionality



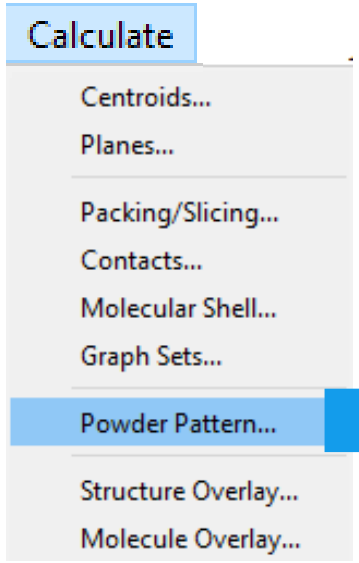


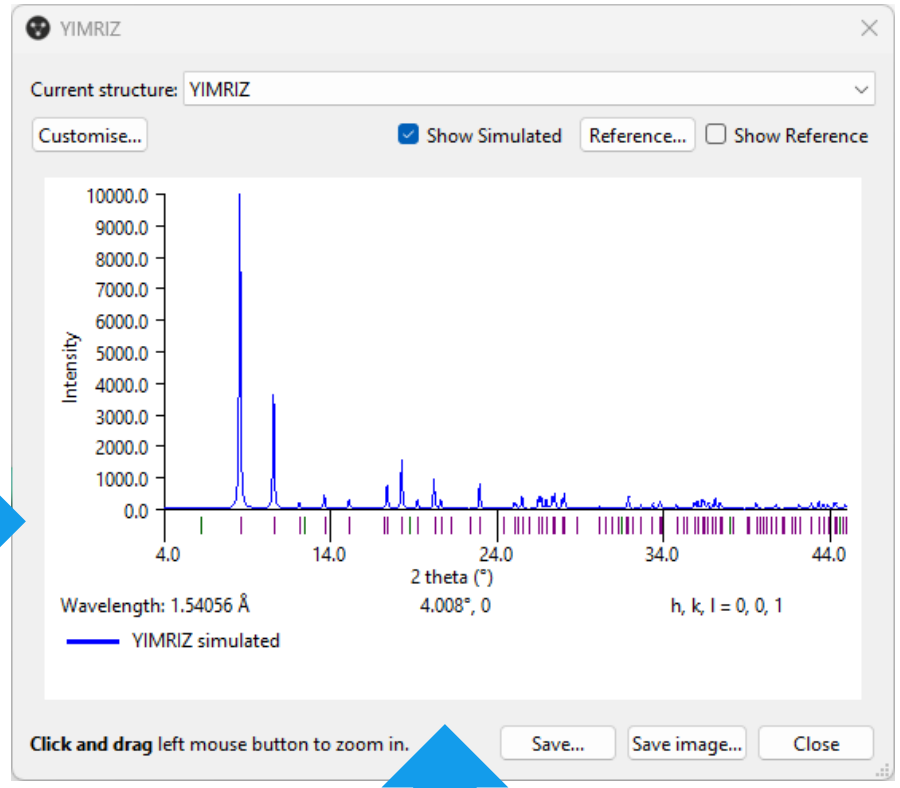
Use for structure analysis





Calculate powder pattern

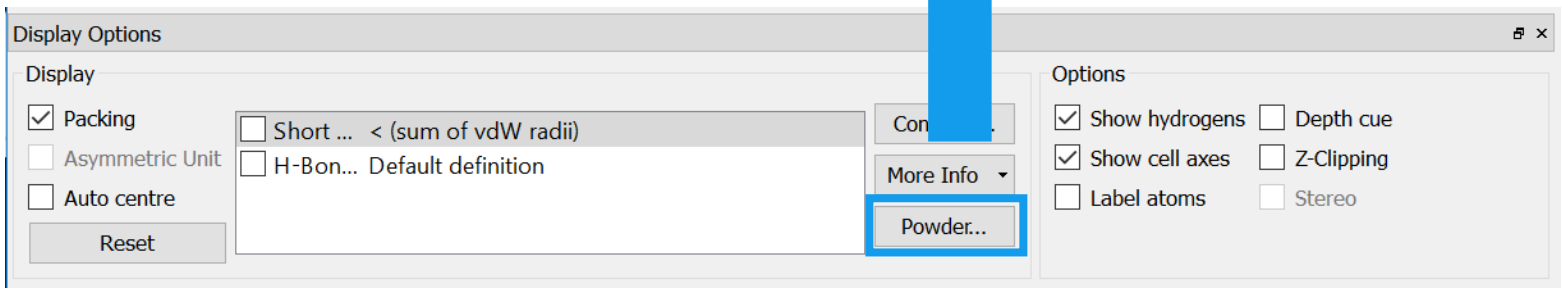




How are the patterns calculated?

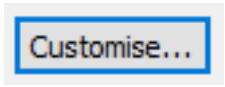
- Fundamental calculation
- No one algorithm exists
- Series of steps need to be taken based on the theories of how X-rays interact with 3D lattices.

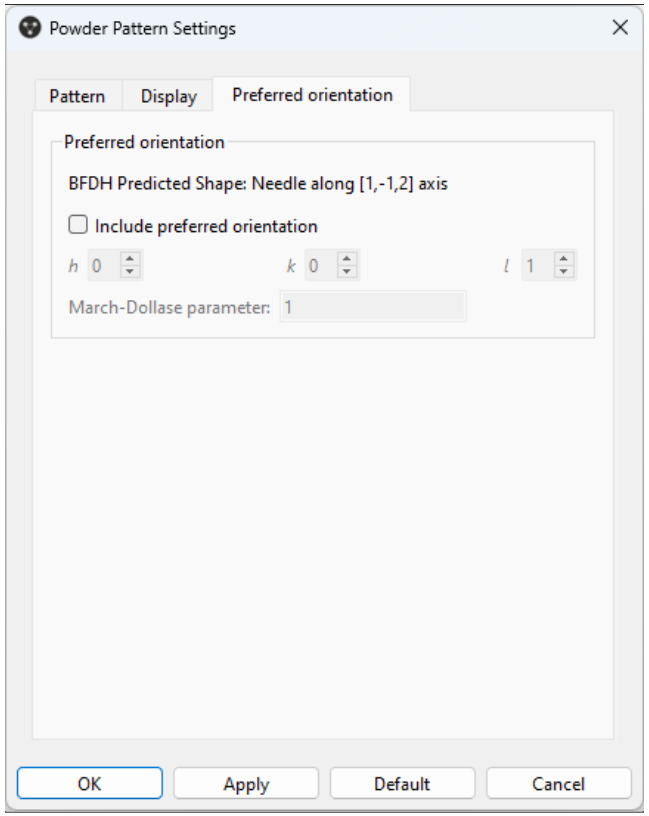
See the book "Fundamentals of Crystallography" by C. Giacovazzo, et al. for an overview.

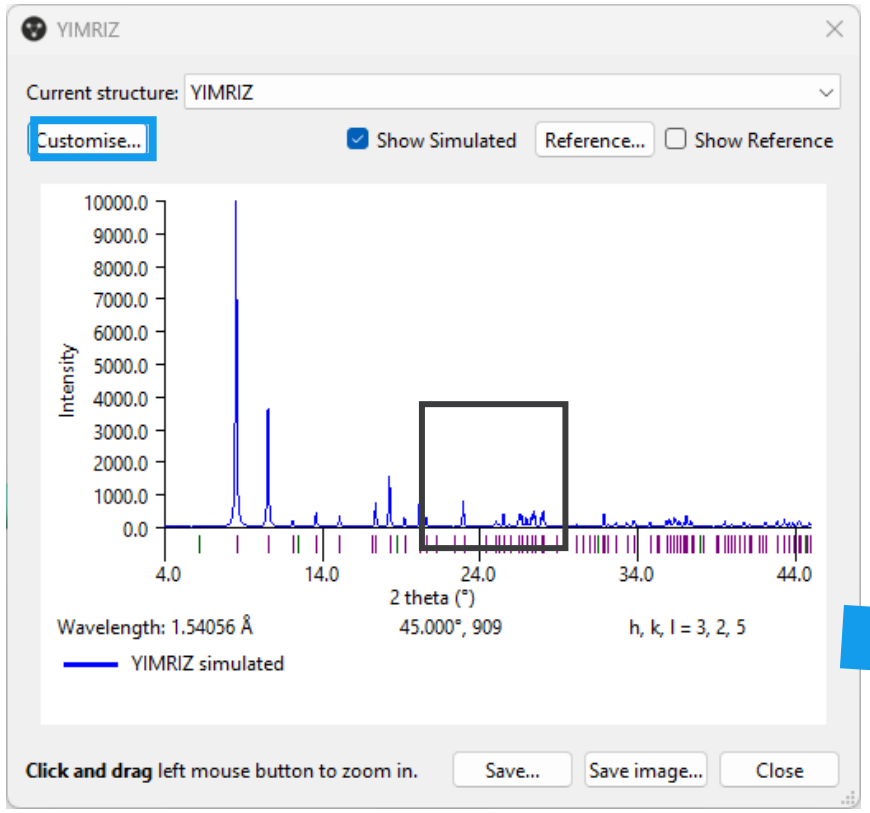


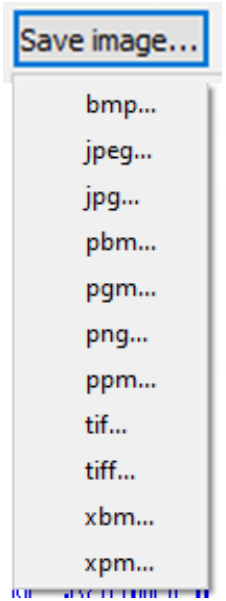
CCDC

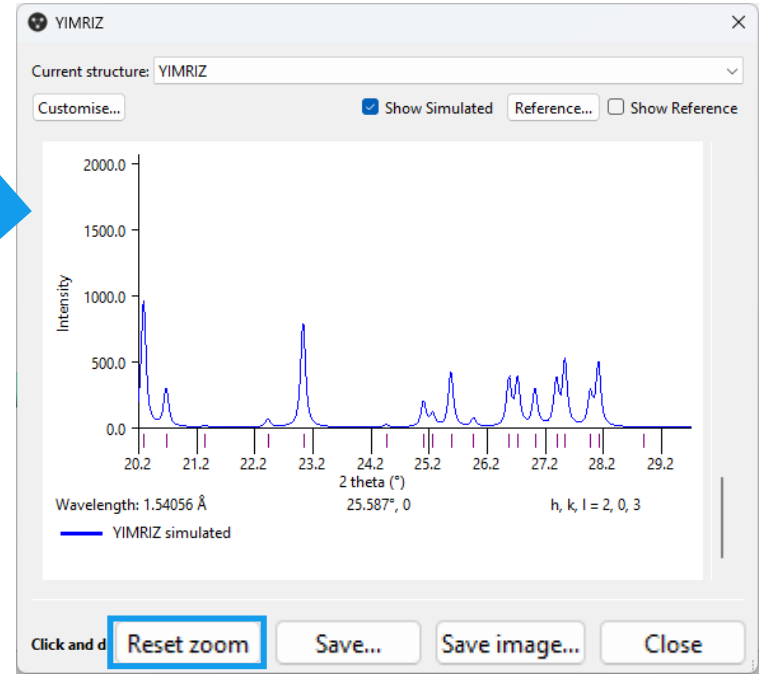
Calculate powder pattern

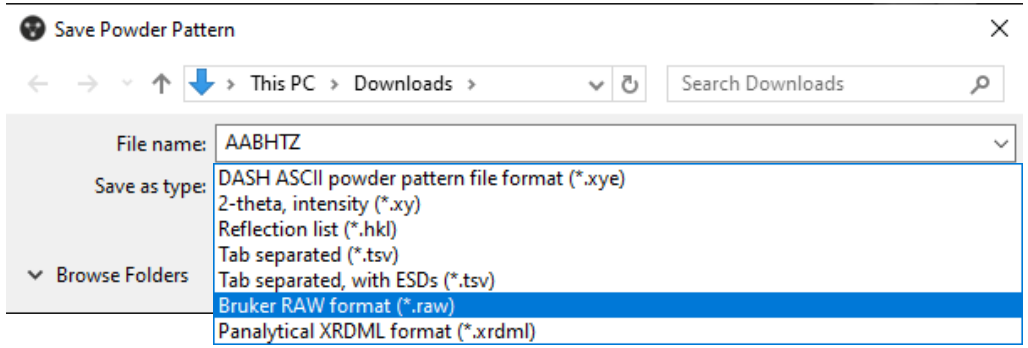


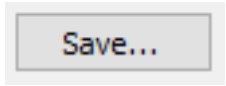




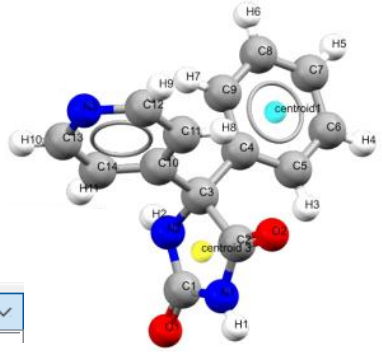






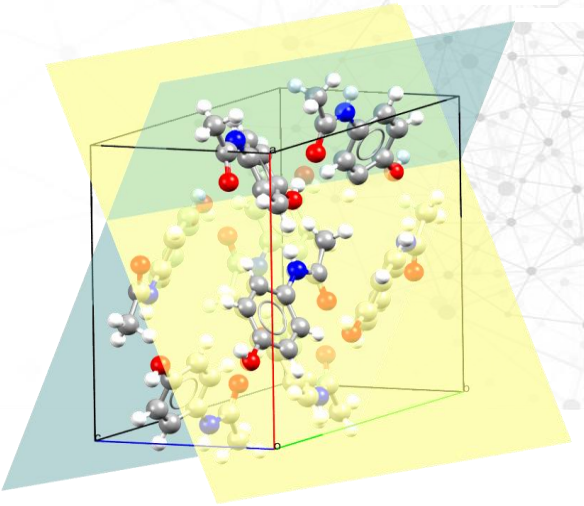


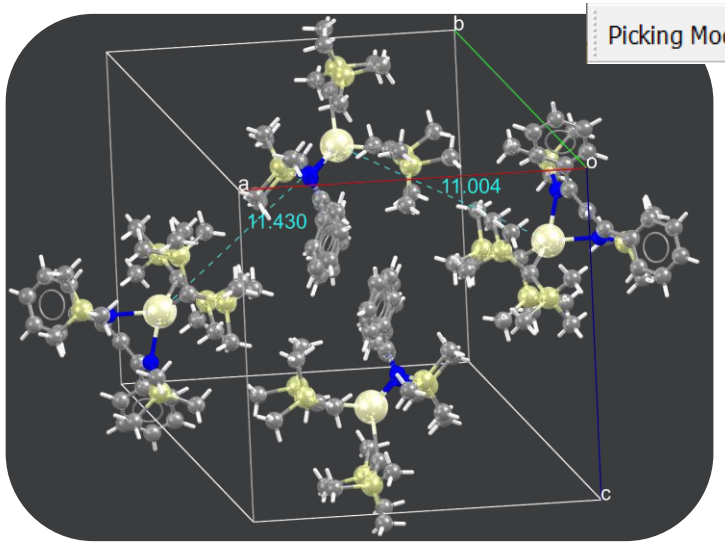
Lots of display and analysis options!



Calculate

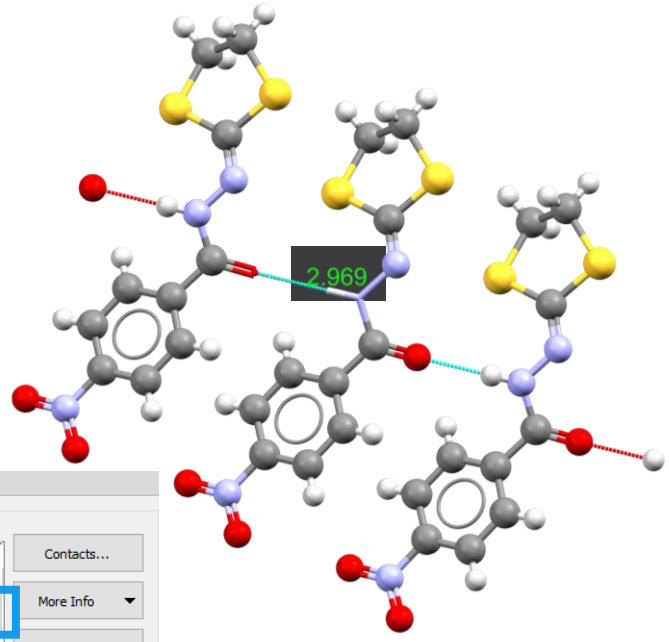
- Centroids...
- Planes...
- Packing/Slicing...
- Contacts...
- Molecular Shell...
- Graph Sets...
- Powder Pattern...
- Pore Analyser...
- Structure Overlay...
- Molecule Overlay...





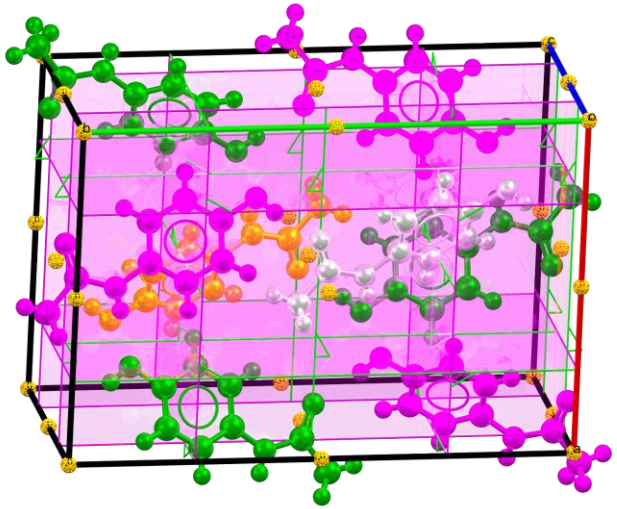
Picking Mode: Pick Atoms

- Pick Atoms
- Lasso Atoms
- Expand Contacts
- Toggle Labels
- Move Labels
- Measure Distances
- Measure Angles
- Measure Torsions
- Reveal Symmetry-Generated Molecules



Display

- Styles
- Labels
- Colours
- Show/Hide
- More Information
- Symmetry Elements...
- Voids...
- Display Options...
- Manage Styles...
- View along
- Dial box...
- Splash screen
- Toolbars



Display Options

Display

Packing
Asymmetric Unit
Auto centre

Short Contact < (sum of vdW radii)
H-Bond Default definition

Contacts...
More Info
Powder...

Reset

Read the publication on symmetry tools

CCDC

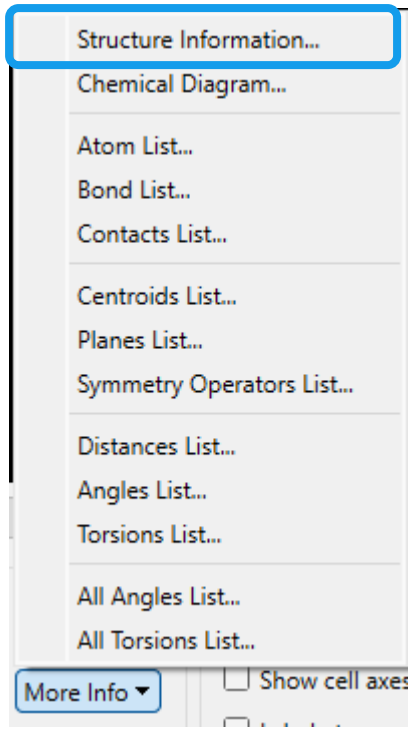
Data fields

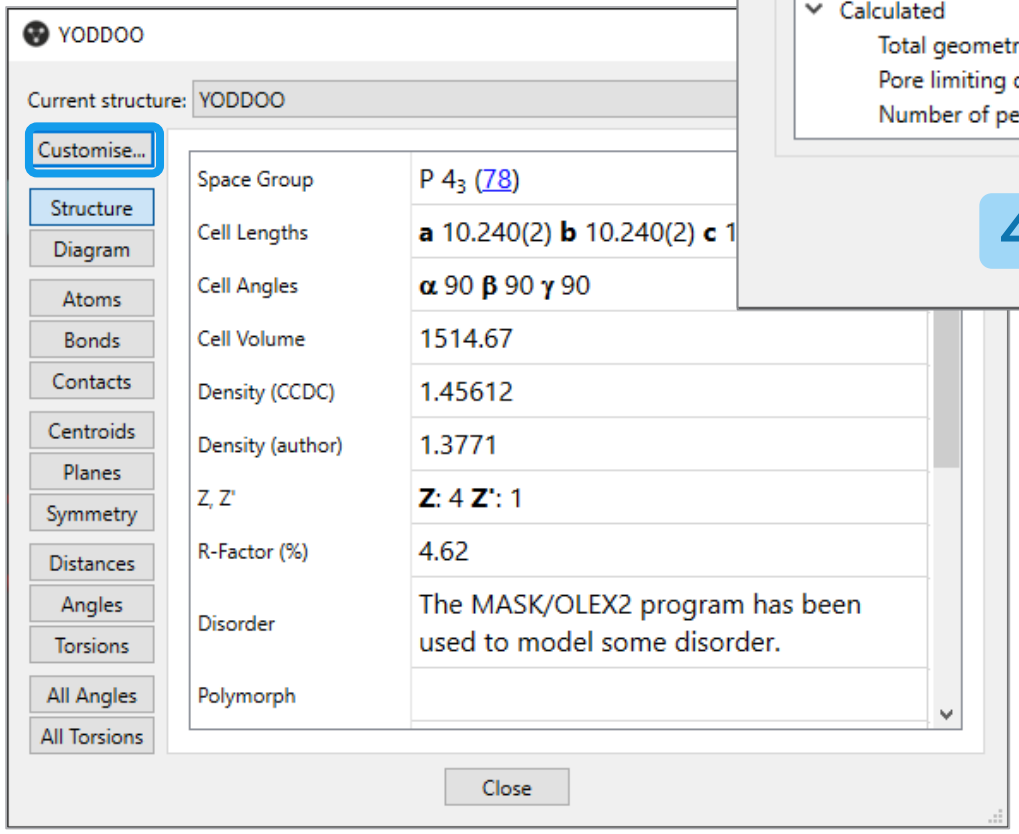
1

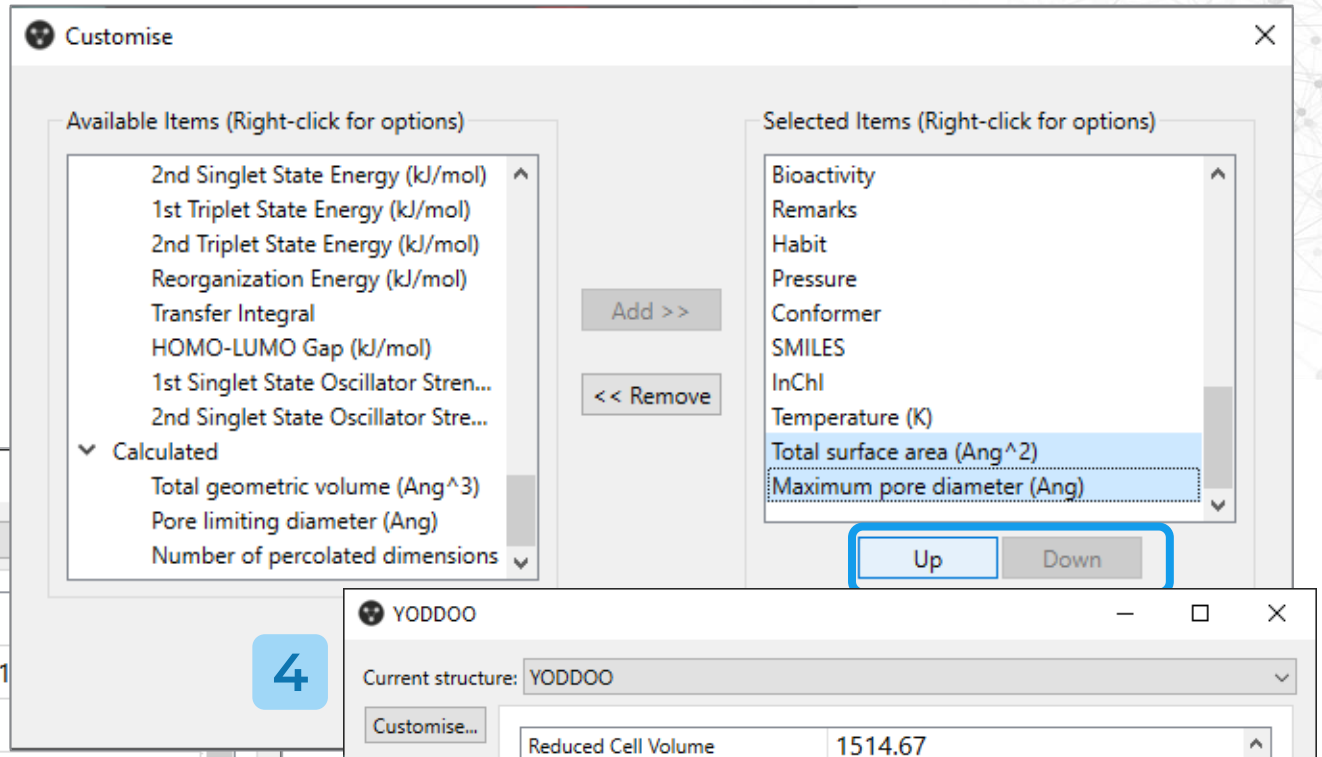
2

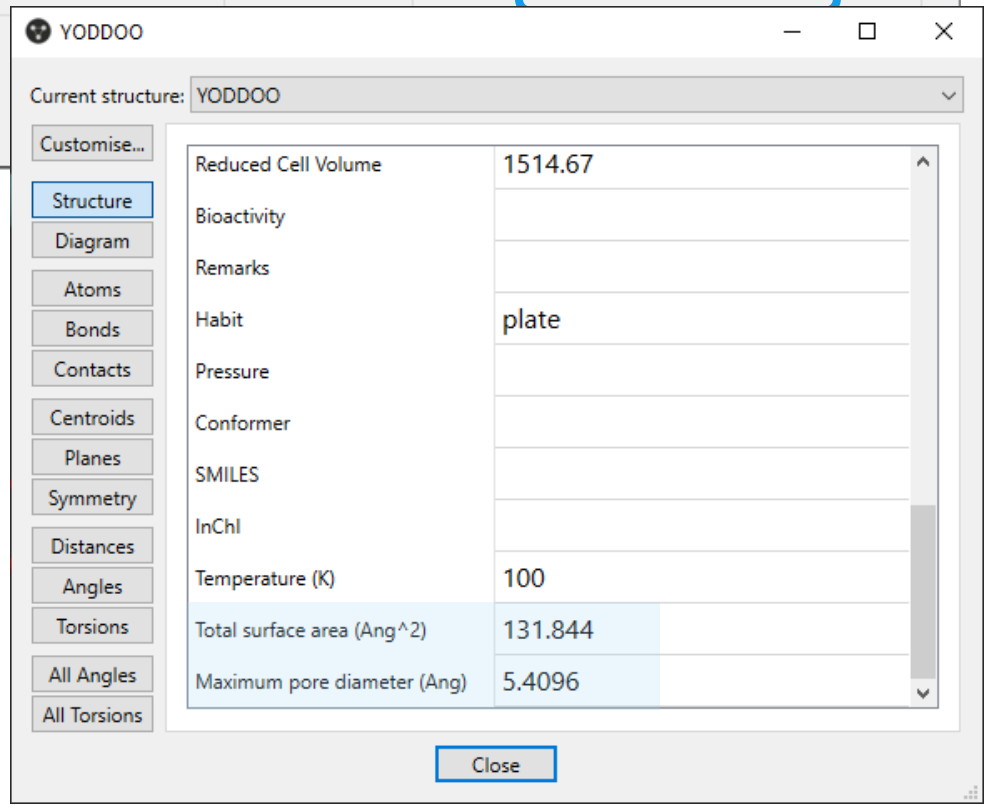
3

4









Data fields in the CSD Python API

The new properties can also be accessed using the CSD Python API from the Entry Module.

Calculated Properties

class ccdc.entry.**CrystalCalculatedProperties**(*_properties=None*, *_entry=None*) [source]

A container for different types of calculated properties of a database entry.

Currently ccdc.descriptors.CrystalDescriptors.PoreAnalyser is the only type implemented.

property **pore_analyser**: *PoreAnalyser* | *None*

Returns a ccdc.descriptors.CrystalDescriptors.PoreAnalyser with stored calculated properties for the entry.

The pore analyser properties stored are:

- ccdc.descriptors.CrystalDescriptors.PoreAnalyser.total_surface_area
- ccdc.descriptors.CrystalDescriptors.PoreAnalyser.total_geometric_volume
- ccdc.descriptors.CrystalDescriptors.PoreAnalyser.pore_limiting_diameter
- ccdc.descriptors.CrystalDescriptors.PoreAnalyser.max_pore_diameter
- ccdc.descriptors.CrystalDescriptors.PoreAnalyser.num_percolated_dimensions

Other properties of the pore analyser are not stored and will be calculated on demand.

property **refinement_goodness_of_fit**

Returns the refinement goodness of fit. Describes how well the structural model fits the experimental data.

```
>>> from ccdc.io import EntryReader
>>> csd_reader = EntryReader('CSD')
>>> arutue = csd_reader.entry('ARUTUE')
>>> arutue.refinement_goodness_of_fit
1.085
```

property **refinement_max_shift**

Returns the refinement max shift. Describes the largest ratio of the final least-squares parameter shift to the final standard uncertainty.

```
>>> from ccdc.io import EntryReader
>>> csd_reader = EntryReader('CSD')
>>> arutue = csd_reader.entry('ARUTUE')
>>> arutue.refinement_max_shift
0.001
```

property **refinement_number_of_constraints**

Returns the refinement number of constraints. Describes the number of constrained (non-refined or dependent) parameters in the least-squares process. If a model has been highly constrained users may wish to ignore this structure. Similar to No. of restraints.

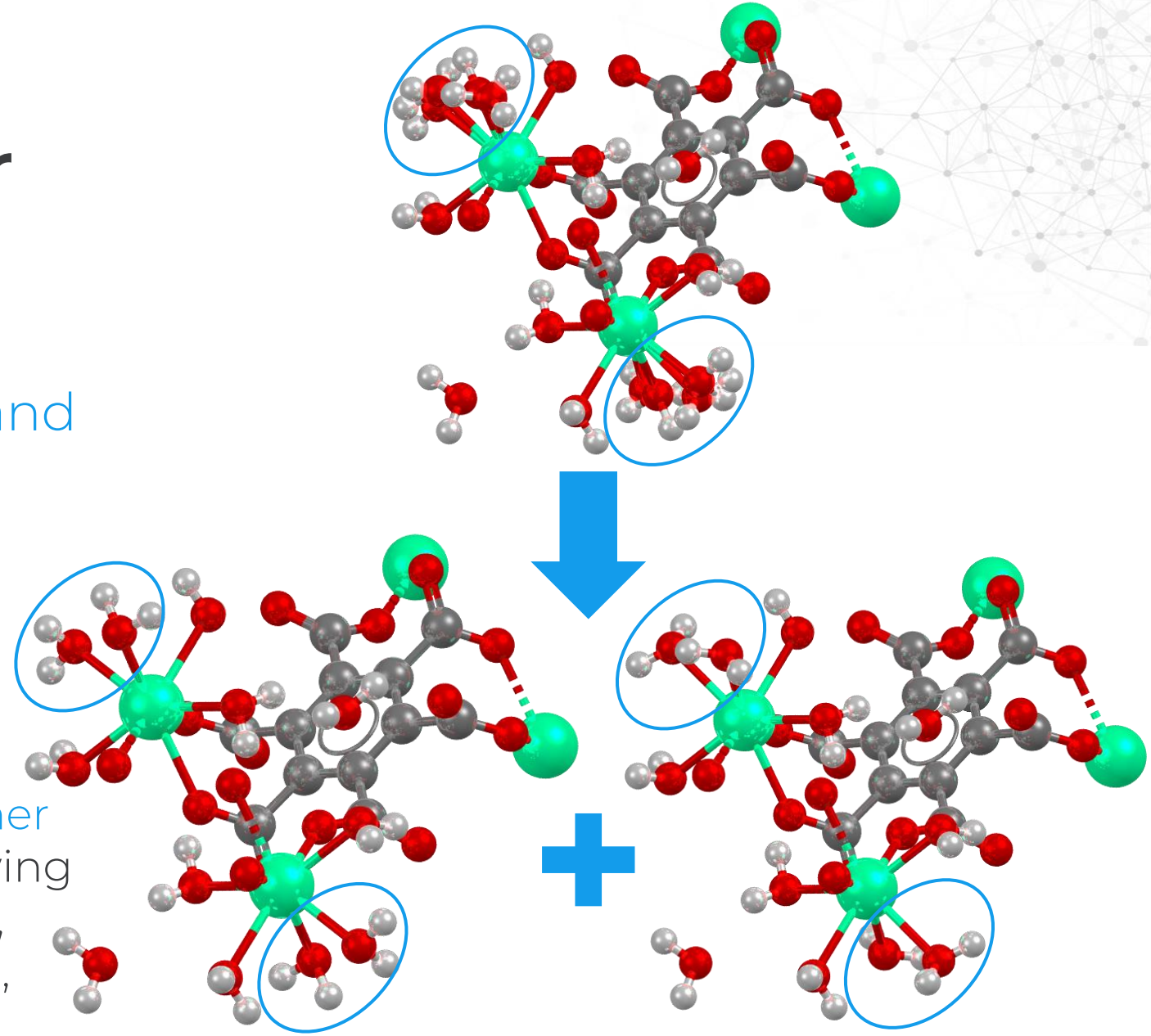
```
>>> from ccdc.io import EntryReader
>>> csd_reader = EntryReader('CSD')
>>> arutue = csd_reader.entry('ARUTUE')
>>> arutue.refinement_number_of_constraints
0
```

(Entry API Module)
https://downloads.ccdc.cam.ac.uk/documentation/API/modules/entry_api.html

CCDC

Visualizing disorder

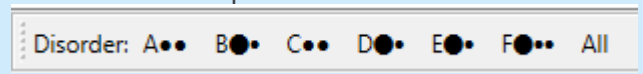
- Functionality available when you load a CIF into Mercury and for structures in the CSD: NEW
  - Visualise the different disorder models.
  - Visualise the contacts for each disorder model.
  - Works in combination with other functionality, such as Edit, viewing and selecting hydrogen bonds, displaying Voids, Pore Analyser, Mogul, CSD-Materials & CSD-Particle tools.



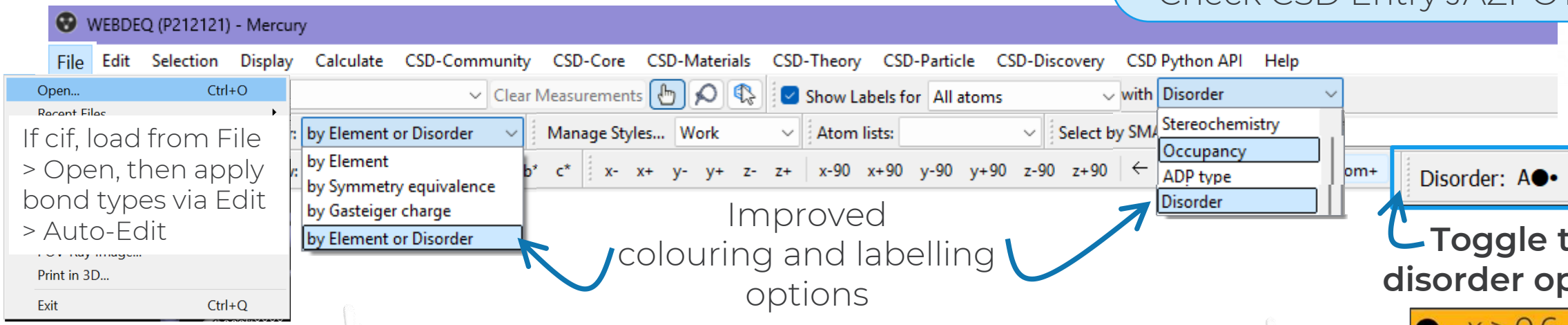
CCDC

Visualizing disorder

Note: you can have more disordered groups and more than two options each.



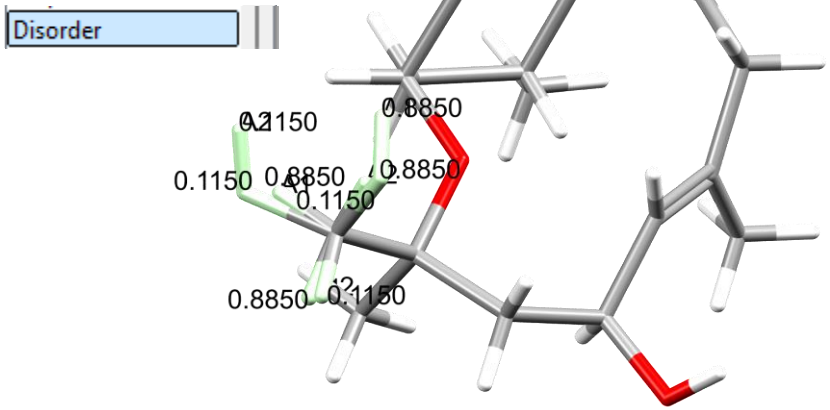
Check CSD Entry JAZPOT!

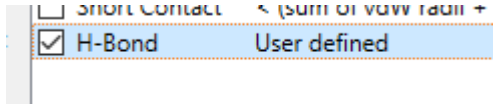


If cif, load from File > Open, then apply bond types via Edit > Auto-Edit

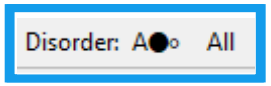
Improved colouring and labelling options

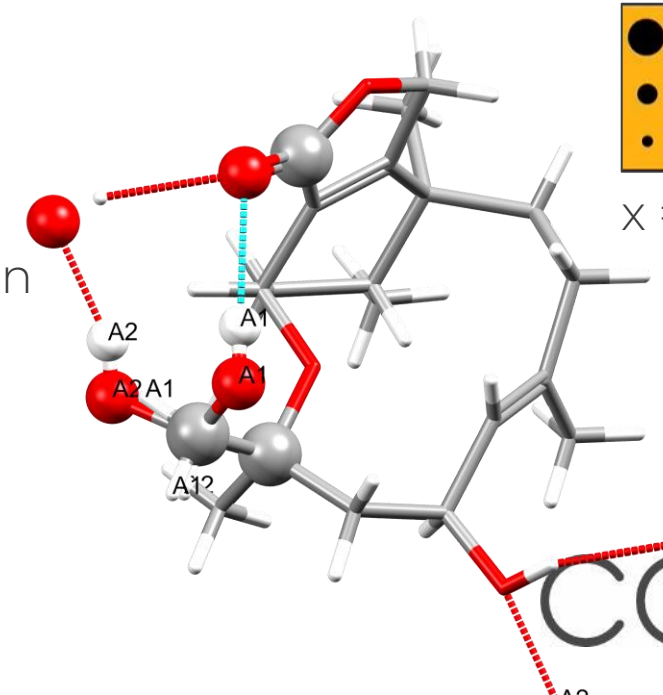
Toggle the disorder options

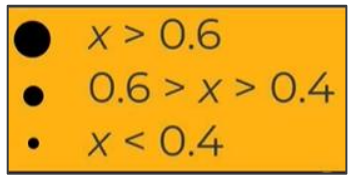




Example of combination with hydrogen bonds







x = occupancy

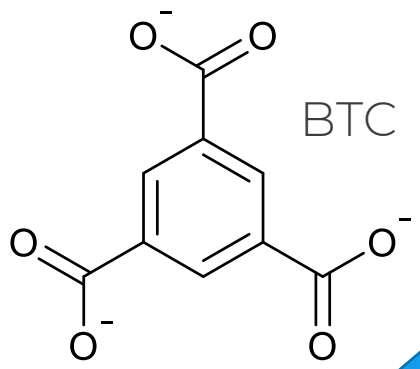
CCDC

Demo of Mercury

Search and analysis of polymeric structures in Mercury

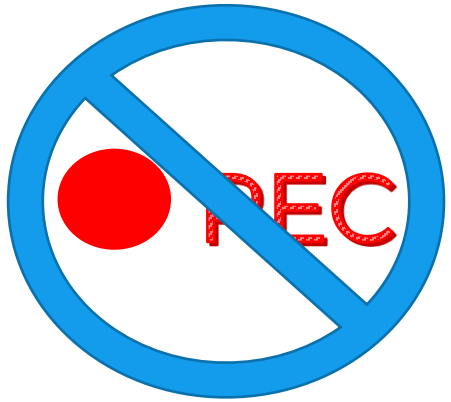
We will:

- Use Mercury to visualise the polymeric frameworks of LnBTC (Ln = lanthanoid, Y) compounds
- Compare structures with a Structure Overlay
- Calculate Pore properties
- Simulate and compare PXRD patterns



We are recording today's session and will make the recording available to you in the next few days

Try One

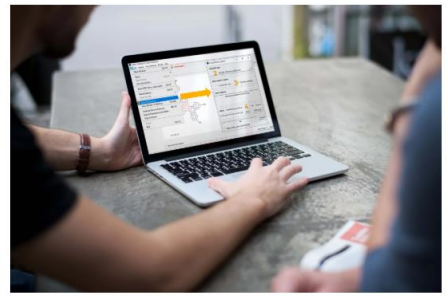


It's your turn!

- Try one example from the handout
- Your tutors are on hand to help you!
- To ask questions during this time type a message in the chat box
- If you finish early, try the other example or try analyse your structures

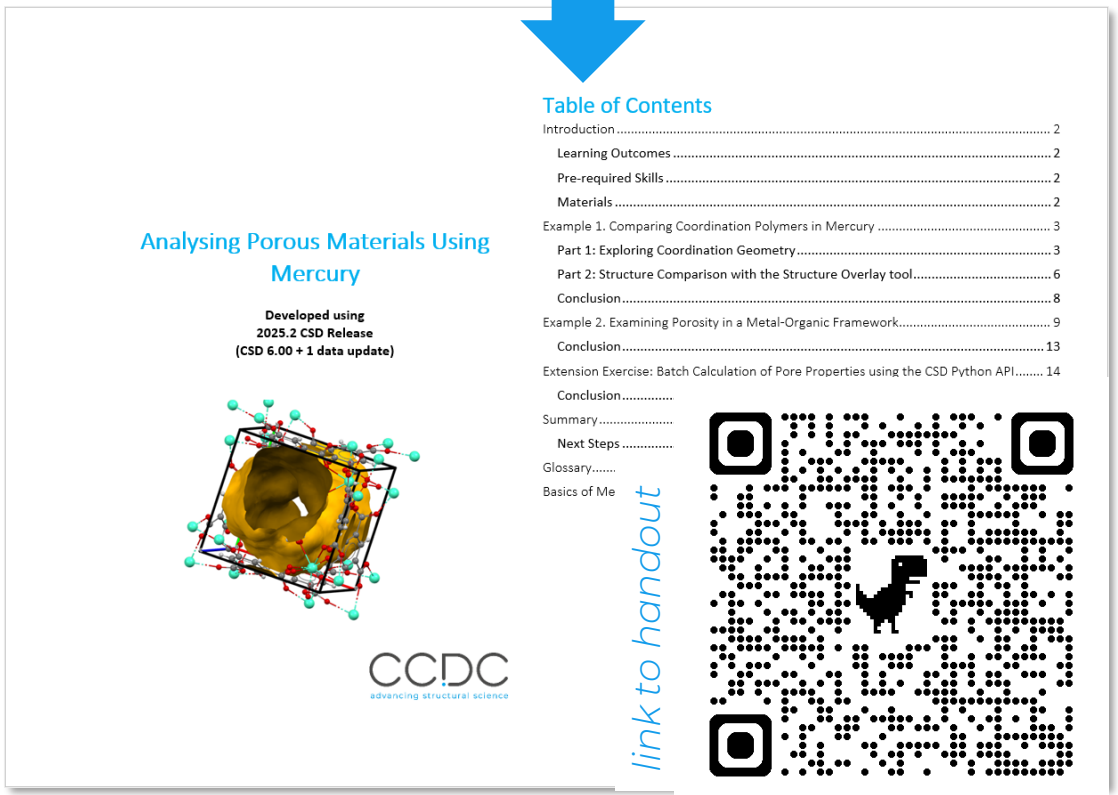
https://info.ccdc.cam.ac.uk/2025-autumn-virtual-workshops

link to webpage



Advanced Structural Visualization and Analysis of Porous Materials Using Mercury

Tuesday, 7th October
13:30-14:15 (BST)



Analysing Porous Materials Using Mercury

Developed using 2025.2 CSD Release (CSD 6.00 + 1 data update)

CCDC
advancing structural science

Table of Contents

link to handout

Functionality in the CSD Python API

CSD Python API 3.5.0 documentation » API documentation » Descriptors API

CCDC
advancing structural science

Table of Contents

Quick search

Go

Hide Search Matches

Descriptors API

Introduction

The ccdc.descriptors module contains classes for calculating descriptors.

The main classes in the ccdc.descriptors module are:

- ccdc.descriptors.MolecularDescriptors.
- ccdc.descriptors.GeometricDescriptors.
- ccdc.descriptors.CrystalDescriptors.PowderPattern.
- ccdc.descriptors.CrystalDescriptors.Morphology.
- ccdc.descriptors.CrystalDescriptors.GraphSetSearch.
- ccdc.descriptors.CrystalDescriptors.HBondCoordination.
- ccdc.descriptors.CrystalDescriptors.HBondPropensities.
- ccdc.descriptors.StatisticalDescriptors.

class **PoreAnalyser**(*crystal*, *settings=None*)

Calculates Pore Analysis. crystal is ccdc.crystal.Crystal

class **Flags**

Flags for validlity of cached variables

property **calculator_is_valid**

grid spacing (A)

class **Settings**

Settings for PoreAnalyser

property **cutoff_distance**

Cut-off distance (A)

property **grid_spacing**

grid spacing (A)

property **he_probe_epsilon**

UFF L-J epsilon/k for He probe (K)

property **he_probe_sigma**

UFF L-J sigma for He probe (A)

property **n_probe_sigma**

UFF L-J sigma for N probe (A)

property **samples_per_atom**

Sample size for surface area calculation

ation » API documentation » Crystal API

Crystal API

Introduction

The main class of the ccdc.crystal module is ccdc.crystal.Crystal.

ccdc.crystal.Crystal contains attributes and functions that relate to crystallography. An ex
graphic attribute is the ccdc.crystal.Crystal.cell_volume.

```
>>> from ccdc.io import CrystalReader
>>> csd_crystal_reader = CrystalReader('CSD')
>>> first_csd_crystal = csd_crystal_reader[0]
>>> round(first_csd_crystal.cell_volume, 3)
769.978
```

void_volume(*probe_radius=1.2, grid_spacing=0.7, mode='contact'*)

Determine the void volume of the crystal.

```
>>> from ccdc.io import EntryReader
>>> entry_reader = EntryReader('CSD')
>>> abawop_crystal = entry_reader.crystal('ABAWOP')
>>> round(abawop_crystal.void_volume(), 2)
13.99
```

Parameters:
- **probe_radius** – float, size of the probe
- **grid_spacing** – float, fineness of the grid on which the calculation is
- **mode** – either 'accessible' or 'contact' according to whether the cen
  the whole probe must be accomodated.

Returns: void volume as a percentage of the unit cell volume

Polymer Expansion

Polymeric crystals may be expanded, as in Mercury, with the method ccdc.crystal.
For example:

```
>>> crystal = csd_reader.crystal('ABALOF')
>>> print(len(crystal.molecule.atoms))
36
>>> mol = crystal.polymer_expansion()
>>> print(len(mol.atoms))
100
```

There is an optional repetitions argument to ccdc.crystal.polymer_expansion() to co
sions performed:

```
>>> mol = crystal.polymer_expansion(repetitions=5)
>>> print(len(mol.atoms))
356
```

There is another keyword argument to ccdc.crystal.Crystal.polymer_expansion(), atom
control over the expansion is required. The atoms provided will be matched by la
crystal's polymeric bonds, so the repetitions argument can still be honoured.

```
>>> original_mol = crystal.molecule
>>> atom_to_expand = original_mol.atom('I4C')
>>> mol = crystal.polymer_expansion(atoms=[atom_to_expand])
>>> print(len(mol.atoms))
68
>>> mol = crystal.polymer_expansion(atoms=[atom_to_expand], repetitions=5)
>>> print(len(mol.atoms))
196
>>> print(len([a for a in mol.atoms if a.label == 'Cu1A']))
0
>>> print(len([a for a in mol.atoms if a.label == 'Cu3A']))
0
```

CSD Python API 3.5.0 documentation » Search

Search

Searching for multiple words only shows matches that contain all words.

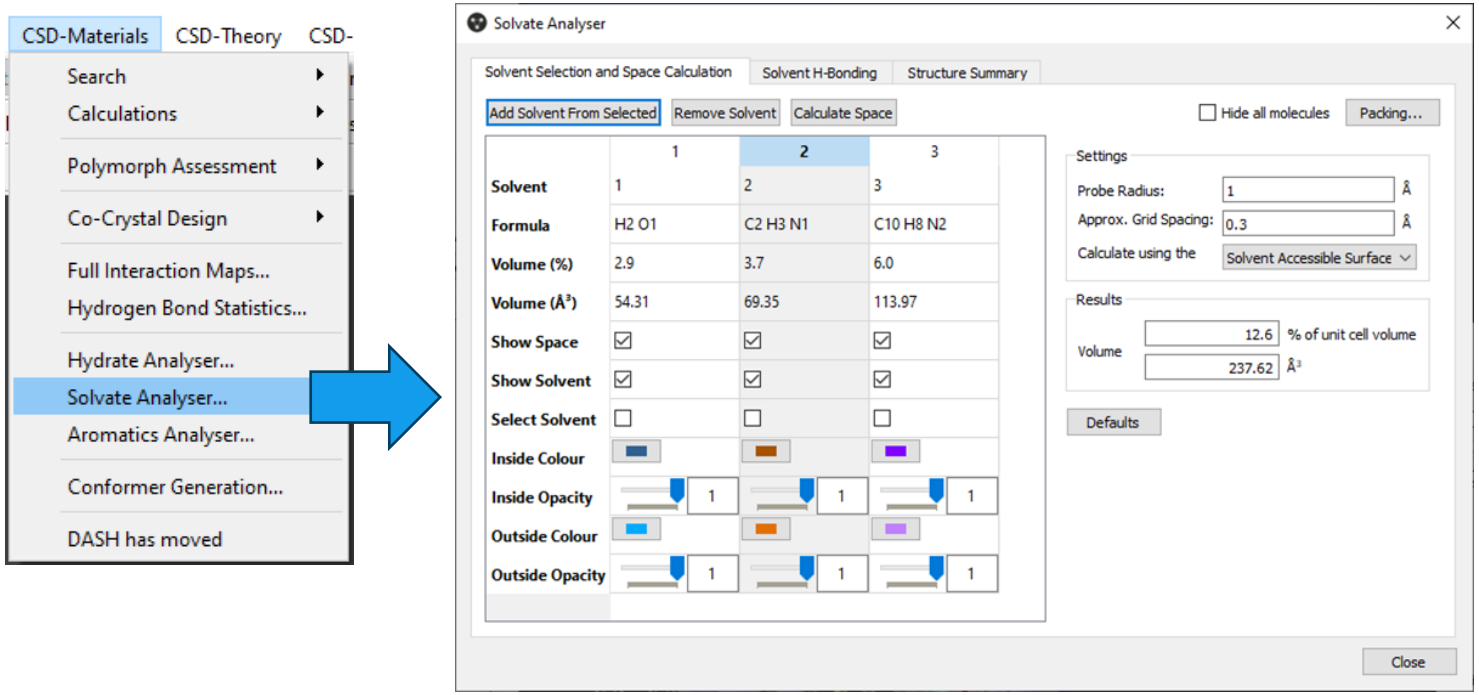
search

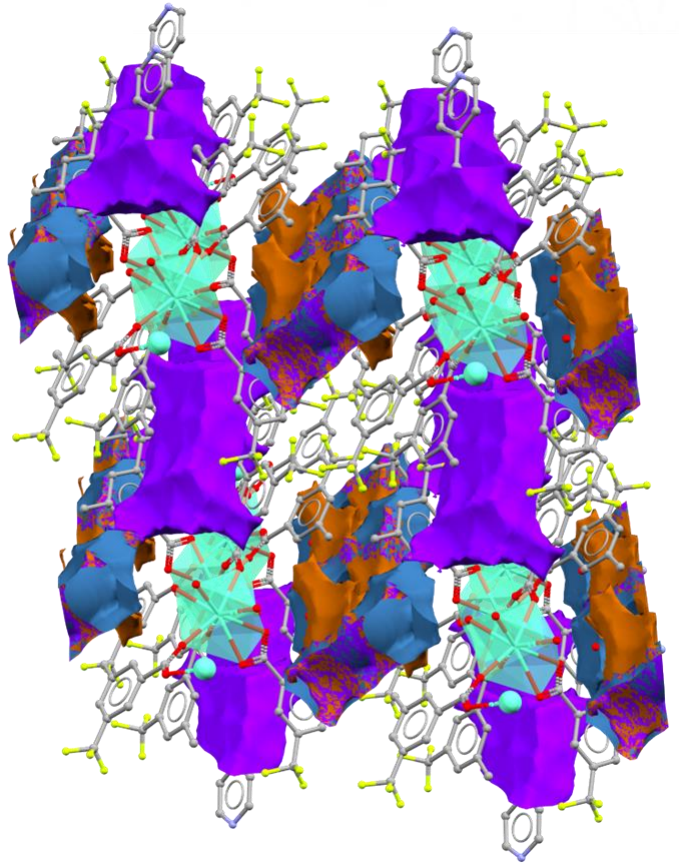
Find more functionality using the search bar!

https://downloads.ccdc.cam.ac.uk/documentation/API/search.html?q=

Solvate Volume

Solvate Analyser enables to visualize the volume occupied by inclusion molecules such as solvates

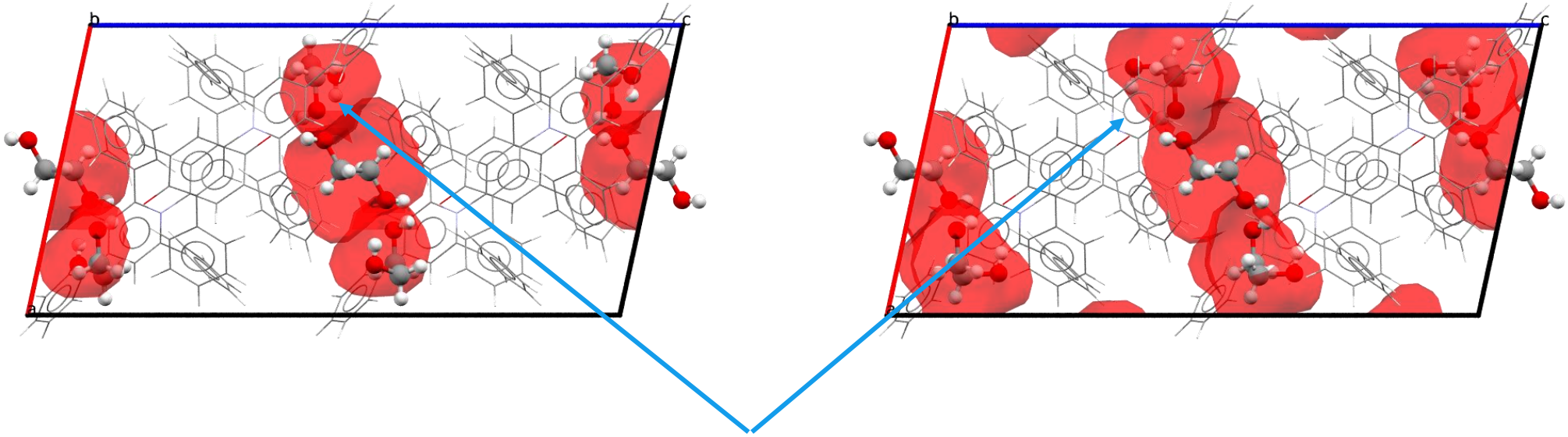




CCDC

Disorder and Solvate Analyser

- Solvate Analyser works with disorder producing different results based on the disorder group selected



Space calculated with different disorder groups active.

CCDC

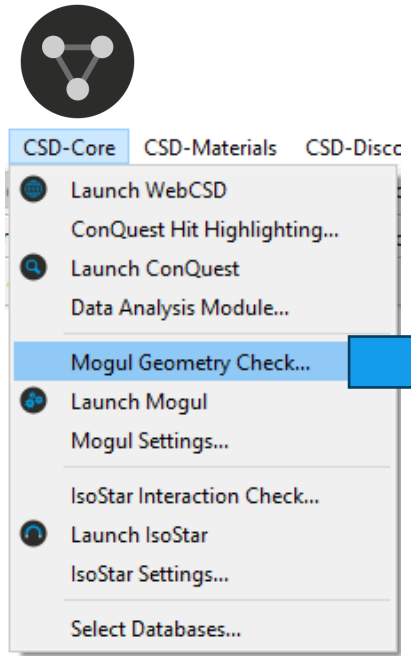
Disorder and Mogul

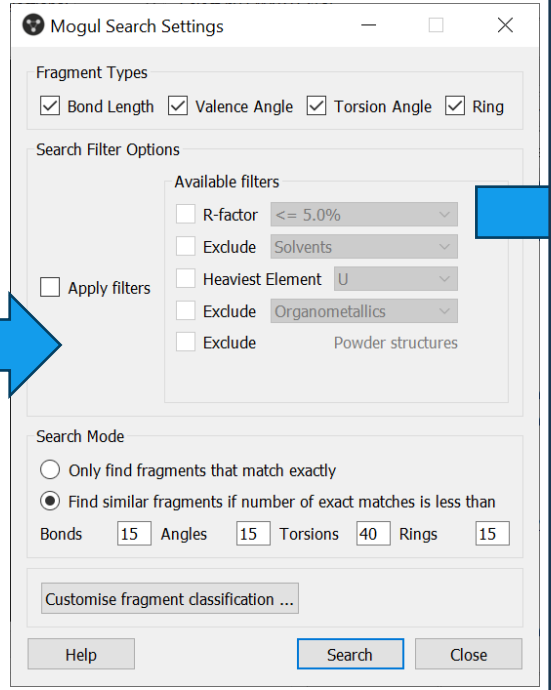
Mogul provides precise information on preferred molecular geometries.

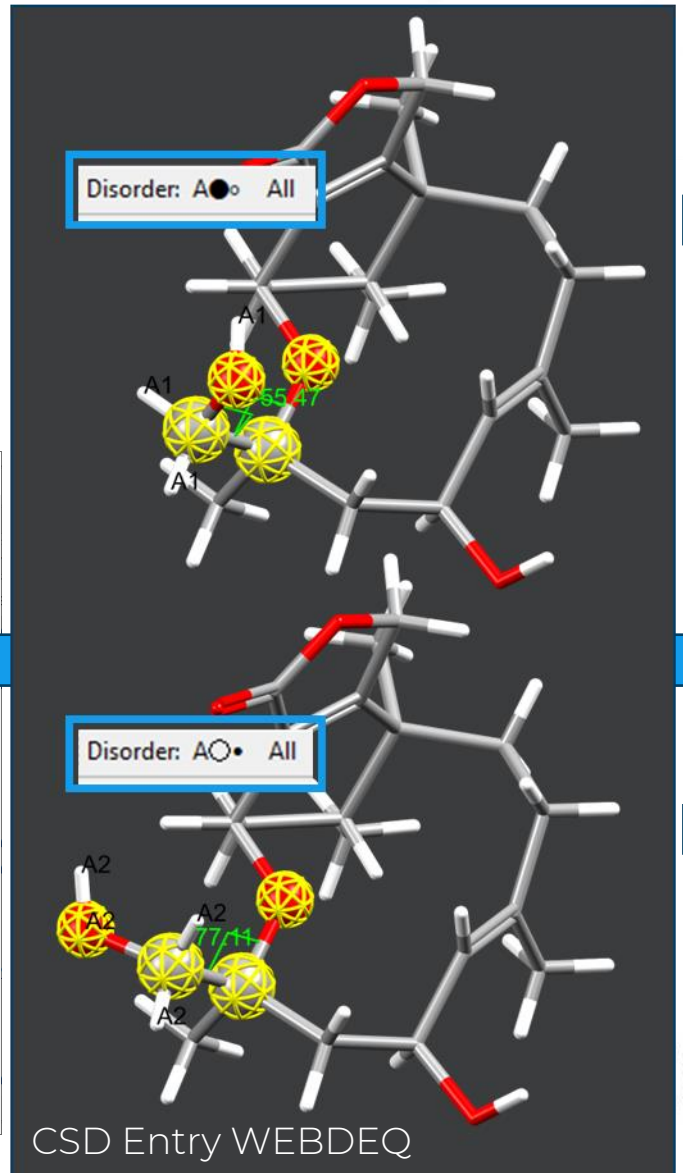
- Enables you to validate the chemistry of your structure against the experimental data in the CSD

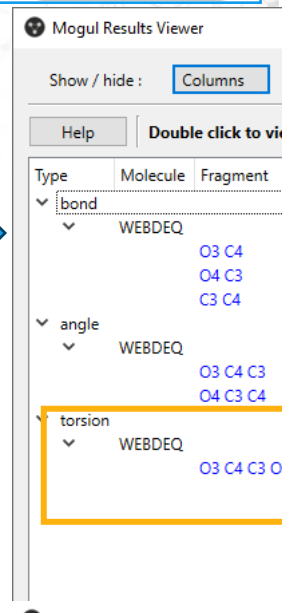
To see the results for different disorder options:

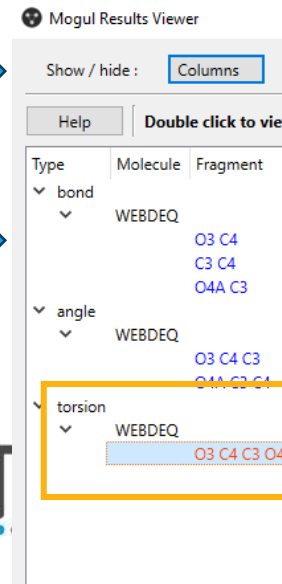
- Toggle the disorder button
- Then use Mogul Geometry Check as usual
- Repeat for the other options











Bruno *et al., J. Chem. Inf. Comput. Sci.*, 44, 2133-2144, 2004

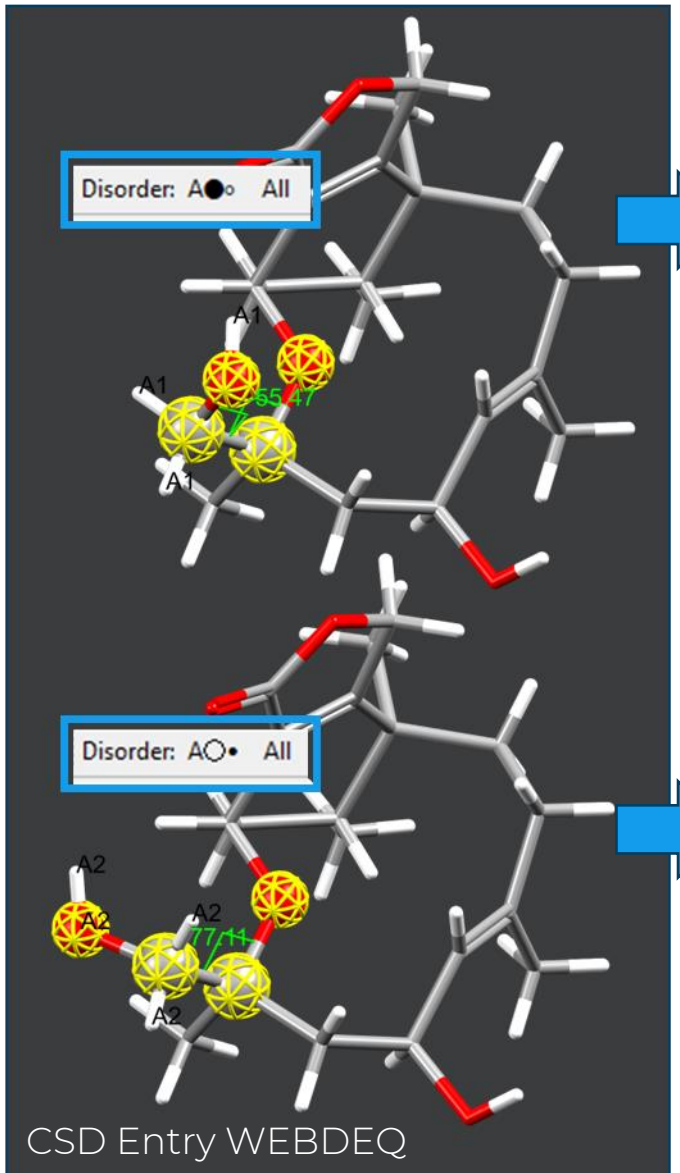
On-demand training module

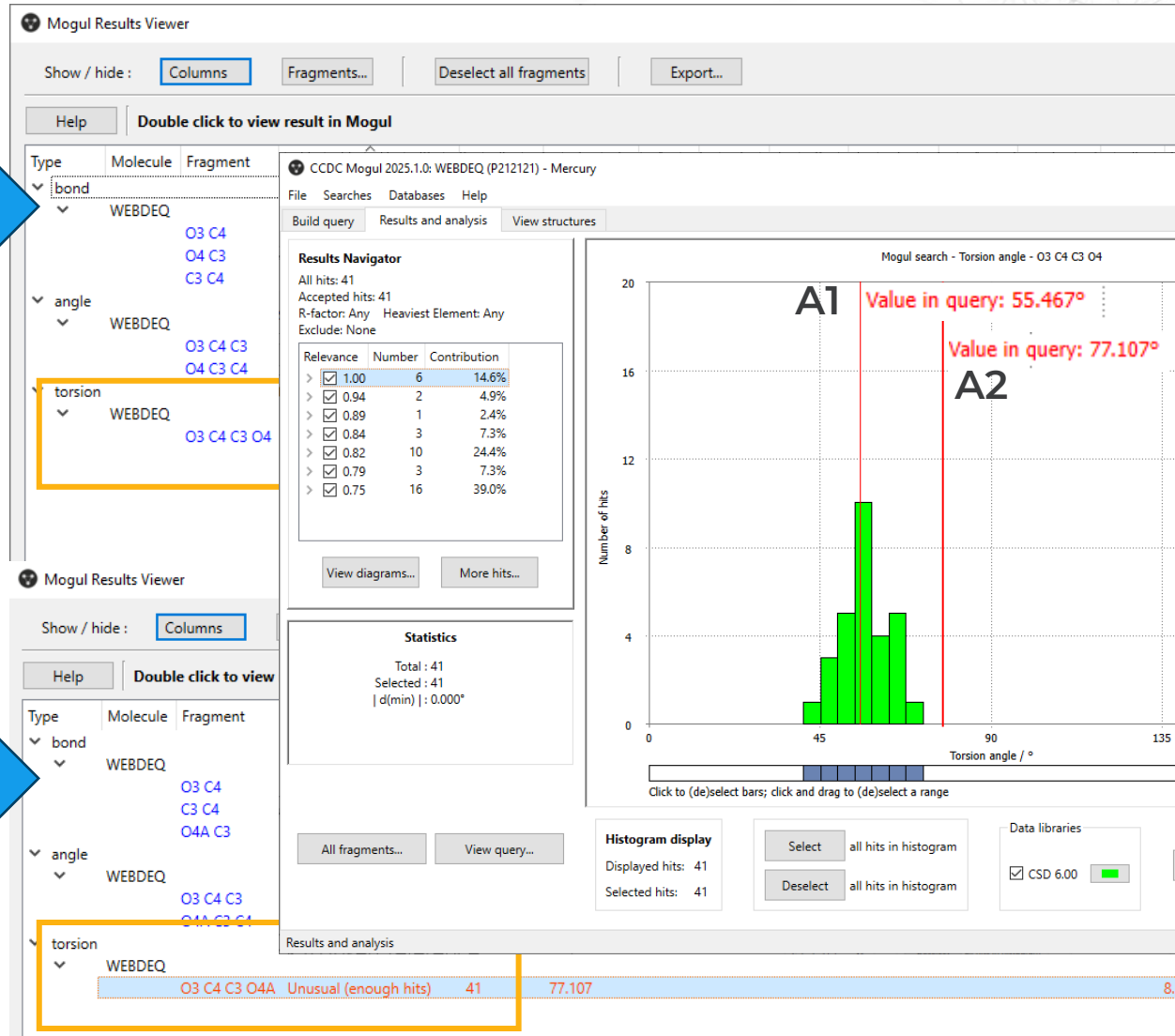
Disorder and Mogul

On-demand training module

To see the results for different disorder options:

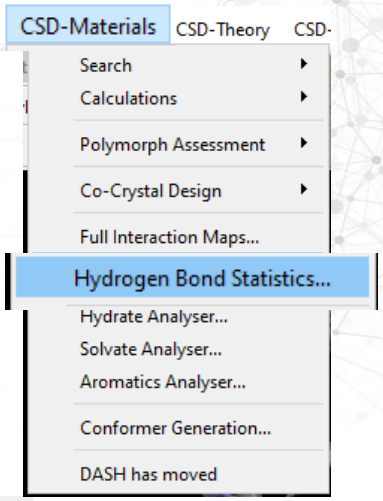
- Toggle the disorder button
- Then use Mogul Geometry Check as usual
- Repeat for the other options





Bruno *et al., J. Chem. Inf. Comput. Sci.*, 44, 2133-2144, 2004

Disorder and Hydrogen Bond Statistics

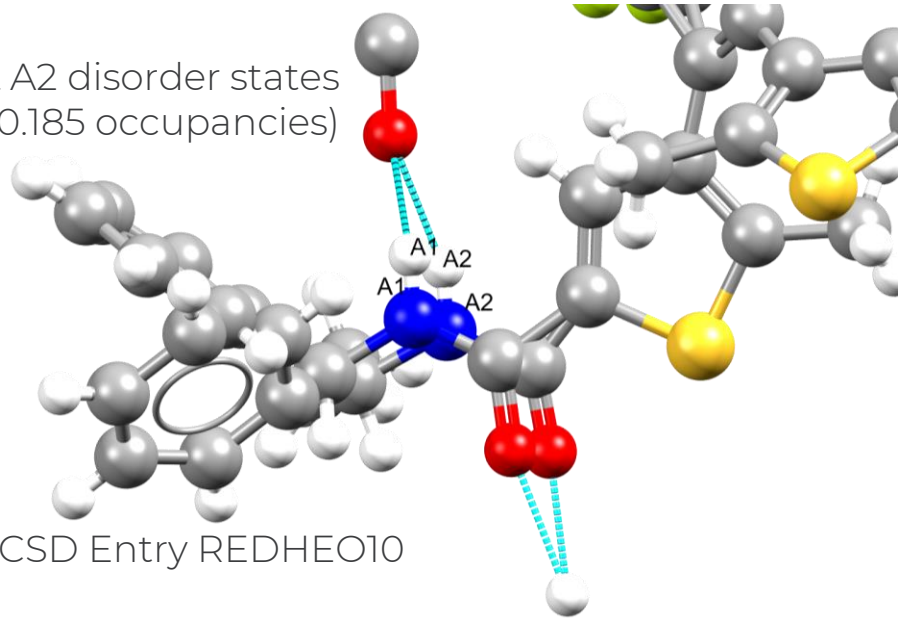


Hydrogen Bond Statistics

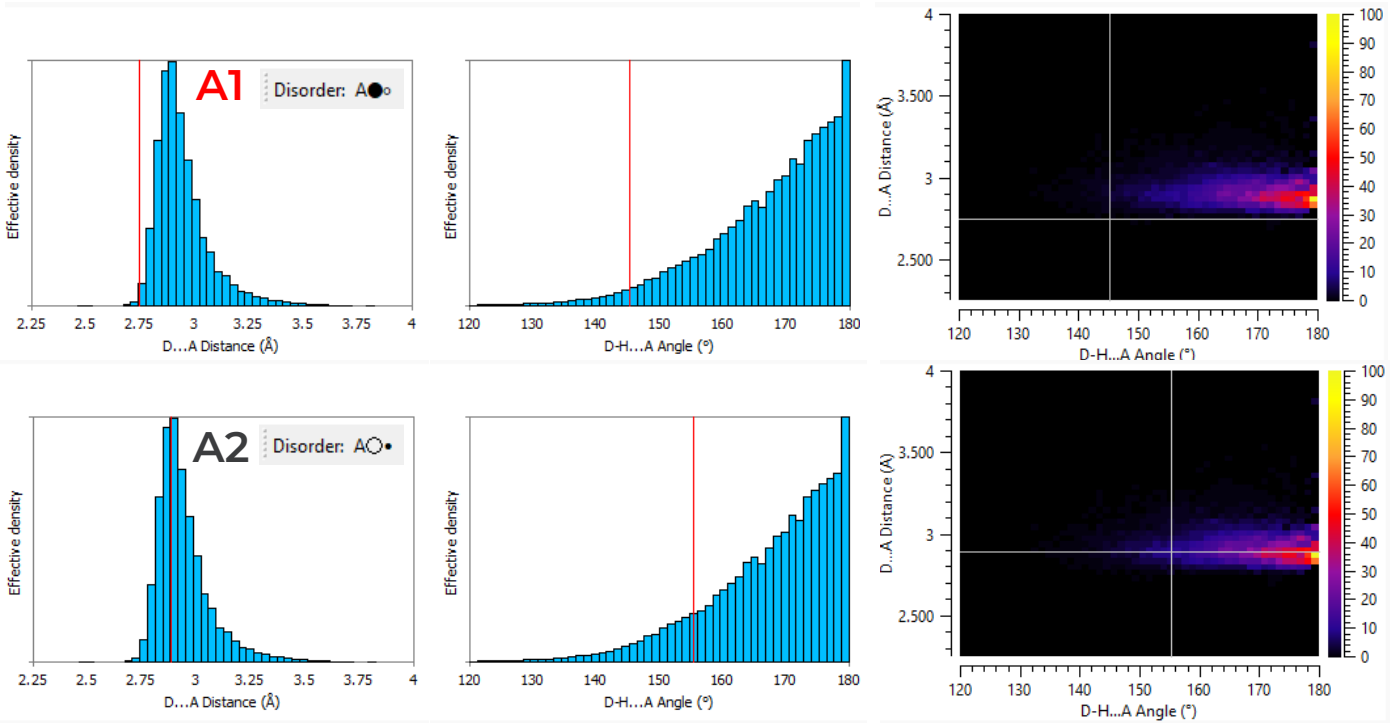
- Interactions geometries in the context of million+ crystal structures in the CSD
- Analysis of the usual- or unusual-ness of observed hydrogen bond lengths/angles in a crystal structure

	Donor	Acceptor	Distance D-A	Distance classification	Distance threshold	Angle D-H...A	Angle classification	Angle threshold
A1	N4A (amide_carbonyl) (A1)	O1 (amide_carbonyl)	2.74	Unusual	(2.80, 3.19)	145.21	Unusual	147.52
A2	N4B (amide_carbonyl) (A2)	O1 (amide_carbonyl)	2.89	Not Unusual	(2.80, 3.19)	155.33	Not Unusual	147.52

A1 & A2 disorder states (0.815:0.185 occupancies)

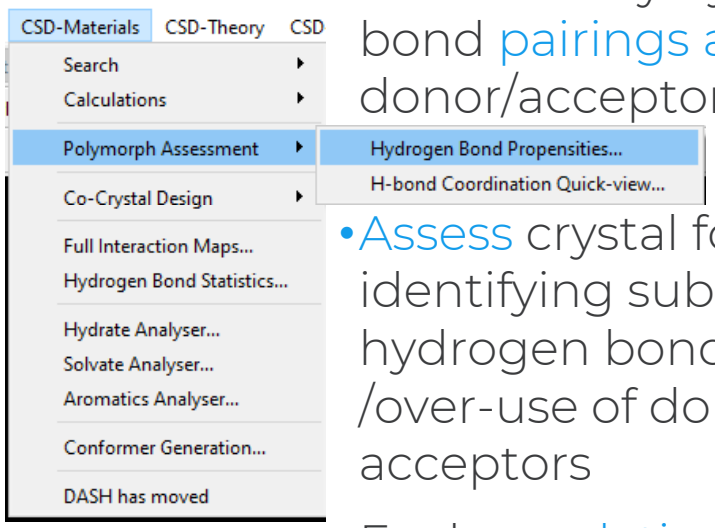


CSD Entry REDHEO10



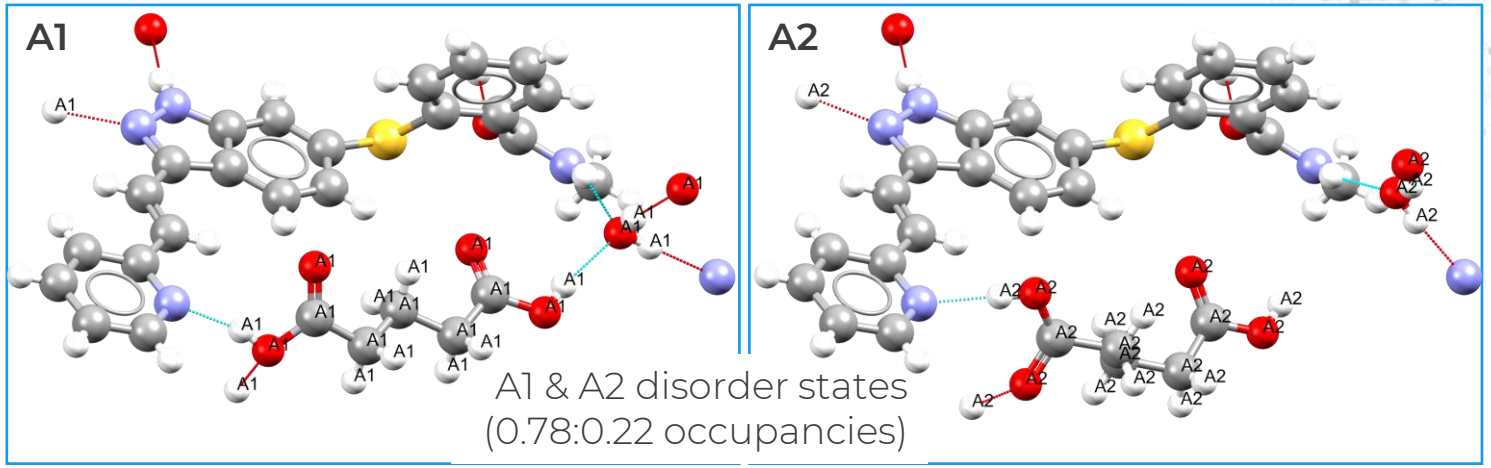
Disorder and Hydrogen Bond Propensities

- **Hydrogen Bond Propensities**
  - Predict likely hydrogen bond pairings and donor/acceptor coordination likelihoods
  - Assess crystal forms, e.g. identifying sub-optimal hydrogen bonding or under-/over-use of donors and acceptors
  - Explore relative likelihoods of putative h-bond networks

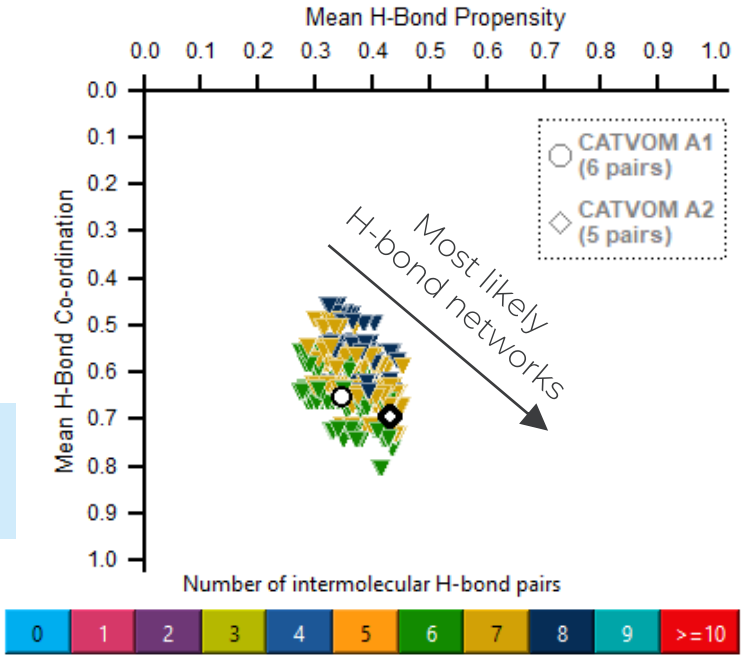


Other tools include Full Interaction Maps (FIMs) and Aromatics Analyser

CSD Entry CATVOM



A1 & A2 disorder states (0.78:0.22 occupancies)



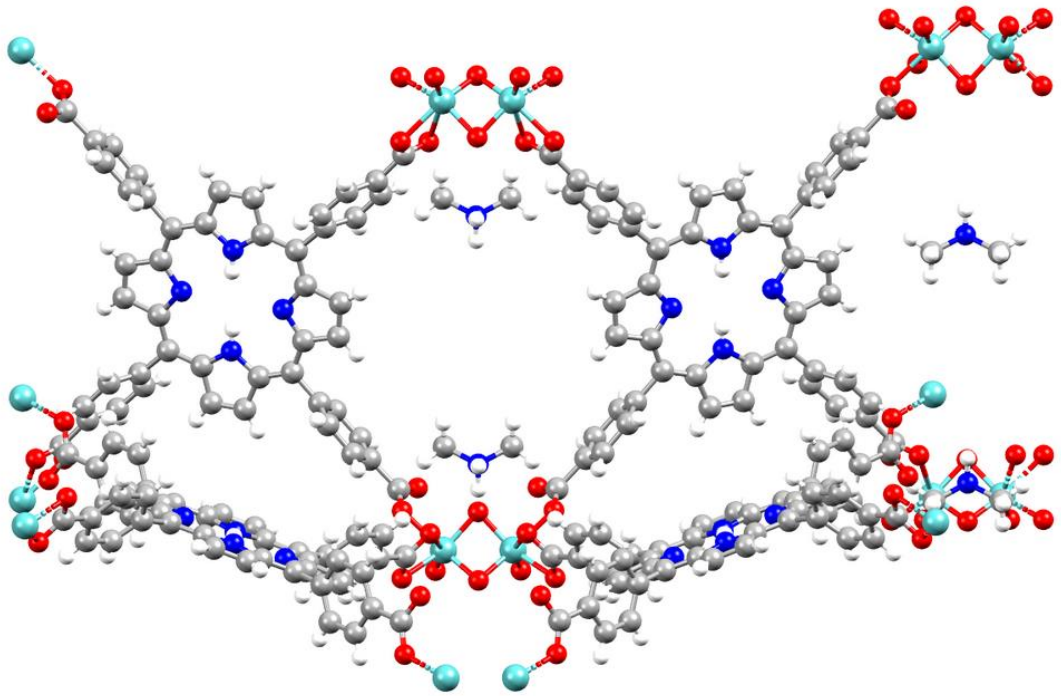
- A1 low propensity pair, donors used optimal # times
- A2 higher ranked pair, less pairs, under-used donor

CCDC

CSD MOF Collection

- Complements our CSD MOF subsets embedded in CSD Portfolio
- A good starting point for high-throughput analysis
- 12,505 3D porous MOFs
  - All 3D MOFs containing over 10 % void space
  - Converted to space group P1
  - Non-bonded solvent removed
  - Missing hydrogen atoms added

3D MOF Collection Entry



CCDC

CSD MOF Collection

- The CSD MOF Collection contains reference spreadsheets
  - Allows further categorisation within the collection

CSD refcode	CIF filename	original crystal system	Sohncke space group	Percentage void spa
ABAVIJ	abavij_P1	monoclinic	-	15.
ABAVOP	abavop_P1	monoclinic	-	13.
ABAYIO	abayio_P1_charged	cubic	-	62.
ABAZAF	abazaf_P1	monoclinic	-	21.
ABAZAF01	abazaf01_P1	monoclinic	-	22.

- The CSD MOF Collection has a Creative Commons licence: CC BY-NC-SA 4.0
- https://www.ccdc.cam.ac.uk/free-products/csd-mof-collection/
- https://doi.org/10.1016/j.matt.2021.03.006

Matter

CellPress

Volume 4, Issue 4, 7 April 2021, Pages 1105-1106

Preview

The launch of a freely accessible MOF CIF collection from the CSD

Aurelia Li [1], Rocio Bueno Perez [1], Seth Wiggin [2], Suzanna C. Ward [2], Peter A. Wood [2], David Fairen-Jimenez [1]

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.matt.2021.03.006 Get rights and content

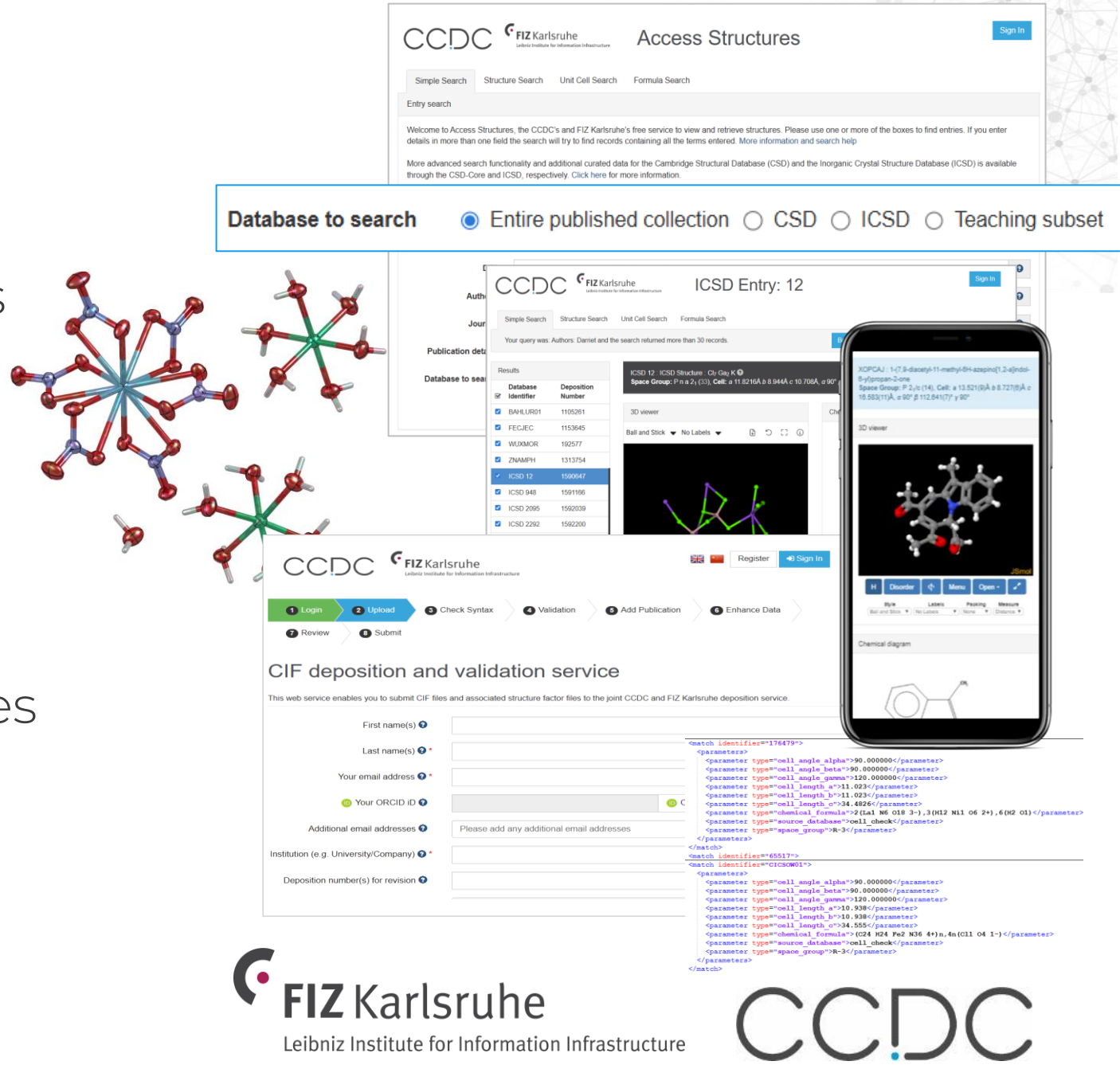
Under an Elsevier user license open archive

The computational study of metal-organic frameworks (MOFs) reliesheavily on the availability, quality, and simulation-readiness of MOF structural data. We announce here the release of the freely accessible and regularly updated "CSD MOF Collection," a dataset of >10,000 3D porous computation-ready MOFs, derived from our automatically updated CSD MOF subset.

CCDC

CSD and ICSD

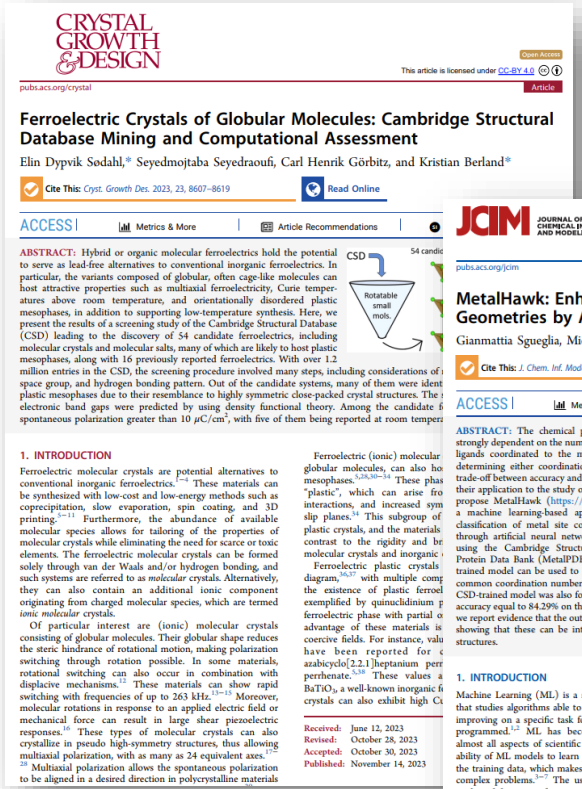
- Porous materials encompass structures that may be considered inorganic, metal-organic or organic
- It is possible to search for entries in both the CSD and ICSD using Access Structures
- Joint deposition system too!



FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

CCDC

Creating knowledge from the collection



CRYSTAL GROWTH & DESIGN

pubs.acs.org/crystal

Open Access

This article is licensed under CC-BY 4.0

Article

Ferroelectric Crystals of Globular Molecules: Cambridge Structural Database Mining and Computational Assessment

Elin Dypvik Sødahl,* Seyedmojtaba Seyedraoufi, Carl Henrik Görbitz, and Kristian Berland*

Cite This: *Cryst. Growth Des.* 2023, 23, 8607−8619

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Hybrid or organic molecular ferroelectrics hold the potential to serve as lead-free alternatives to conventional inorganic ferroelectrics. In particular, the variants composed of globular, often cage-like molecules can host attractive properties such as multiaxial ferroelectricity, Curie temperatures above room temperature, and orientationally disordered plastic mesophases, in addition to supporting low-temperature synthesis. Here, we present the results of a screening study of the Cambridge Structural Database (CSD) leading to the discovery of 54 candidate ferroelectrics, including molecular crystals and molecular salts, many of which are likely to host plastic mesophases, along with 16 previously reported ferroelectrics. With over 1.2 million entries in the CSD, the screening procedure involved many steps, including considerations of space group, and hydrogen bonding pattern. Out of the candidate systems, many of them were identified as plastic mesophases due to their resemblance to highly symmetric close-packed crystal structures. The electronic band gaps were predicted by using density functional theory. Among the candidate ferroelectrics, spontaneous polarization greater than 10 $\mu C/cm^2$, with five of them being reported at room temperature.

54 candidates

1. INTRODUCTION

Ferroelectric molecular crystals are potential alternatives to conventional inorganic ferroelectrics.[1−4] These materials can be synthesized with low-cost and low-energy methods such as coprecipitation, slow evaporation, spin coating, and 3D printing.[5−11] Furthermore, the abundance of available molecular species allows for tailoring of the properties of molecular crystals while eliminating the need for scarce or toxic elements. The ferroelectric molecular crystals can be formed solely through van der Waals and/or hydrogen bonding, and such systems are referred to as *molecular* crystals. Alternatively, they can also contain an additional ionic component originating from charged molecular species, which are termed *ionic molecular* crystals.

Of particular interest are (ionic) molecular crystals consisting of globular molecules. Their globular shape reduces the steric hindrance of rotational motion, making polarization switching through rotation possible. In some materials, rotational switching can also occur in combination with displacive mechanisms.[12] These materials can show rapid switching with frequencies of up to 263 kHz,[13−15] Moreover, molecular rotations in response to an applied electric field or mechanical force can result in large shear piezoelectric responses.[16] These types of molecular crystals can also crystallize in pseudo high-symmetry structures, thus allowing multiaxial polarization, with as many as 24 equivalent axes.[17−20] Multiaxial polarization allows the spontaneous polarization to be aligned in a desired direction in polycrystalline materials

Ferroelectric (ionic) molecular

globular molecules, can also ho

mesophases.[3,28,30−34] These phas

"plastic", which can arise fro

interactions, and increased sym

slip planes.[34] This subgroup of

plastic crystals, and the materials

contrast to the rigidity and br

molecular crystals and inorganic

Ferroelectric plastic crystals

diagram,[36,37] with multiple comp

the existence of plastic ferroel

exemplified by quinuclidinium p

ferroelectric phase with partial o

advantage of these materials is

coercive fields. For instance, valu

have been reported for

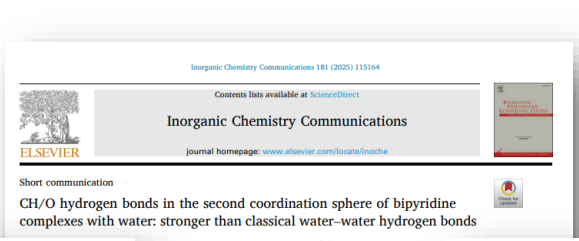
azabicyclo[2.2.1]heptanium perr

perrhenate.[5,38] These values a

$BaTiO_3$, a well-known inorganic f

crystals can also exhibit high Cu

Received: June 12, 2023
Revised: October 28, 2023
Accepted: October 30, 2023
Published: November 14, 2023



Inorganic Chemistry Communications 181 (2025) 115164

Contents lists available at ScienceDirect

Inorganic Chemistry Communications

journal homepage: www.elsevier.com/locate/inoche

ELSEVIER

Short communication

CH/O hydrogen bonds in the second coordination sphere of bipyridine complexes with water: stronger than classical water–water hydrogen bonds

Jelena M. Živković[a], Snežana D. Zarić[b,*]

[a] Faculty of Chemistry, Studentski trg 12-16, 11000 Belgrade, Serbia
[b] Faculty of Chemistry, Studentski trg 12-16, 11000 Belgrade, Serbia

ABSTRACT

Although CH/O interactions are generally recognized as weak noncovalent interactions, they can form strong interactions; the calculated interaction energies for CH/O interactions in copper(II), and palladium(II) complexes with a water molecule range from [illegible]. The interaction energies are influenced by the coordination number of the central metal atom, interacting hydrogen atom. As the coordination number decreases, the CH/O interactions become bifurcated interactions are stronger than linear ones. The strongest calculated CH/O interaction in a square-planar palladium(II) complex. The calculated energies and bipy complexes are consistent with the data observed by analyzing crystal Structural Database. These findings suggest that such interactions could be [illegible] terials and functional molecular systems where noncovalent bonding plays a [illegible]

of the nonconventional hydrogen bonds [illegible] interaction [1−3]. The analysis of the [illegible] Database (CSD) showed that CH/O [illegible], hence they have to be considered [illegible] Waals interactions [4]. In spite of being [illegible] cooperative with CH/pi interactions [illegible] structuring [5]. Various experimental [illegible] [13−16] were used to study CH/O

[illegible] molecules, aromatic molecules form the [illegible] hydrogen atoms in aromatic mole- [illegible] electrostatic potential. Also, in aromatic [illegible] can be bifurcated and even stronger. In [illegible] of linear CH/O interaction is −1.28 [illegible] CH/O interaction is −1.38 kcal/mol [illegible] in organic molecules strengthens the [illegible] forms stronger interactions than [illegible] pyridine/water interaction is −1.97 [illegible] bifurcated interactions is −2.30 kcal/mol [18].

Substituents on the aromatic ring as [illegible] teractions influence the interaction energy [illegible] CH/O interactions were observed [illegible] fluorobenzene-acetone complex, with a [illegible] kcal/mol [21] and for bifurcated interactions [illegible] theophylline complex, with an interaction [illegible] [22].

Also, other noncovalent interactions [illegible] contribute to the strength of the CH/O interaction [illegible] we observed an increase in the strength [illegible] when pyridine nitrogen is involved in a [illegible] pyridine molecule forms a bifurcated CH [illegible] of −2.30 kcal/mol, while the pyridine/water [illegible] forms a classical hydrogen bond with water [illegible] interaction with another water molecule [illegible] mol [18].

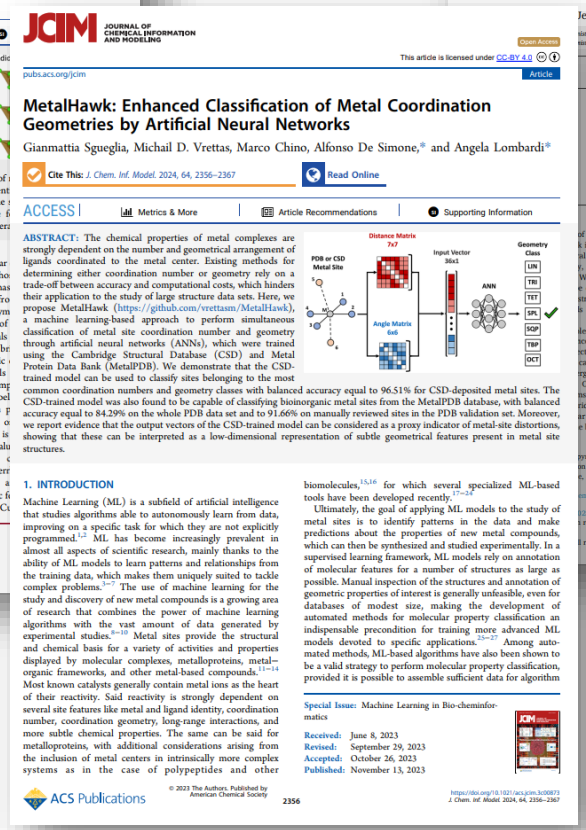
In our previous studies we showed that [illegible] stronger classical hydrogen bonds than [illegible]

pyridine; CH/O, Carbon–Hydrogen/Oxygen (noncovalent hydrogen bond); CSD, Cambridge Structural Database; [illegible] on Number; ESP, Electrostatic Potential; CBS, Complete Basis Set; DLPNO-CCSD(T), Domain [illegible], and perturbative Triple excitations; TPSS-D3BJ/def2-TZVP, Specific functional and basis set [illegible]

[illegible] (D.B. Ninković), szaric@chem.bg.ac.rs (S.D. Zarić).

[illegible] revised form 10 July 2025; Accepted 21 July 2025

[illegible] rights are reserved, including those for text and data mining, AI training, and similar technologies.



JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim

Open Access

This article is licensed under CC-BY 4.0

Article

MetalHawk: Enhanced Classification of Metal Coordination Geometries by Artificial Neural Networks

Gianmattia Sgueglia, Michail D. Vrettas, Marco Chino, Alfonso De Simone,* and Angela Lombardi*

Cite This: *J. Chem. Inf. Model.* 2024, 64, 2356−2367

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The chemical properties of metal complexes are strongly dependent on the number and geometrical arrangement of ligands coordinated to the metal center. Existing methods for determining either coordination number or geometry rely on a trade-off between accuracy and computational costs, which hinders their application to the study of large structure data sets. Here, we propose MetalHawk (https://github.com/vrettasm/MetalHawk), a machine learning-based approach to perform simultaneous classification of metal site coordination number and geometry through artificial neural networks (ANNs), which were trained using the Cambridge Structural Database (CSD) and Metal Protein Data Bank (MetalPDB). We demonstrate that the CSD-trained model can be used to classify sites belonging to the most common coordination numbers and geometry classes with balanced accuracy equal to 96.51% for CSD-deposited metal sites. The CSD-trained model was also found to be capable of classifying bioinorganic metal sites from the MetalPDB database, with balanced accuracy equal to 84.29% on the whole PDB data set and to 91.66% on manually reviewed sites in the PDB validation set. Moreover, we report evidence that the output vectors of the CSD-trained model can be considered as a proxy indicator of metal-site distortions, showing that these can be interpreted as a low-dimensional representation of subtle geometrical features present in metal site structures.

1. INTRODUCTION

Machine Learning (ML) is a subfield of artificial intelligence that studies algorithms able to autonomously learn from data, improving on a specific task for which they are not explicitly programmed.[1,2] ML has become increasingly prevalent in almost all aspects of scientific research, mainly thanks to the ability of ML models to learn patterns and relationships from the training data, which makes them uniquely suited to tackle complex problems.[3−7] The use of machine learning for the study and discovery of new metal compounds is a growing area of research that combines the power of machine learning algorithms with the vast amount of data generated by experimental studies.[8−10] Metal sites provide the structural and chemical basis for a variety of activities and properties displayed by molecular complexes, metalloproteins, metal−organic frameworks, and other metal-based compounds.[11−14] Most known catalysts generally contain metal ions as the heart of their reactivity. Said reactivity is strongly dependent on several site features like metal and ligand identity, coordination number, coordination geometry, long-range interactions, and more subtle chemical properties. The same can be said for metalloproteins, with additional considerations arising from the inclusion of metal centers in intrinsically more complex systems as in the case of polypeptides and other biomolecules,[15,16] for which several specialized ML-based tools have been developed recently.[17−24]

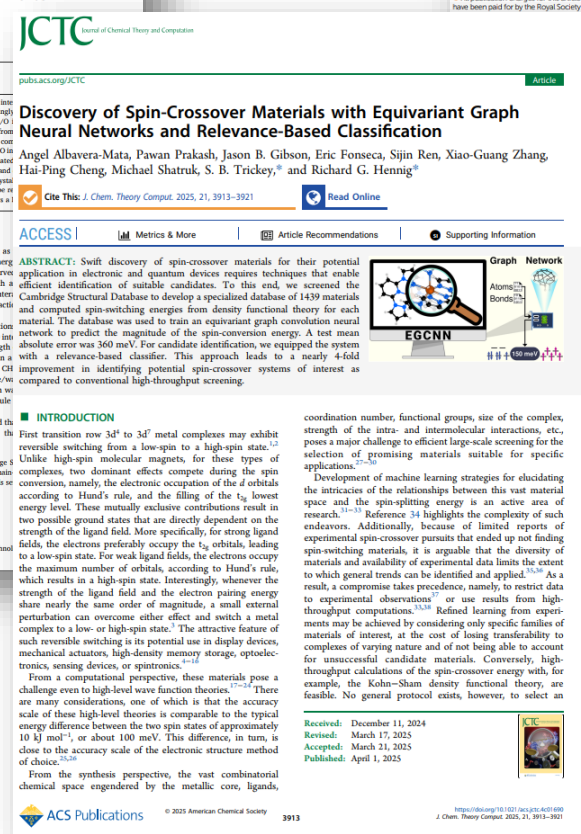
Ultimately, the goal of applying ML models to the study of metal sites is to identify patterns in the data and make predictions about the properties of new metal compounds, which can then be synthesized and studied experimentally. In a supervised learning framework, ML models rely on annotation of molecular features for a number of structures as large as possible. Manual inspection of the structures and annotation of geometric properties of interest is generally unfeasible, even for databases of modest size, making the development of automated methods for molecular property classification an indispensable precondition for training more advanced ML models devoted to specific applications.[25−27] Among automated methods, ML-based algorithms have also been shown to be a valid strategy to perform molecular property classification, provided it is possible to assemble sufficient data for algorithm

Special Issue: Machine Learning in Bio-cheminformatics

Received: June 8, 2023
Revised: September 29, 2023
Accepted: October 26, 2023
Published: November 13, 2023

ACS Publications

https://doi.org/10.1021/acs.jcim.3c00873
J. Chem. Inf. Model. 2024, 64, 2356−2367



JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC

Article

Discovery of Spin-Crossover Materials with Equivariant Graph Neural Networks and Relevance-Based Classification

Angel Albavera-Mata, Pawan Prakash, Jason B. Gibson, Eric Fonseca, Sijin Ren, Xiao-Guang Zhang, Hai-Ping Cheng, Michael Shatruk, S. B. Trickey,* and Richard G. Hennig*

Cite This: *J. Chem. Theory Comput.* 2025, 21, 3913−3921

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Swift discovery of spin-crossover materials for their potential application in electronic and quantum devices requires techniques that enable efficient identification of suitable candidates. To this end, we screened the Cambridge Structural Database to develop a specialized database of 1439 materials and computed spin-switching energies from density functional theory for each material. The database was used to train an equivariant graph convolution neural network to predict the magnitude of the spin-conversion energy. A test mean absolute error was 360 meV. For candidate identification, we equipped the system with a relevance-based classifier. This approach leads to a nearly 4-fold improvement in identifying potential spin-crossover systems of interest as compared to conventional high-throughput screening.

INTRODUCTION

First transition row $3d^4$ to $3d^7$ metal complexes may exhibit reversible switching from a low-spin to a high-spin state.[1,2] Unlike high-spin molecular magnets, for these types of complexes, two dominant effects compete during the spin conversion, namely, the electronic occupation of the d orbitals according to Hund's rule, and the filling of the t_{2g} lowest energy level. These mutually exclusive contributions result in two possible ground states that are directly dependent on the strength of the ligand field. More specifically, for strong ligand fields, the electrons preferably occupy the t_{2g} orbitals, leading to a low-spin state. For weak ligand fields, the electrons occupy the maximum number of orbitals, according to Hund's rule, which results in a high-spin state. Interestingly, whenever the strength of the ligand field and the electron pairing energy share nearly the same order of magnitude, a small external perturbation can overcome either effect and switch a metal complex to a low- or high-spin state.[3] The attractive feature of such reversible switching is its potential use in display devices, mechanical actuators, high-density memory storage, optoelectronics, sensing devices, or spintronics.[4−14]

From a computational perspective, these materials pose a challenge even to high-level wave function theories.[15−18] There are many considerations, one of which is that the accuracy scale of these high-level theories is comparable to the typical energy difference between the two spin states of approximately 10 kJ mol^{-1}, or about 100 meV. This difference, in turn, is close to the accuracy scale of the electronic structure method of choice.[19,20]

From the synthesis perspective, the vast combinatorial chemical space engendered by the metallic core, ligands, coordination number, functional groups, size of the complex, strength of the intra- and intermolecular interactions, etc., poses a major challenge to efficient large-scale screening for the selection of promising materials suitable for specific applications.[21−30]

Development of machine learning strategies for elucidating the intricacies of the relationships between this vast material space and the spin-splitting energy is an active area of research.[31−33] Reference 34 highlights the complexity of such endeavors. Additionally, because of limited reports of experimental spin-crossover pursuits that ended up not finding spin-switching materials, it is arguable that the diversity of materials and availability of experimental data limits the extent to which general trends can be identified and applied.[35,36] As a result, a compromise takes precedence, namely, to restrict data to experimental observations[37] or use results from high-throughput computations.[33,38] Refined learning from experiments may be achieved by considering only specific families of materials of interest, at the cost of losing transferability to complexes of varying nature and of not being able to account for unsuccessful candidate materials. Conversely, high-throughput calculations of the spin-crossover energy with, for example, the Kohn−Sham density functional theory, are feasible. No general protocol exists, however, to select an

Received: December 11, 2024
Revised: March 17, 2025
Accepted: March 21, 2025
Published: April 1, 2025

ACS Publications

https://doi.org/10.1021/acs.jctc.4c01600
J. Chem. Theory Comput. 2025, 21, 3913−3921

Chemical Science

ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Chem. Sci.*, 2022, 13, 13507

All publication charges for this article have been paid for by the Royal Society of Chemistry

Identifying porous cage subsets in the Cambridge Structural Database using topological data analysis†

Aurelia Li, Rocio Bueno-Perez and David Fairen-Jimenez *

As rationally designable materials, the variety and number of synthesised metal–organic cages (MOCs) and organic cages (OCs) are expected to grow in the Cambridge Structural Database (CSD). In this regard, two of the most important questions are, which structures are already present in the CSD and how can they be identified? Here, we present a cage mining methodology based on topological data analysis and a combination of supervised and unsupervised learning that led to the derivation of – to the best of our knowledge – the first and only MOC dataset of 1839 structures and the largest experimental OC dataset of 7736 cages, as of March 2022. We illustrate the use of such datasets with a high-throughput screening of MOCs and OCs for xenon/krypton separation, important gases in multiple industries, including healthcare.

of microporous materials, metal- [illegible]anic cages (OCs) are of particular [illegible] more established metal–organic [illegible]ent organic frameworks (COFs), [illegible]e structures constructed from [illegible]cks, MOCs and OCs are discrete [illegible]cage-like shape. When packed, [illegible]rough non-covalent interactions [illegible] Therefore, there are two types of [illegible] with cages: (i) the molecule's [illegible] porosity and (ii) the interstitial [illegible]*extrinsic* porosity. The combined [illegible]justify the growing research for [illegible]ular[4−7] or gas separations,[1,2,8,9] [illegible] molecular sensing,[3,10,11] and as [illegible]OFs, COFs, MOCs and OCs all [illegible]are all tuneable porous materials, [illegible]nd OCs means they are process- [illegible] Giri *et al.* reported the prepara- [illegible]ssolving custom-designed rigid [illegible]t (15-crown-5) that cannot enter [illegible] soluble in 15-crown-5 (6.7 μmol [illegible]ges increased the solubility by [illegible]). Cages can also be incorporated

into polymers[46] and form mixed-matrix membranes.[47] Choosing between a MOC and an OC depends on the requirements of the application. For example, the open metal sites available in MOCs offer more tuning possibilities for catalysis, whereas OCs provide an overall lighter material.

Like in MOFs and COFs, the modularity of MOCs and OCs can lead to a large space of possible structures and has thus attracted the attention of computational researchers. The two types of porosities – intrinsic and extrinsic – add yet another customisable dimension. Evans *et al.* estimated that using only small organic molecules as building blocks for cage-based porous molecular materials, there could be 10^{107} potential candidates.[48] Inspired by the success of high-throughput screenings (HTS) on extended porous materials, several research groups started to apply the same data mining methods to organic molecular materials, whether their porosity is intrinsic and/or extrinsic. McKeown *et al.* carried out a targeted structure search in the Cambridge Structural Database (CSD) to identify promising organic microporous crystals for nitrogen and hydrogen adsorption.[49] In particular, the authors were looking for structures that might possess enhanced microporosity compared to existing examples of microporous crystals. Therefore, to narrow their search in the database, they looked for structures (1) with densities lower than 0.9 g cm^{-3}, as the lowest density of any known microporous organic crystal was 0.96 g cm^{-3} for *p-tert*-butylcalix[4]dihydroquinone after water removal, (2) containing mostly aromatic rings as these play an essential role in the structures' stability (or shape persistence) and (3) with pore diameters smaller than 10 Å, as it was shown it ensures stronger gas adsorption for hydrogen storage. Following this data sieving and after the elimination of additional structures with questionable data quality, 23 organic and

by the Royal Society of Chemistry

CCDC

A wealth of structures, stories and possibilities

CSD one million special issue

- 33 articles from industry and academia

1 million structures.
stories.
possibilities.

Cambridge Structural Database, CSD

CCDC
advancing structural science

Legacy and Future Impact of the Cambridge Structural Database: A Tribute to Dr. Olga Kennard - special issue

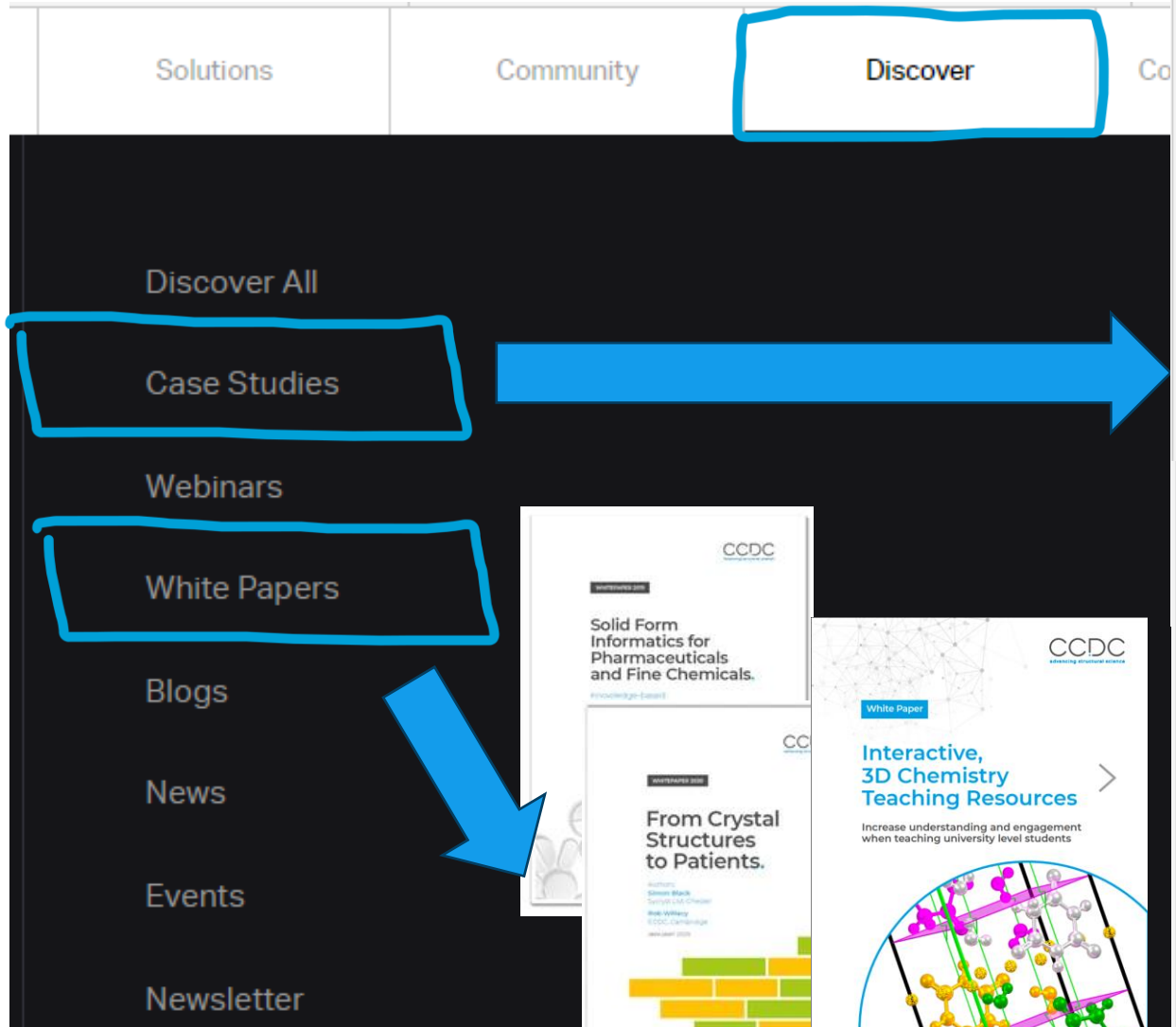
- 29 articles from industry and academia

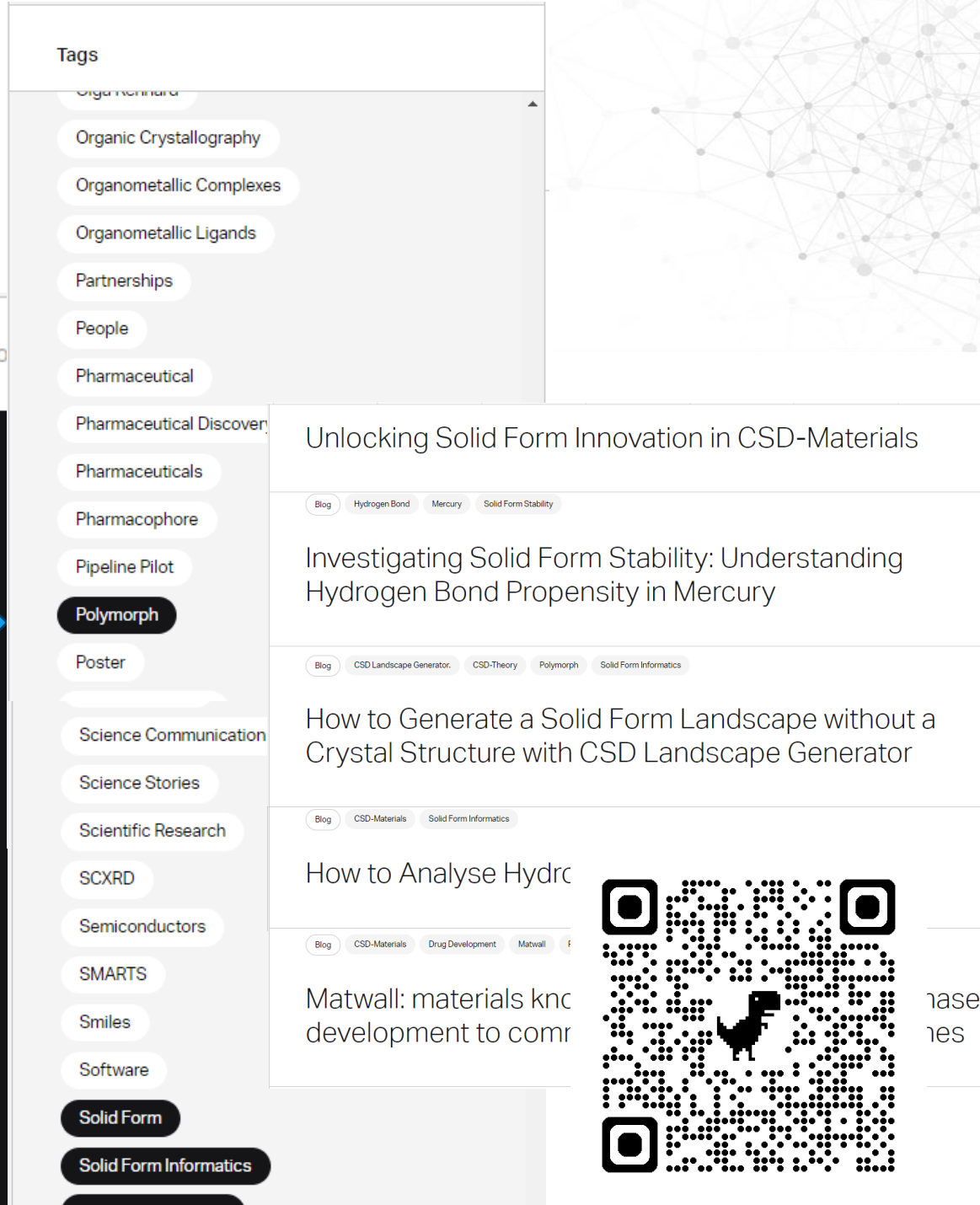
CRYSTAL GROWTH & DESIGN

The CSD - A wealth of knowledge gained from a million structures. *CrystEngComm*, 2020,22, 7131-77502
The CSD. C.R.Groom, I.J.Bruno, M.P.Lightfoot and S.C.Ward, *Acta Cryst.* (2016). B72, 171-179 DOI: 10.1107/S2052520616003954

CCDC

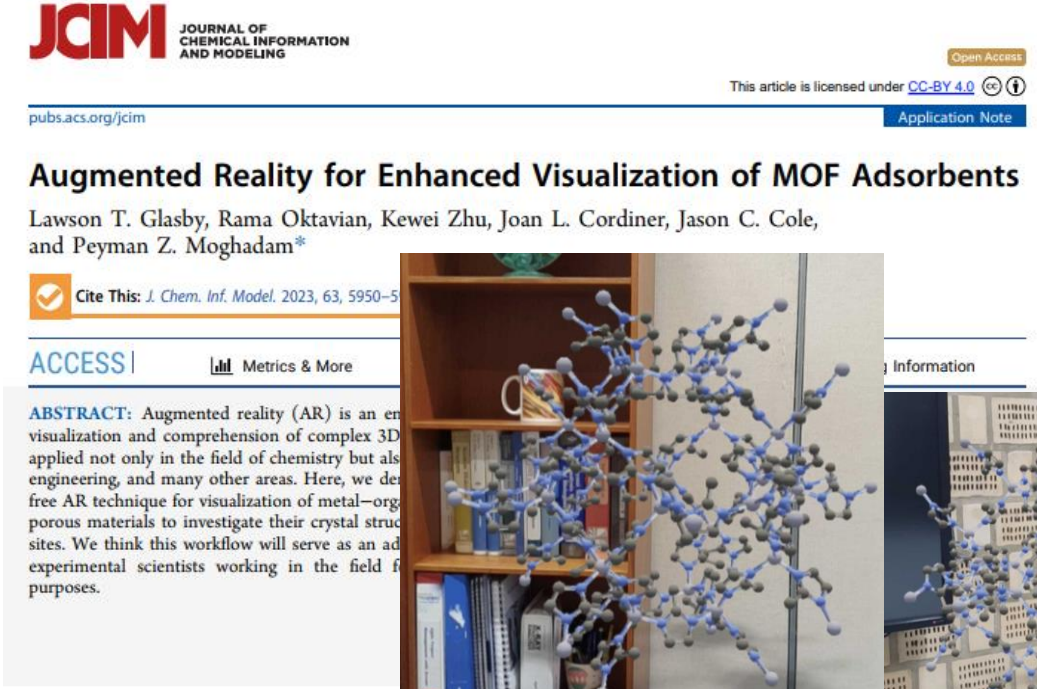
More case studies

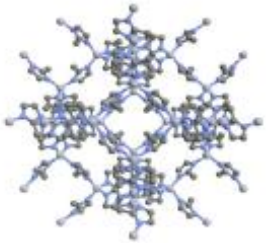




Augmented reality?

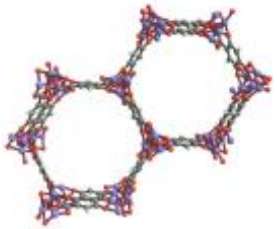
DigiMOF – a database of properties for MOFs found in the CSD, data mined from journal articles using ChemDataExtractor



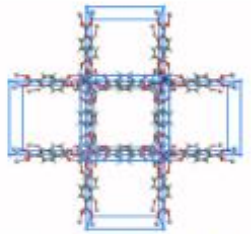


ZIF-8 (FAWCEN)

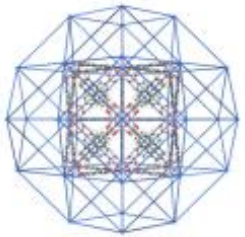
MIL-53 (ASOHUL)



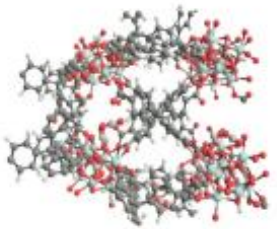
CPO-27-Co (MOHGEY)



IRMOF-1 (SAHYIK)



UiO-67(WIZMAV)



MOF-812 (BOHWOM)

Lawson T. Glasby, Rama Oktavian, Kewei Zhu, Joan L. Cordiner, Jason C. Cole, Peyman Z. Moghadam, J. Chem. Inf. Model. 2023, 63, 19, 5950–5955

CCDC

How to cite the tools seen today

- **Mercury**
  Mercury 4.0: from visualization to analysis, design and prediction
  C. F. Macrae, I. Sovago, S. J. Cottrell, P. T. A. Galek, P. McCabe, E. Pidcock, M. Platings, G. P. Shields, J. S. Stevens, M. Towler and P. A. Wood, *J. Appl. Cryst.*, **53**, 226-235, 2020
  DOI: 10.1107/S1600576719014092
- **CSD**
  The Cambridge Structural Database
  C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Cryst.*(2016). B72, 171-179
  DOI: 10.1107/S2052520616003954
- **MOF subset**
  P. Z. Moghadam, A. Li, S. B. Wiggin, A. Tao, A. G. P. Maloney, P. A. Wood, S. C. Ward, and D. Fairen-Jimenez, *Chemistry of Materials* 2017 *29* (7), 2618-2625
  DOI: 10.1021/acs.chemmater.7b00441

More reference papers for CSD Software

CCDC

What we have learnt

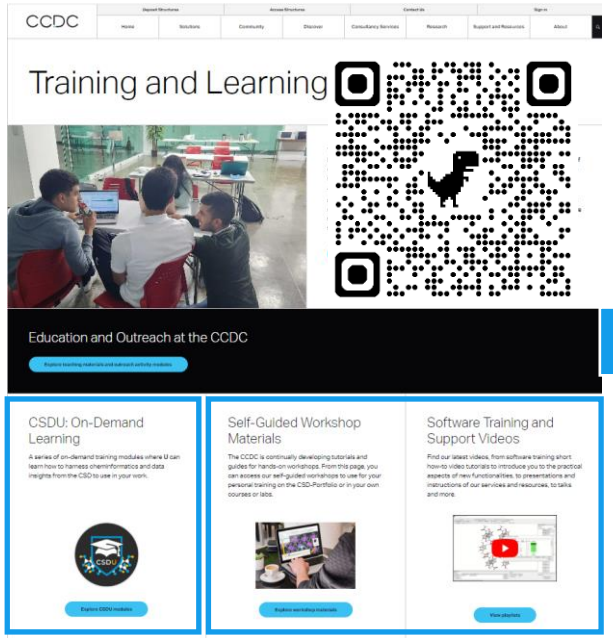
- ✓Become familiar with the **CSD**.
- ✓Which **subsets** of the CSD are available, how to access them in ConQuest, Mercury and the CSD Python API, and how they can help you in your research.
- ✓How to **visualise** porous materials effectively in **Mercury**.
- ✓How to **analyse** pore space in your structures using **Pore Analyser**.
- ✓About applications of the **disorder visualization feature**, including recent advances and integrations with other functionality available in Mercury.
- ✓And more Mercury **tips and tricks**!

CCDC

How many points did you get?

Want to explore more?

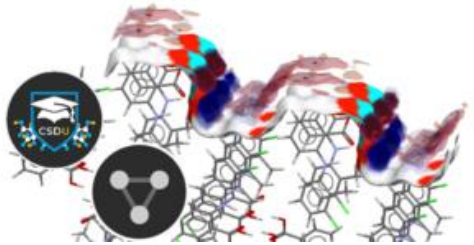
Training and learning



- Covering use cases with guided examples.
- Including examples from today as well as other CSD Software.

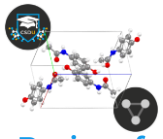
CSDU

Learn on demand with free online training courses

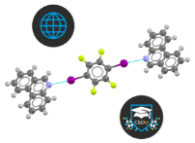


- Free
- Completion certificate upon passing a final test.
- In-depth review of Mercury and more software and functionality.

CSD Python API

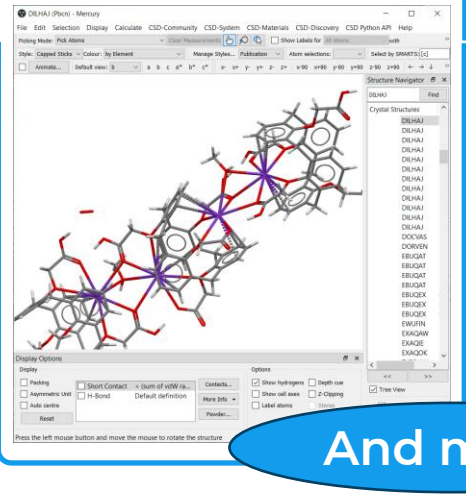


Basics of Mercury



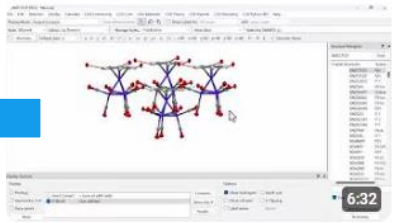
WebCSD

Self-guided workshops



Videos

YouTube channel



How to: Explore molecules and create eye-catching graphics with Mercury

How to: Build and combine searches in ConQuest

And more!

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

Want to explore more?

CCDC Virtual Workshops

- **21st October –** How to Use **Particle Informatics Tools** to Connect Crystal Structure to Particle Behaviour
- **4th November –** How to Use the Docking Software **GOLD** to Perform **Virtual Screening** Simulations

CCDC Webinars

- **13th November** - Using general purpose **force fields** in solid form studies
- **20th November, Panel** – **Frontiers in MOFs**: Creating Real-World Solutions with MOFs

Check our Events online for more virtual and in person events.