

# Searching and analysing metal-organic structures

CCDC Virtual Workshop Summer 2021 – Session 3

Ilaria Gimondi, Suzanna Ward, Yinka Olatunji-Ojo, Seth Wiggin, Eva Myers

July 2021



CCDC  
advancing structural science

# Learning outcomes

Advanced tips and tricks to search and analyse metal-organic structures by:

- Learning tactics to search metal-organic structures in ConQuest.
- Familiarity of metal-organic subsets in ConQuest.
- Learning how to export and analyse metal-organic hitlists in Mercury.

# The CSD and the ICSD

Connecting data resources, increase discoverability and simplify deposition of organic, metal-organic and inorganic structural data.

- Established, trusted databases
- Value quality and high levels of curation
- Comprehensive
- Shared communities

## Alliance Reshapes Crystallography Data Access

— March 27, 2017

Structural chemistry's trusted crystallographic database providers join forces to provide single point access to all of the world's small molecule crystal data.

Cambridge, United Kingdom, and Karlsruhe, Germany, March 27, 2017.

The Cambridge Crystallographic Data Centre (The CCDC) and FIZ Karlsruhe – Leibniz Institute for Information Infrastructure (FIZ Karlsruhe) today announce the start of a joint development project that will deliver for the first time shared deposition and access services for crystallographic data across all domains of chemistry – including organic, inorganic and metal-organic structures. The resulting capability - to search over one million crystallographic structures and to deposit data for the CCDC's Cambridge Structural Database (CSD) and FIZ Karlsruhe's Inorganic Crystal Structure Database (ICSD) and their underlying CIF depots at a single source - will benefit researchers and educators

## Free, unified deposition and access of crystal structure data

— July 12, 2018

The Cambridge Crystallographic Data Centre (CCDC) and FIZ Karlsruhe – Leibniz Institute for Information Infrastructure (FIZ Karlsruhe) today announced the launch of their joint deposition and access services for crystallographic data across all chemistry. These services will enable researchers to share data through a single deposition portal and explore all chemical structures for free worldwide.

FIZ Karlsruhe supports the community's need for a reliable infrastructure for research data from Sabine Brünger-Weilandt, CEO from FIZ Karlsruhe. "Providing freely available research data for all is key to Advancing Science. The announcement of the cooperation between CCDC and FIZ was enthusiastically received by the community. We are convinced that we can meet the high new joint depot."

for the CCDC is equally excited about the impact of this launch to researchers worldwide: "All the other people I know, whether they admit it or not, wish that all of the information that they require was in a single location. They are searching for a 'magic bullet' that will hit exactly what they want; they want to be able to use a single search to get all of their information needs. By unifying the deposition and access of organic, metal-organic and inorganic structures we get a little closer to that magic bullet, at least in the area of crystallography, which makes life much easier," says Judith Curran, Chair of Trustees for the CCDC and Head of the Department of Chemistry at the University of Pennsylvania.

Chemistry have meant that the distinctions between inorganic and organic structures have become blurred. Research to design new batteries, gas storage systems, zeolites, catalysts, magnets, and materials has led to the desire from researchers for more integrated databases, has been the driving force behind these joint services.

CCDC FIZ Karlsruhe Leibniz Institute for Information Infrastructure Access Structures

Simple Search Structure Search Unit Cell Search Formula Search

Entry search

Welcome to Access Structures, the CCDC's and FIZ Karlsruhe's free service to view and retrieve structures. Please use one or more of the boxes to find entries. If you enter details in more than one field the search will try to find records containing all the terms entered. More information and search help

More advanced search functionality and additional curated data for the Cambridge Structural Database (CSD) and the Inorganic Crystal Structure Database (ICSD) is available through the CSD-System and ICSD, respectively. Click here for more information.

Identifiers(s) CCDC Number(s), CSD Number(s), CSD Refcode(s) or ICSD Number(s)

Compound name e.g. sulfadiazine

DOI A single publication DOI, CSD DOI or ICSD DOI

Authors e.g. F.H.Allen

Journal e.g. Journal of the American Chemical Society

Publication details Year Volume Page

Database to search  Entire published collection  CSD  ICSD

Search

CCDC FIZ Karlsruhe Leibniz Institute for Information Infrastructure ICSD Entry: 243366

Simple Search Structure Search Unit Cell Search Formula Search

Your query was: Identifier(s): AHUOZO AHUDIT AHOQIA 243366 254369 254371 and the search returned 10 records.

Back to Search List Modify Search New Search

Results

Database	Identifier	Deposition Number
<input checked="" type="checkbox"/>	ICSD 243366	1577704
<input type="checkbox"/>	ICSD 254369	1565100
<input type="checkbox"/>	ICSD 254371	1565103
<input type="checkbox"/>	AHOQIA	1981609
<input type="checkbox"/>	AHUDIT	1946311
<input type="checkbox"/>	AHUOZO	1946312
<input type="checkbox"/>	BINIS	243368
<input type="checkbox"/>	FEDFEA	254369
<input type="checkbox"/>	FEDFOK	254371
<input type="checkbox"/>	ICSD 243366	1791301

Download JSmol

3D viewer

Chemical diagram

More information available from the ICSD <http://icsd.products.fiz-karlsruhe.de/>

View group symbols key

Additional details

Deposition Number 1577704

Data Citation Hannes Dierkes, Jan van Leusen, Dimitri Bogdanovski, Richard Dronskowski CCDC 1577704: Experimental Crystal Structure Determination, 2017, DOI: 10.5517/ccdc.csd.cc1pyqp1

Deposited on 02/10/2017

Associated publications

Hannes Dierkes, Jan van Leusen, Dimitri Bogdanovski, Richard Dronskowski, *Inorganic Chemistry*, 2017, 56, 1045, DOI: 10.1021/acs.inorgchem.6b02816

# CSD criteria

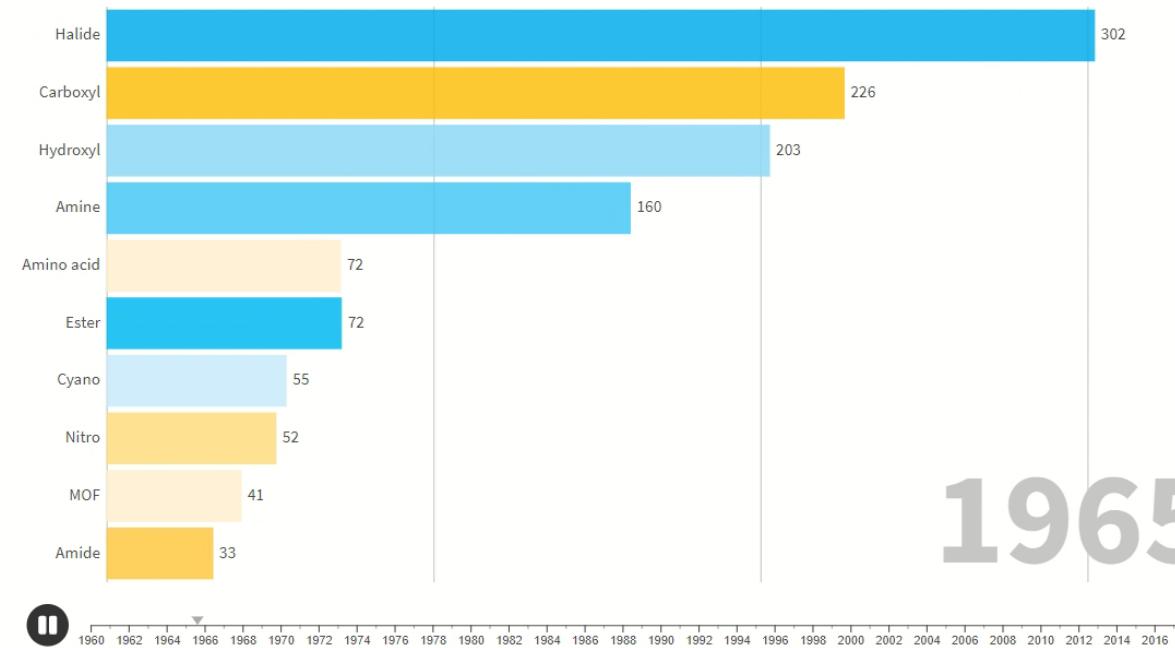
Organic and metal-organic experimental crystal structures

- Types of experimental determination include:
  - a [single crystal study](#) - where cell parameters are reported, or
  - a [powder study](#), where cell parameters, [atomic coordinates](#) and constrained refinement (e.g. Rietveld) are reported
  - from [x-ray](#), [neutron](#) or [electron](#) diffraction
- Types of chemistry include:
  - Metal and metal-organic compounds
  - peptides and saccharides of up to 24 residues
  - mono-, di- and tri-nucleotides
  - metal carbonyls
  - boron compounds containing one or more B-H or B-OH bond and borazines
  - ring compounds containing any two of the following elements: N, P, S, Se and Te

# Exploring the CSD

## Chemistry in the CSD

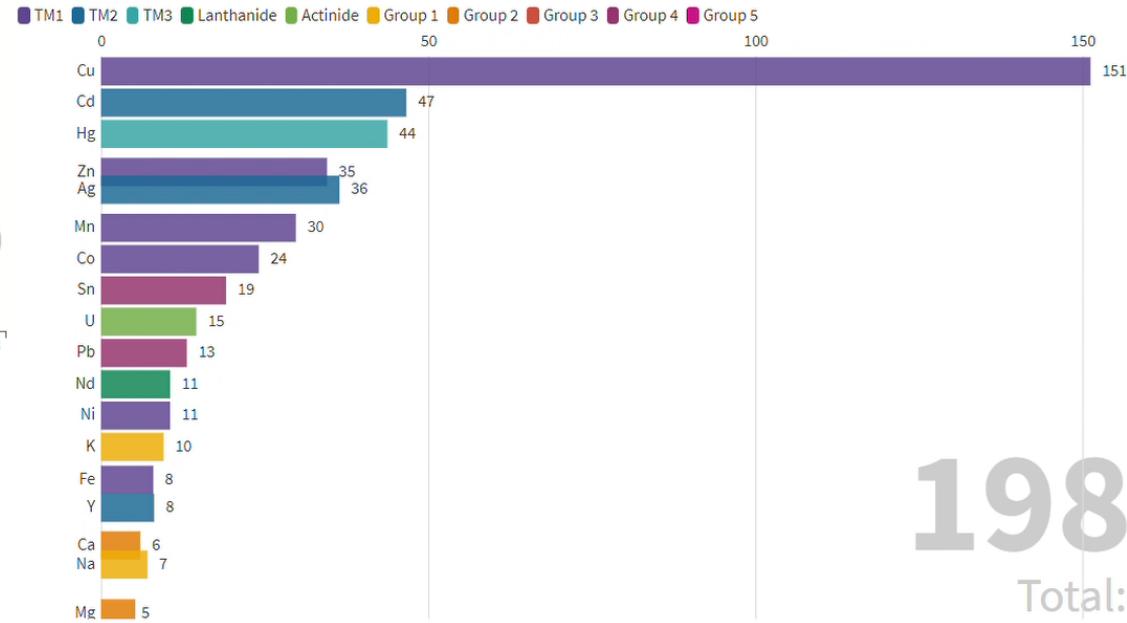
Number of structures containing certain chemical groups



Images and graphics created using Flourish

## Metals in the CSD MOF subset

The number of structures containing a particular metal atom



# What is a MOF?

*"A metal-organic framework, abbreviated to MOF, is a coordination network with organic ligands containing potential voids."*

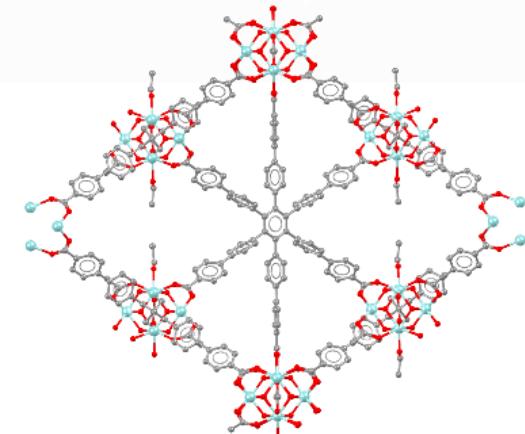
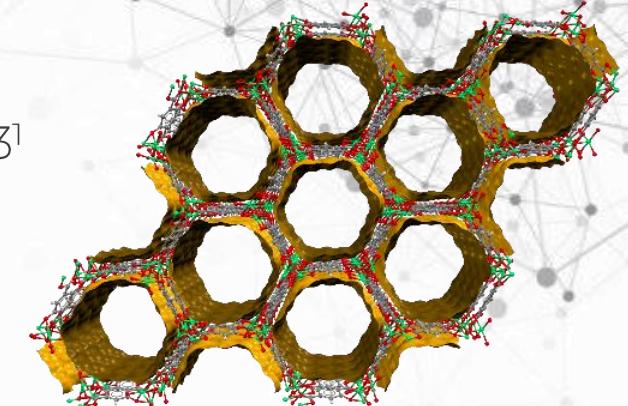
IUPAC Recommendations 2013<sup>1</sup>

Interim 2012 report<sup>2</sup> noted:

- Questionnaire gave no clear consensus if MOFs should include 1D, 2D or 3D networks
- Almost 25% of respondents believed MOFs must have proven porosity
- No IUPAC definition of organic – include oxalates, cyanides etc.?

*"Finally, we should perhaps remind ourselves that what matters, or should matter, in the real world are the properties of our materials not what we chose to call them. As Juliet says: 'A rose by any other name would smell as sweet'"*

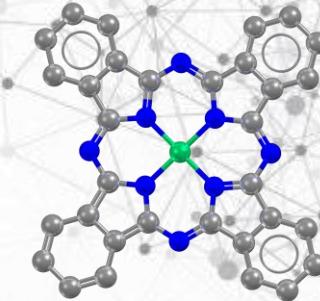
IZA Commission on Metal  
Organic Frameworks<sup>3</sup>



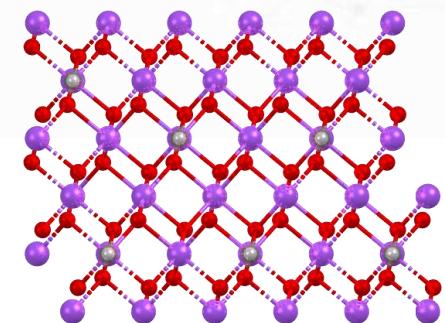
1. S. R. Batten et al. *Pure Appl. Chem.*, 2013, 85, 1715–1724 10.1351/PAC-REC-12-11-20; 2. S.R.Batten et al. *CrystEngComm*, 2012, 14, 3001-3004 10.1039/C2CE06488J; 3. <http://www.iza-online.org/MOF/>

# MOF and metal-organic trivia

- 1<sup>st</sup> metal-organic structure with coords was published in 1938 by Monteath Robertson
- 1<sup>st</sup> MOF in our MOF subset was published in 1940
- >40,000 authors have contributed to CSD MOF structures
- Most prolific author of MOF structures is Pascal D.C.Dietzel
  - Time-resolved synchrotron X-ray diffraction measurements leads to many individual data collections, currently over 1,900!



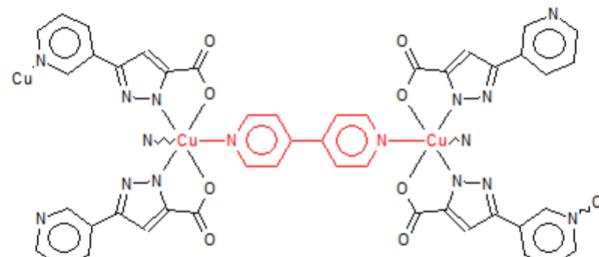
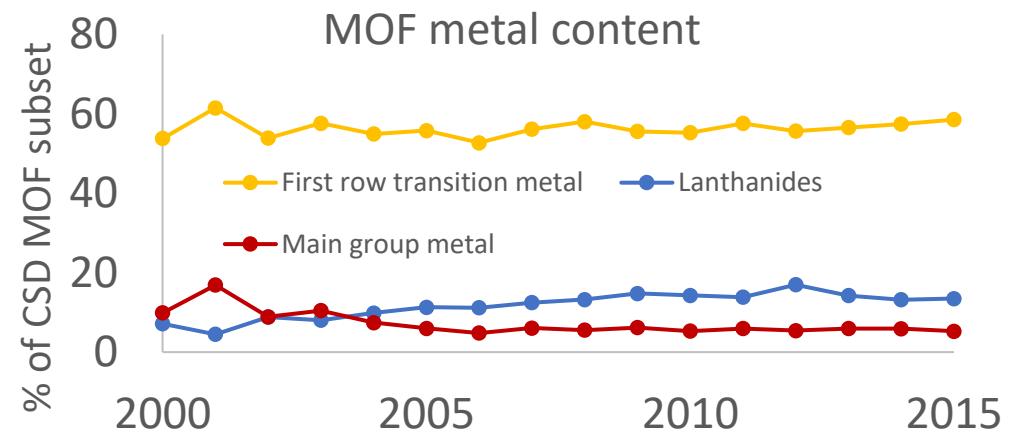
NIPHTC  
1st metal-organic  
structure with 3D coords



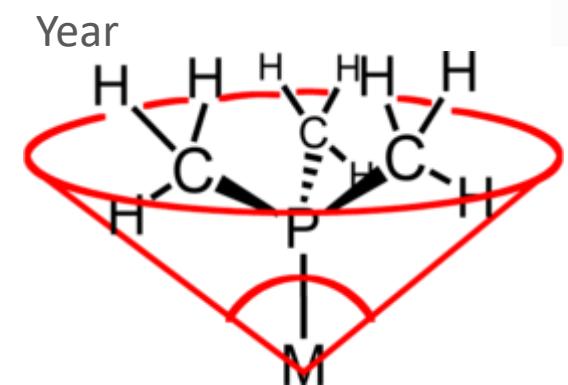
NAFORM  
1st MOF in CSD subset

# Using the CSD to explore metal-coordinations

- Ligand coordination modes
- Metal coordination numbers
- Bond lengths
- Symmetry and shape
- Spin states
- Ligand cone angles
- Polynuclear complexes
- Magnetism
- Software parametrisation

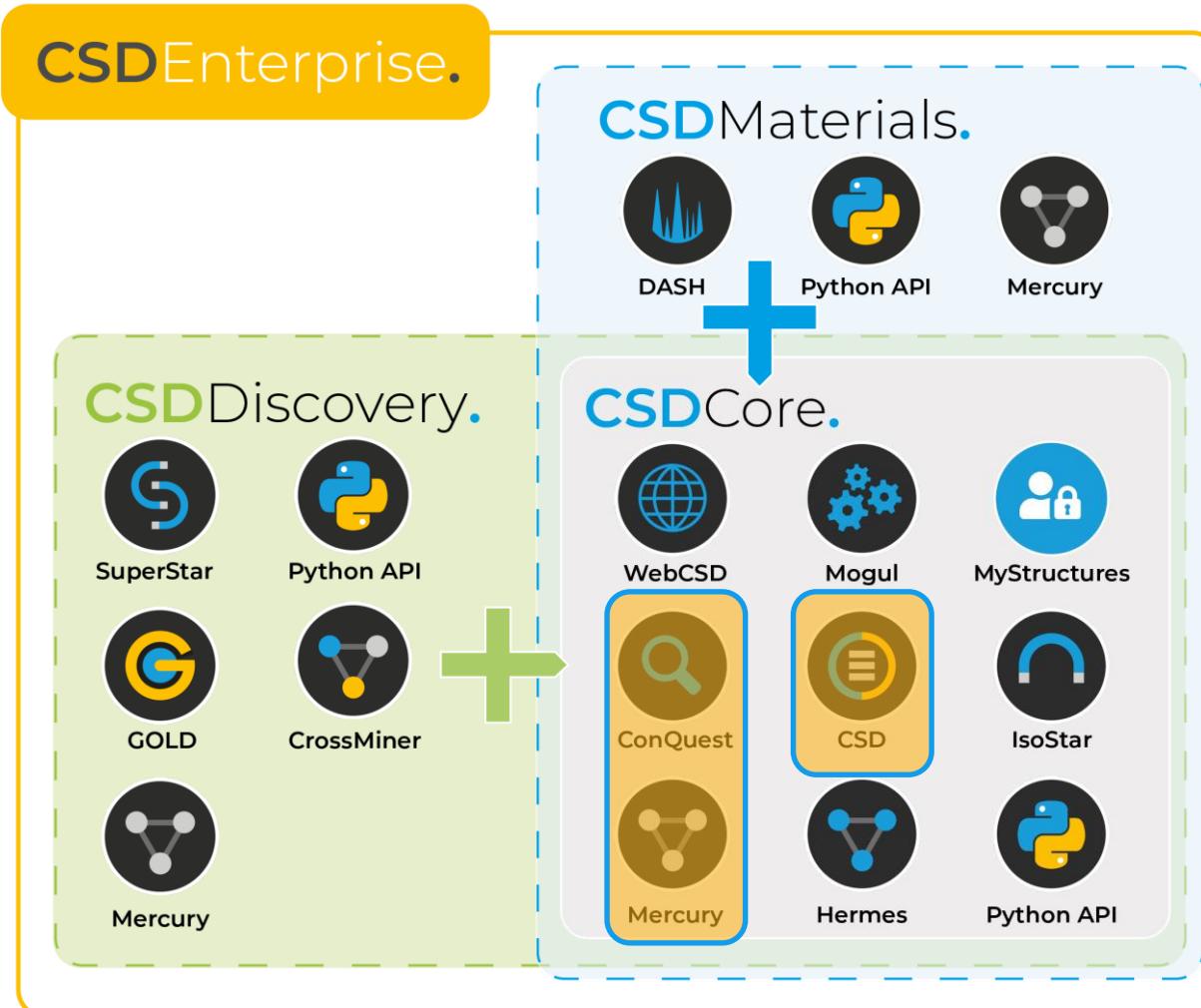


2,2'-bipyridine one of the most frequent ligands in the CSD with >6,000 entries  
~10% mono-coordinated, 90% di-coordinated



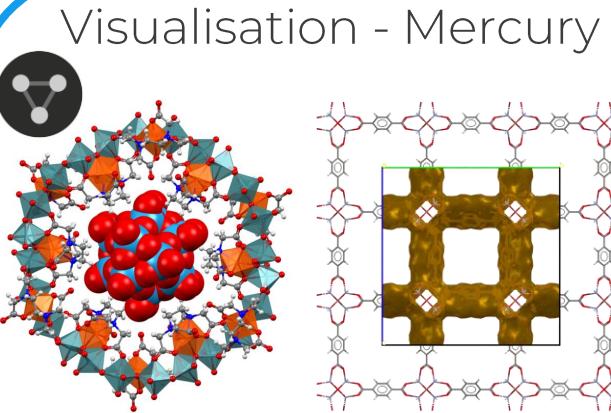
CCDC

# The CSD Portfolio



CCDC

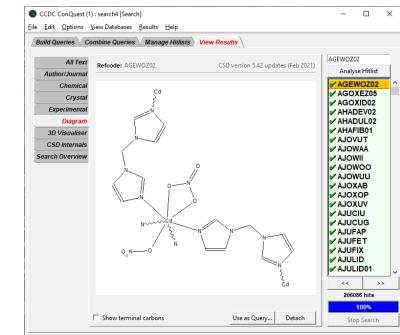
# Search, visualisation and analysis



Visualise data in the CSD or your own structures

- Explore how structures pack together including
  - 1D, 2D and 3D networks, intermolecular and intramolecular bonds
- Visualise important features including
  - Voids, polyhedral display

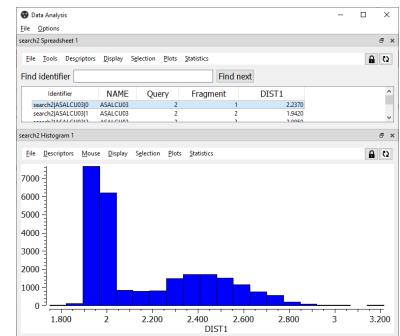
## Search - ConQuest



Search and retrieval of data from the CSD

- Full range of text / numeric database search options
- More complex search functionality
  - Chemical structure searching
  - 3D Geometrical searching
  - Intermolecular non-bonded contact searching

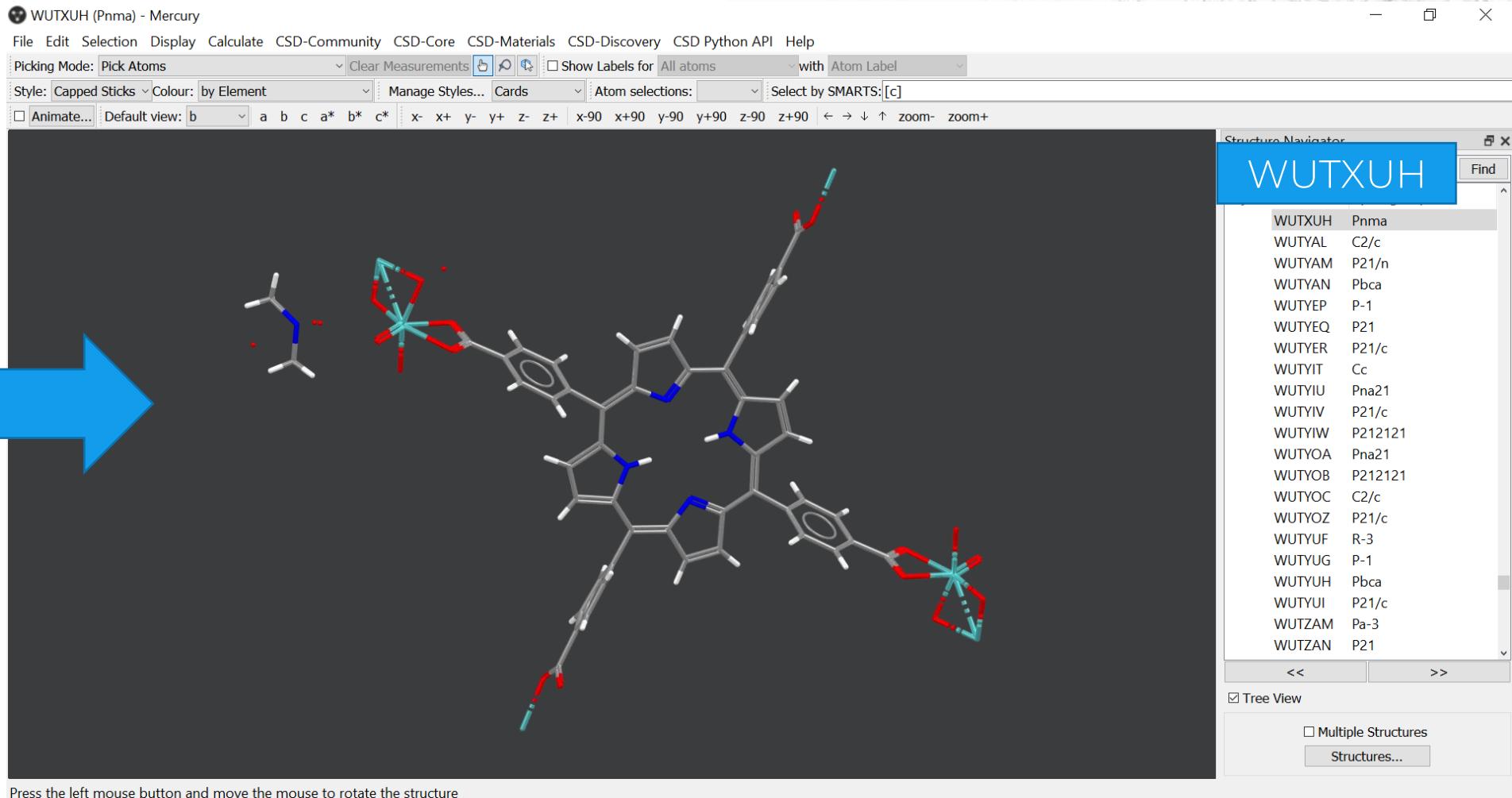
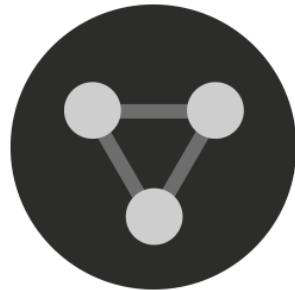
## Analysis – CQ + Mercury



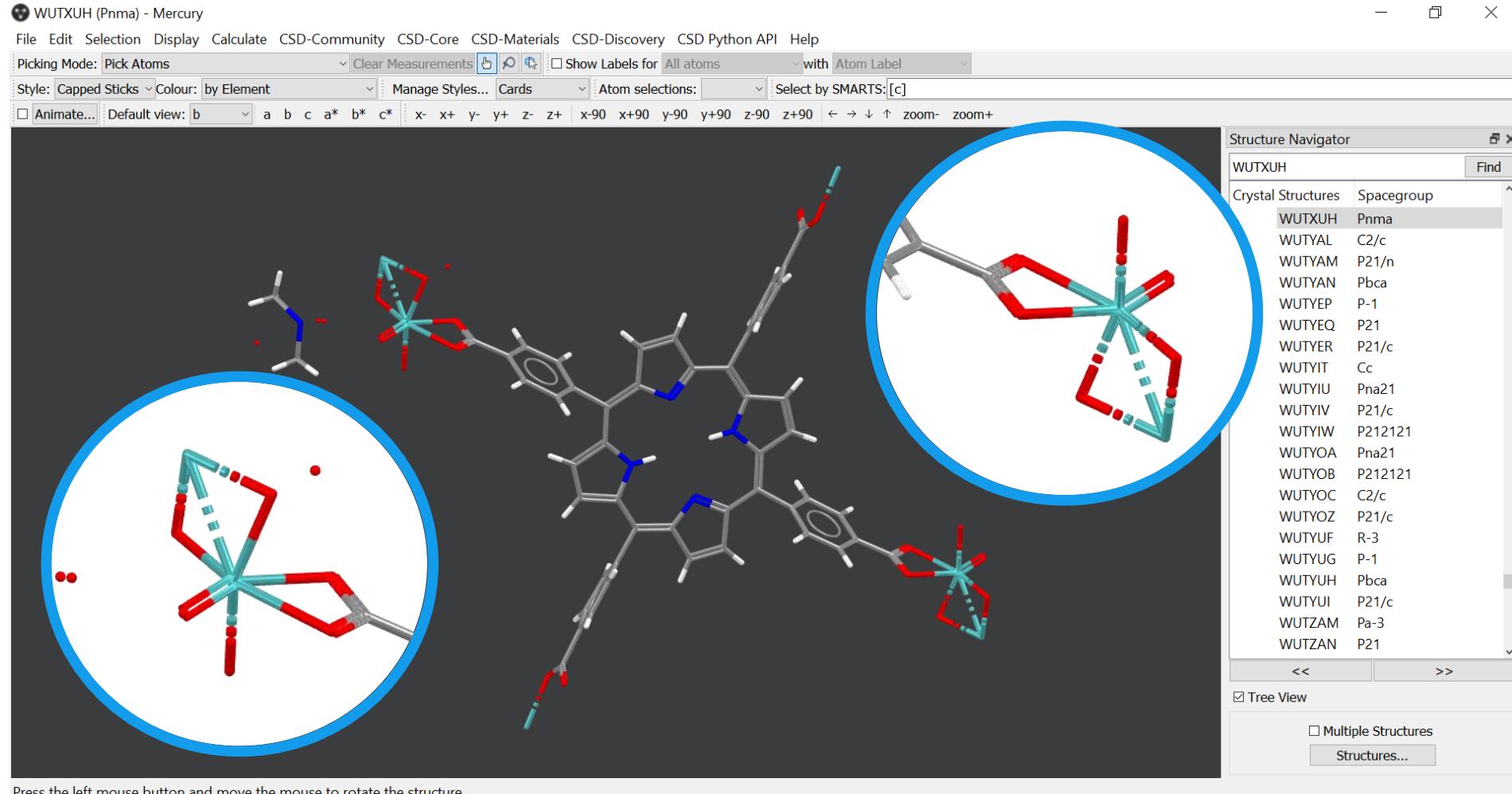
Analyse data in the CSD and compare your own structures

- Learn about metal-coordinations
  - Ligand coordination modes, metal coordination numbers, bond lengths, symmetry and shape, spin states, ligand cone angles, porosity, 1D, 2D and 3D networks

# Show One: Visualisation tips and tricks for metal-organic structures



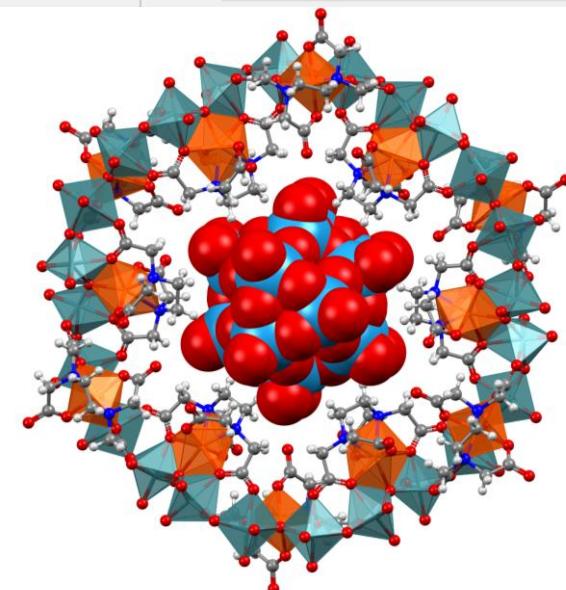
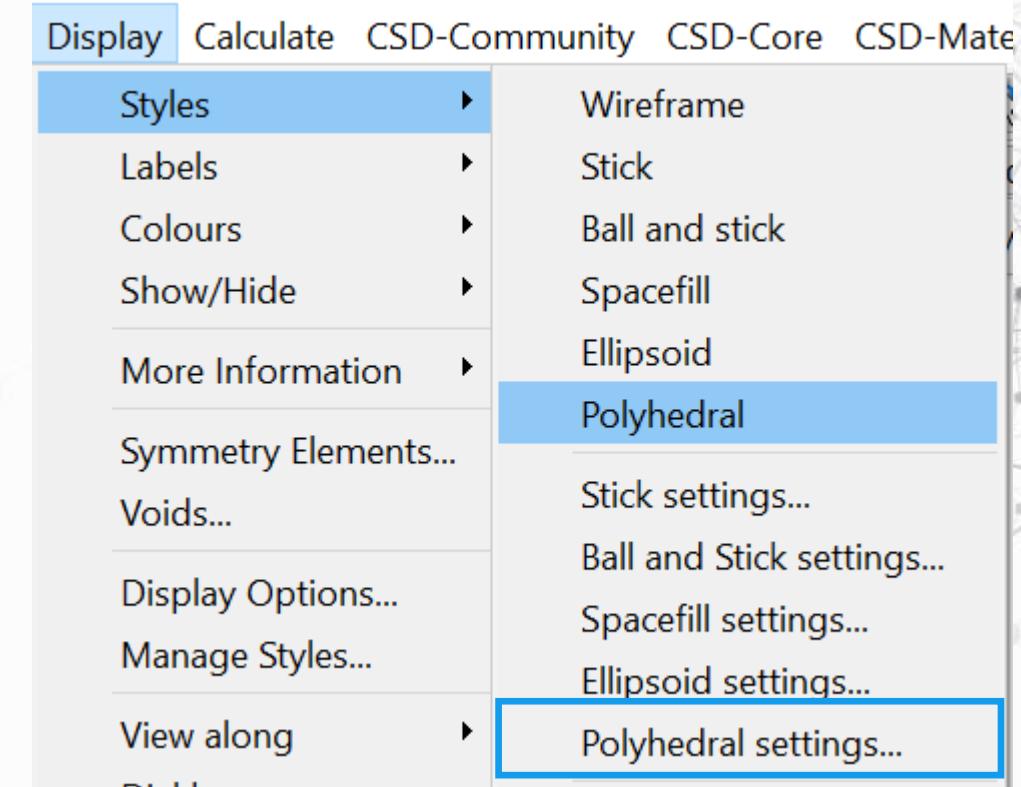
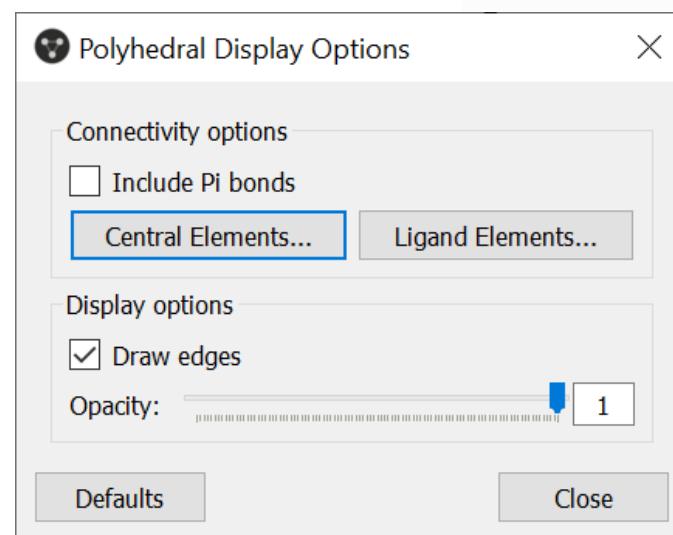
# Polymeric bonds representations



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

# Polyhedral display

- Visualise metal-organic structures using polyhedral representations for the metal-organic coordination centres.
  - Central Elements - atoms around which polyhedra are centred
  - Ligand Elements - neighbouring atoms which create the corners of the polyhedron

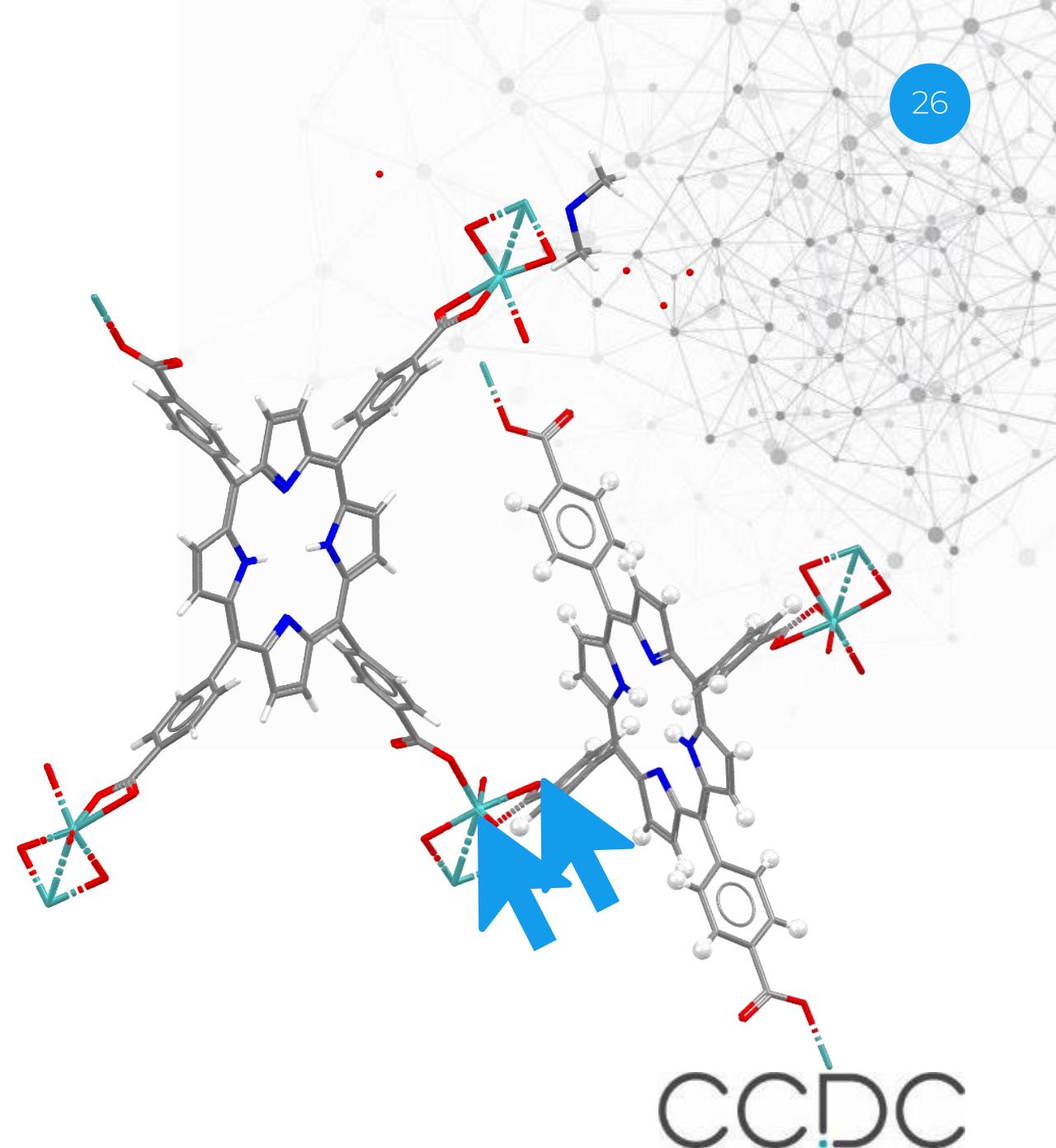
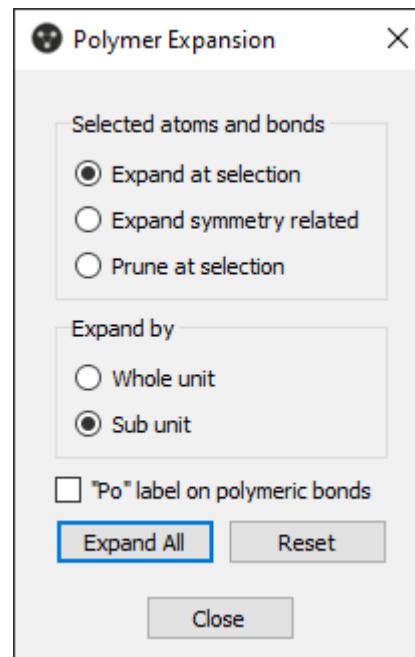


# Polymer Expansion

## Edit

- Undo: Deselect All Ctrl+Z
- Redo: Not Available Ctrl+Y
- Copy Image to Clipboard Ctrl+C
- Edit Structure...
- Auto Edit Structure...
- Normalise Hydrogens...
- Polymer Expansion...**
- Transform Molecules...
- Change Spacegroup Setting...
- Invert Structure
- Change Spacegroup to Subgroup...

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



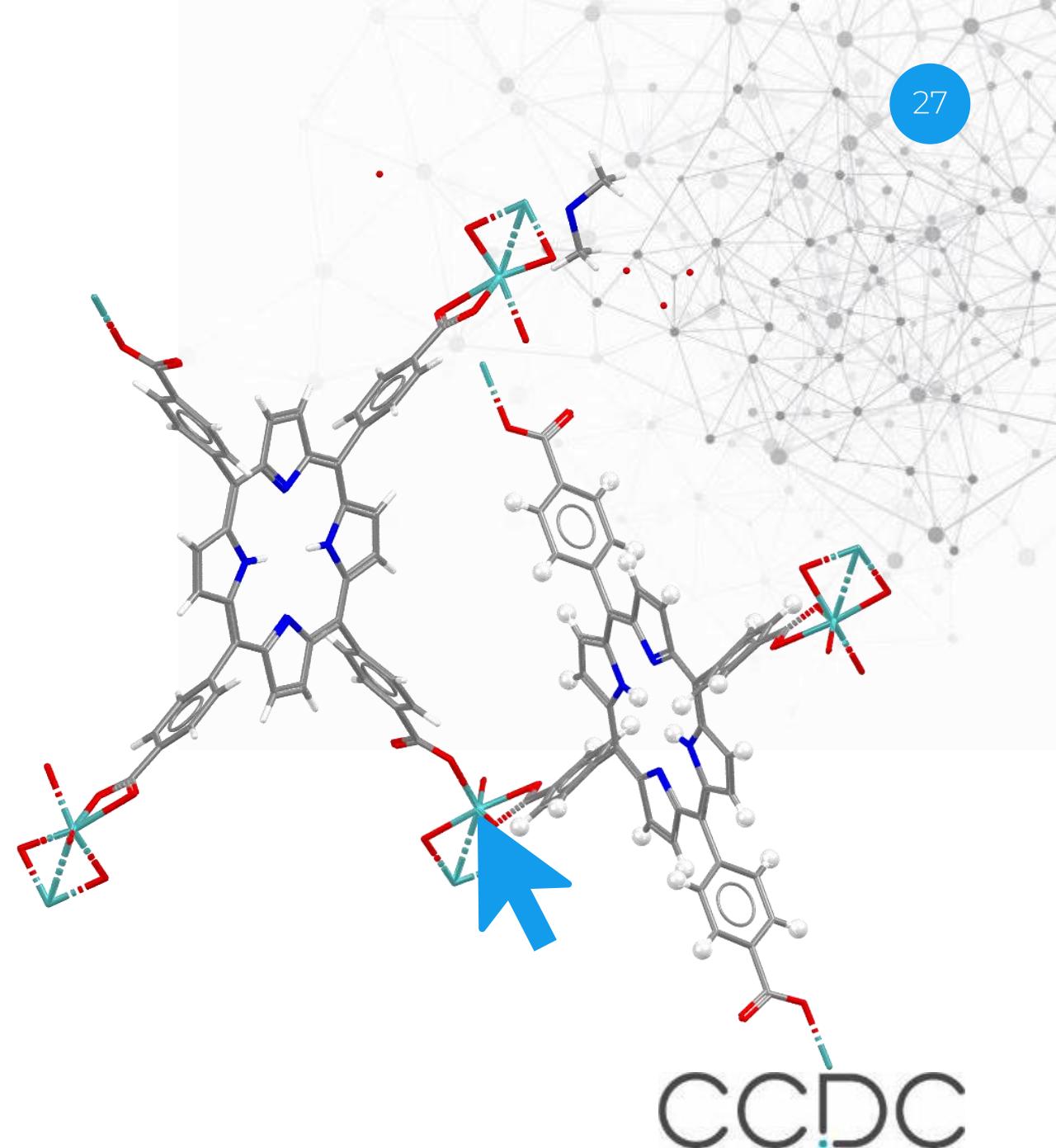
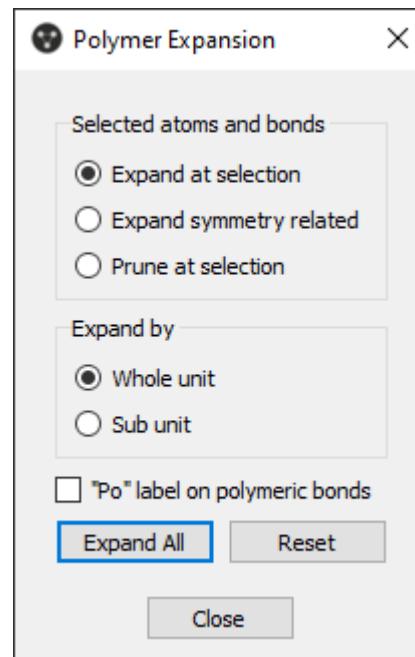
# Polymer Expansion

## Edit

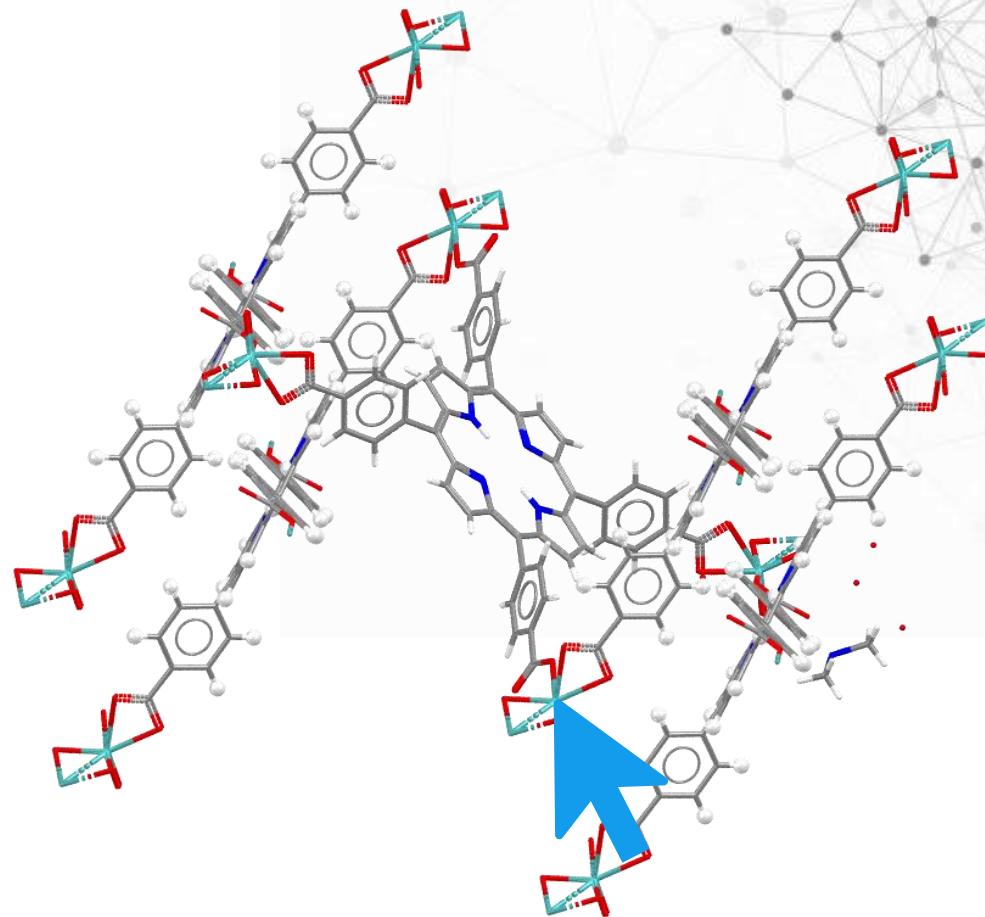
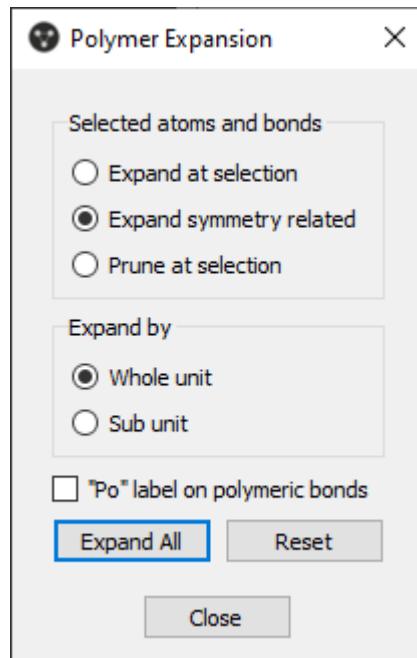
- Undo: Deselect All Ctrl+Z
- Redo: Not Available Ctrl+Y
- Copy Image to Clipboard Ctrl+C
- Edit Structure...
- Auto Edit Structure...
- Normalise Hydrogens...
- Polymer Expansion...**
- Transform Molecules...
- Change Spacegroup Setting...
- Invert Structure
- Change Spacegroup to Subgroup...

**By Whole unit**

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



# Polymer Expansion



# Polymer Expansion

A screenshot of a molecular modeling software interface. On the left, a polymer chain is shown with some atoms and bonds highlighted in red and cyan. A blue arrow points from this visualization to a floating dialog box titled "Polymer Expansion".

**Polymer Expansion**

Selected atoms and bonds

- Expand at selection
- Expand symmetry related
- Prune at selection

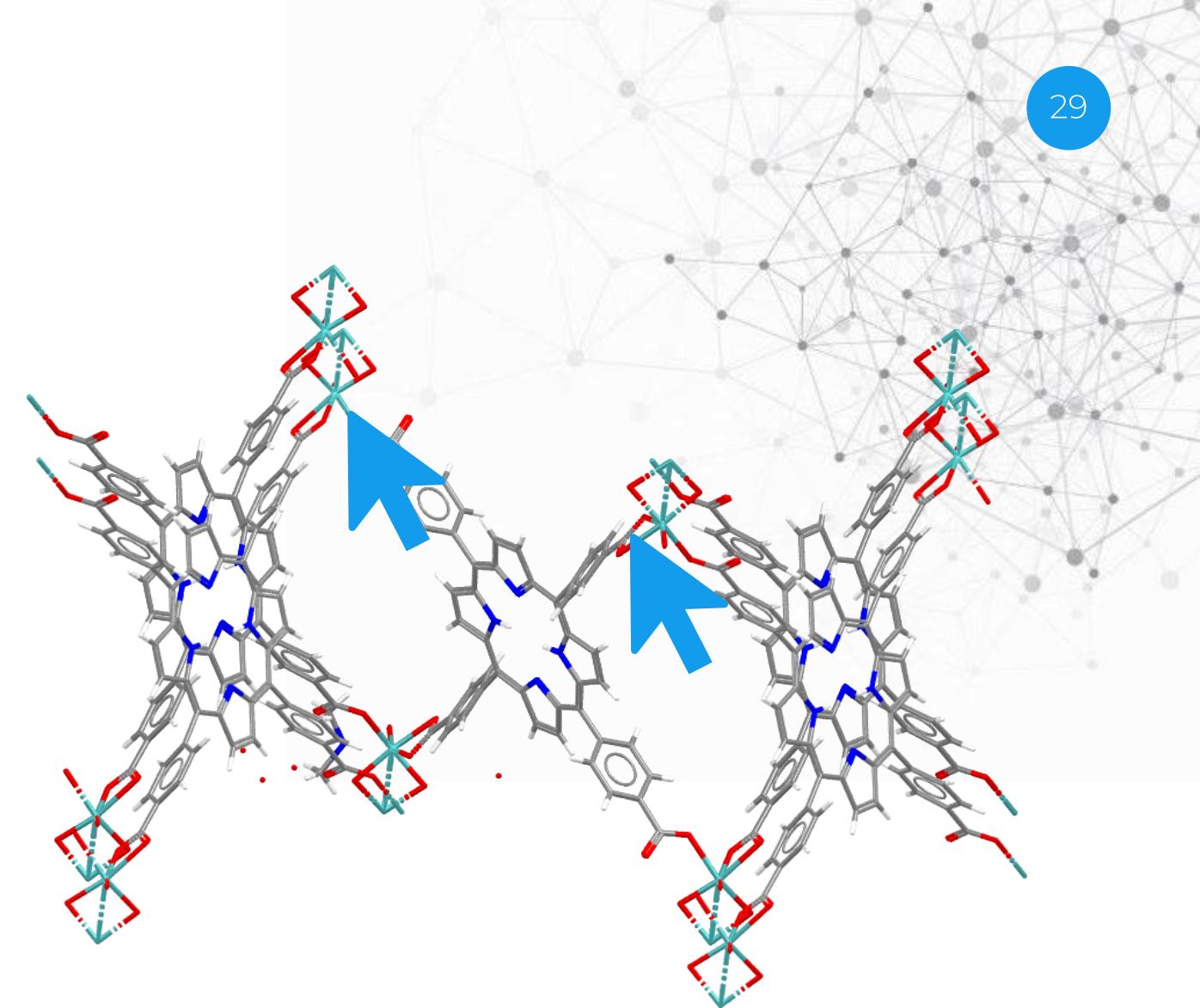
Expand by

- Whole unit
- Sub unit

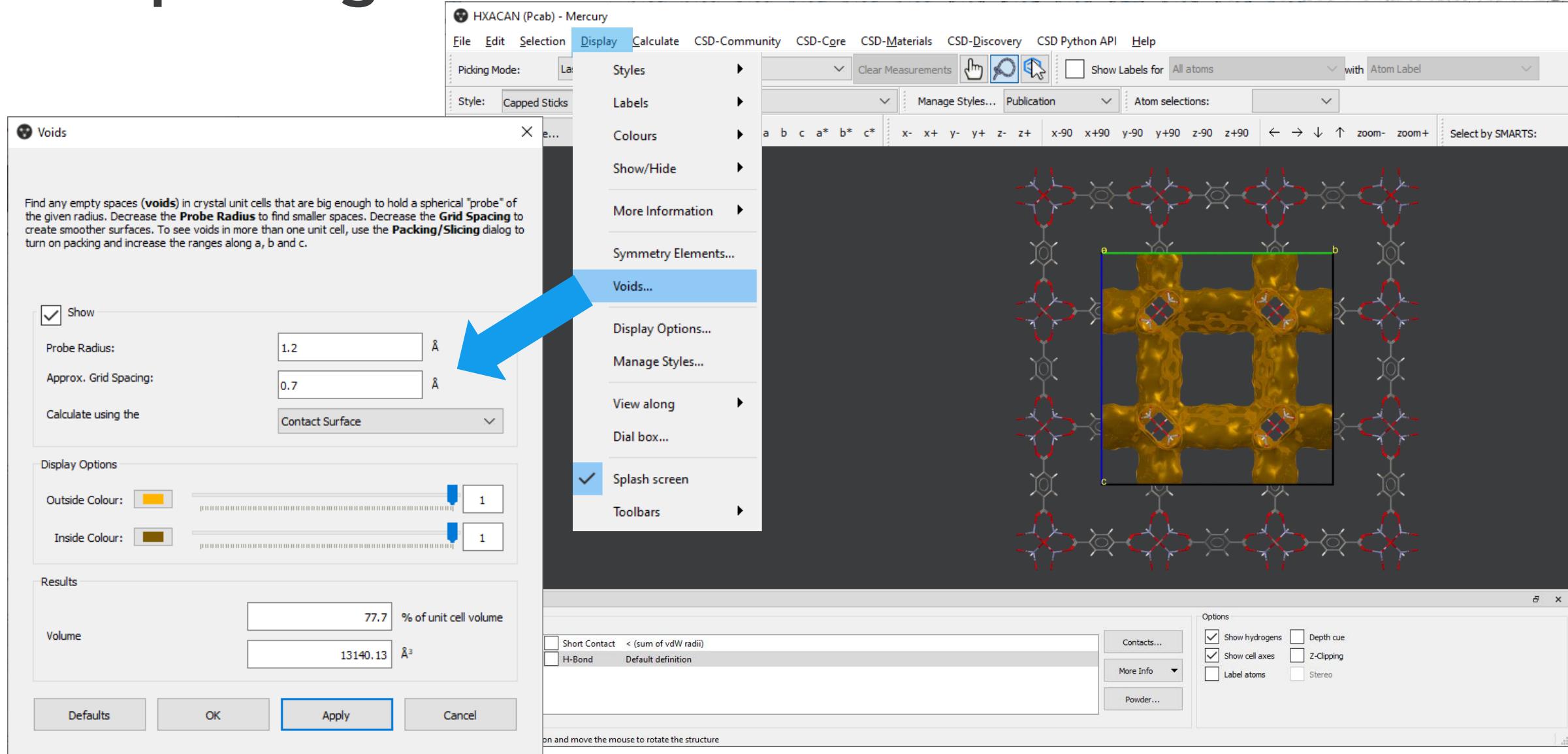
"Po" label on polymeric bonds

**Expand All**   **Reset**

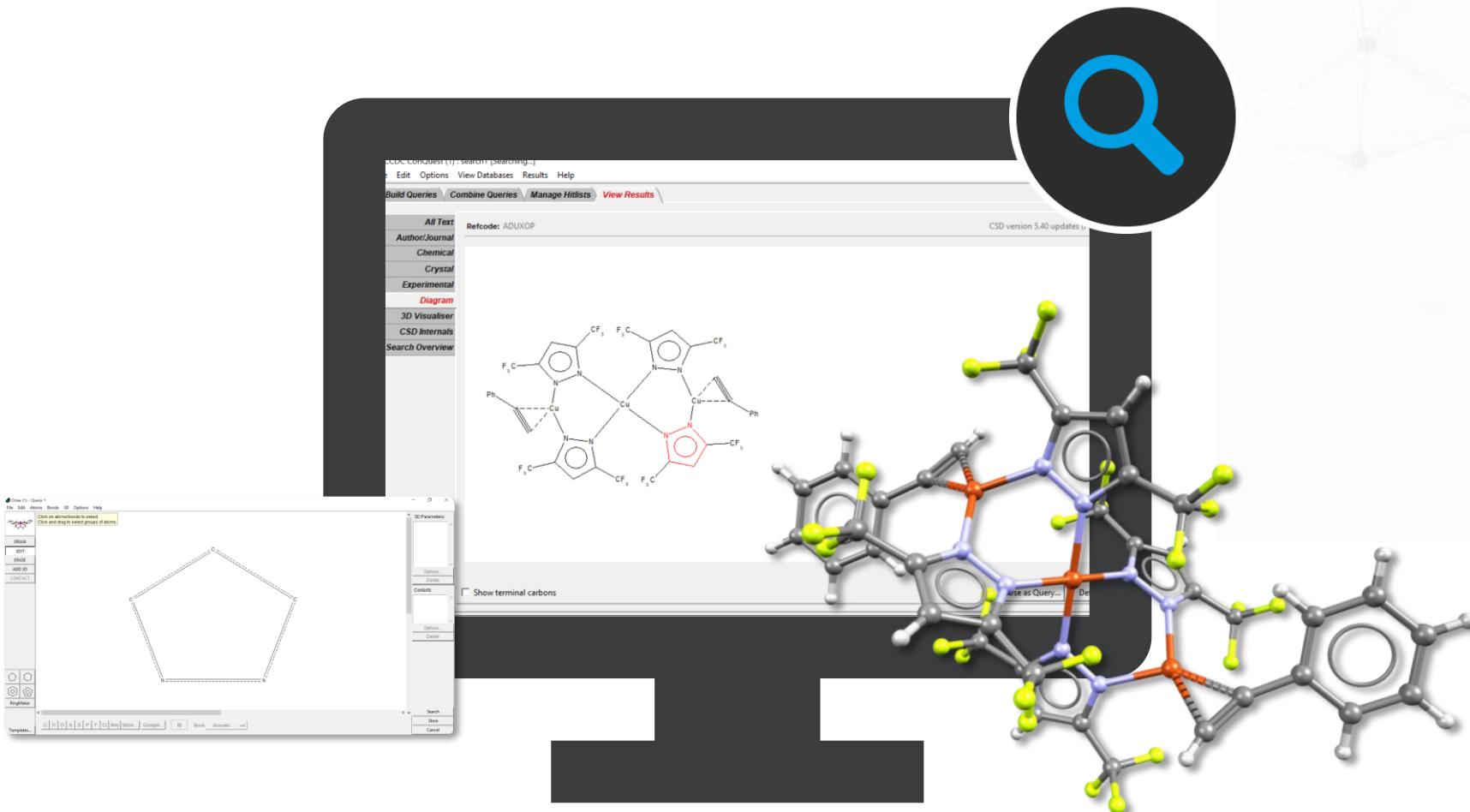
**Close**



# Exploring voids



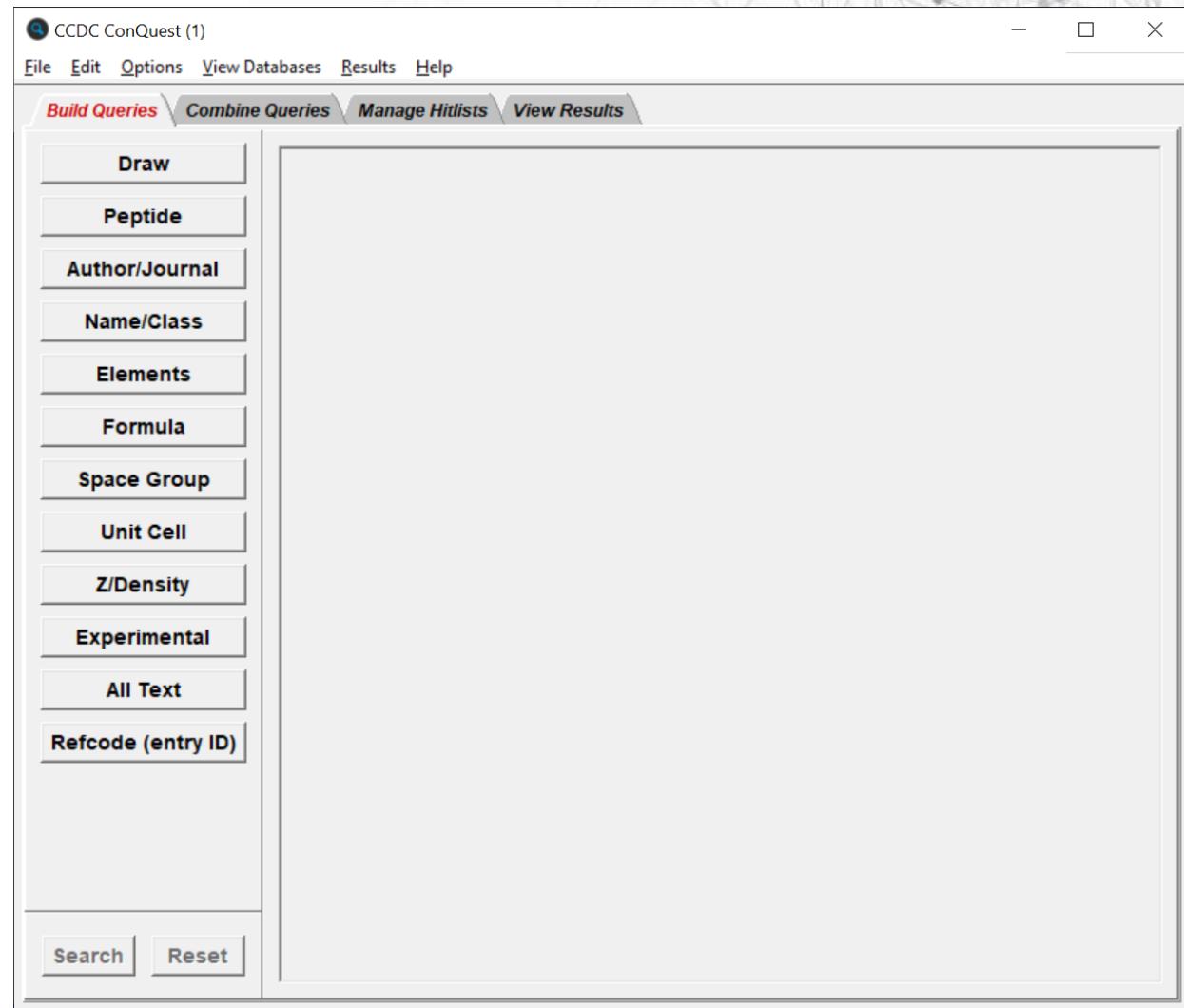
# CSD-Core: Search



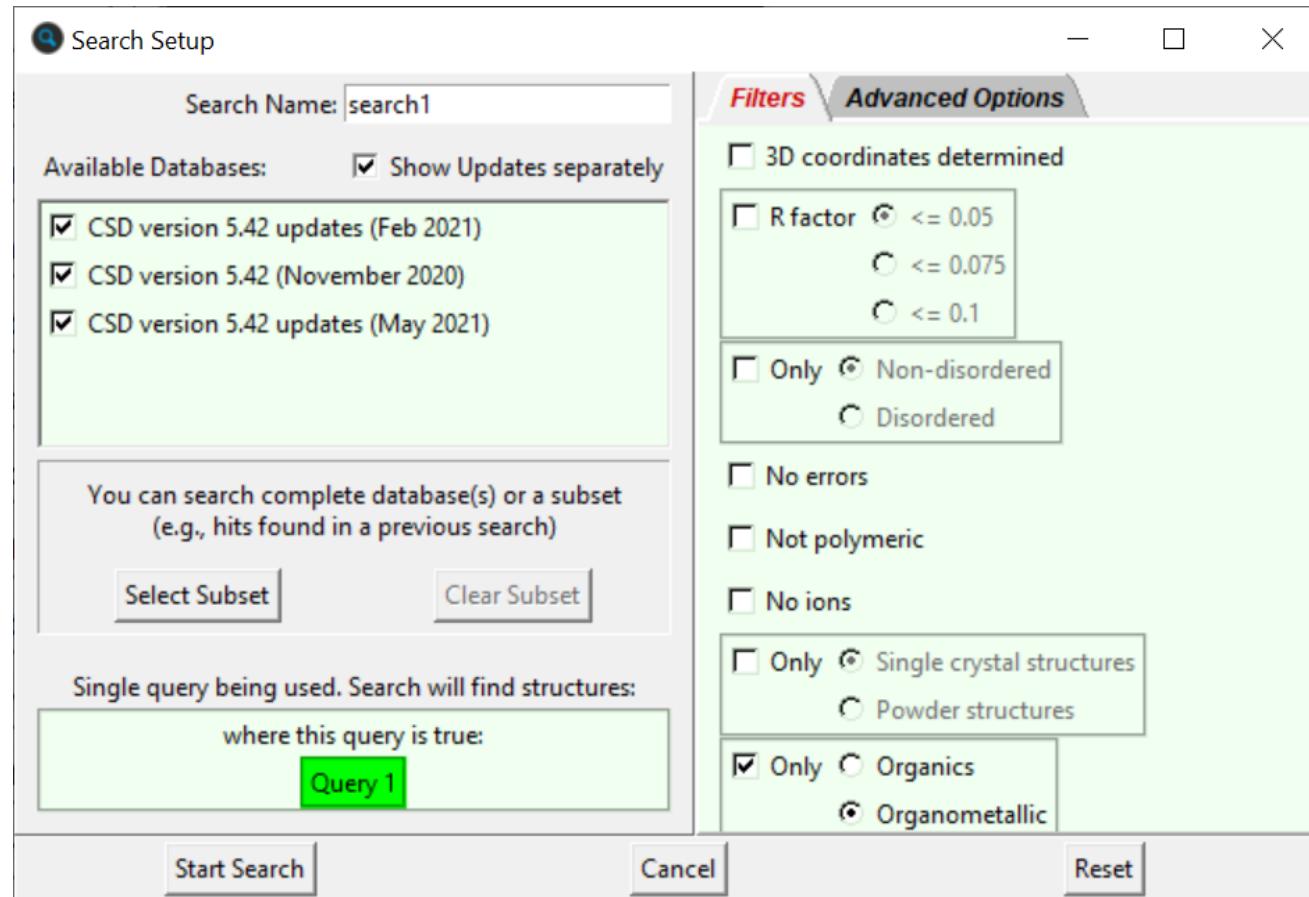
On your desktop: ConQuest

On a browser: [WebCSD](#)  
Or [Free Access Structures](#)

# Show One: Searching metal-organics



# Organometallic filter



- Restricts search to structures containing at least one:
  - Transition metal
  - Lanthanide
  - Actinide
  - Or any of Al, Ga, In, Tl, Ge, Sn, Pb, Sb, Bi, Po

# Oxidation states

- Charges only added to 2D/3D for charge balancing purposes
- Oxidation states are provided in compound name
  - Convention is (i-viii)
- >340,000 entries with oxidation states

The screenshot shows the 'Name/Class' tab selected in the 'Build Queries' section. The search dialog displays a 'Compound Name' field containing '(ii)' and a 'Contains:' dropdown menu. Two checkboxes are visible: 'Ignore non-alphabetic characters' and 'Find exact word'. Below the dialog are 'Search', 'Store', 'Cancel', and 'Reset' buttons.

*Tip:* use parenthesis to avoid

- Other oxidation states
- Polymorphic forms!

The screenshot shows the results for Refcode AGEWOZ02. The compound name 'catena-[bis(μ-1,1'-methylenabis(1H-imidazole))-bis(nitrato)-cadmium(ii)]' is highlighted in yellow. A blue box labeled 'Query highlighted' covers the bottom part of the results table. The right sidebar shows a list of hitlists starting with AGEWOZ02.

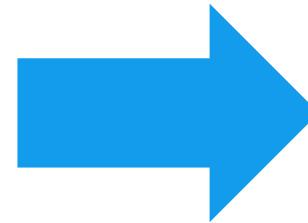
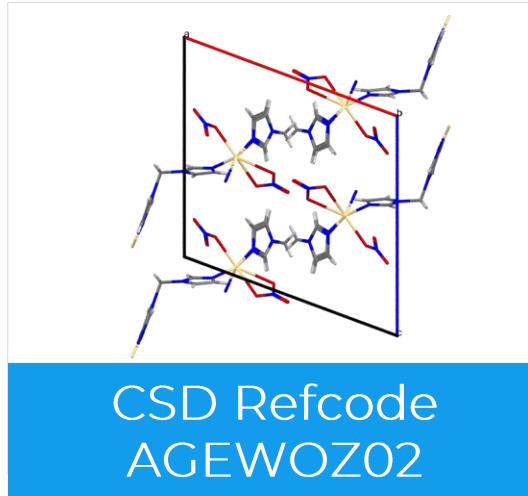
Refcode	Author(s)	Reference	Publication DOI	Deposition	Formula	Compound	Spacegroup	a:	b:	c:	Volume	Molecular Volume	Chemical Units	Z, Z'	R-Factor (%)	Temperature (K)
AGEWOZ02	Shanshan Zhang	J.Mol.Struct. (2021), 1230, 129871	10.1016/j.molstruc.2021.129871	CCDC 2026390	(C <sub>14</sub> H <sub>16</sub> CdN <sub>10</sub> O <sub>6</sub> ) <sub>n</sub>	catena-[bis(μ-1,1'-methylenabis(1H-imidazole))-bis(nitrato)-cadmium(ii)]	P21/c	15.191(<1)	9.159(<1)	14.687(<1)	1916.125	479.031	1	Z: 4.0	Z': 1.0	Room Temp. (283-303)

Query highlighted

catena-[bis(μ-1,1'-methylenabis(1H-imidazole))-bis(nitrato)-cadmium(ii)]

Compound names include information about polymeric nature, oxidation states, ligand names, clathrates and solvates and hydrate

# CSD Refcodes

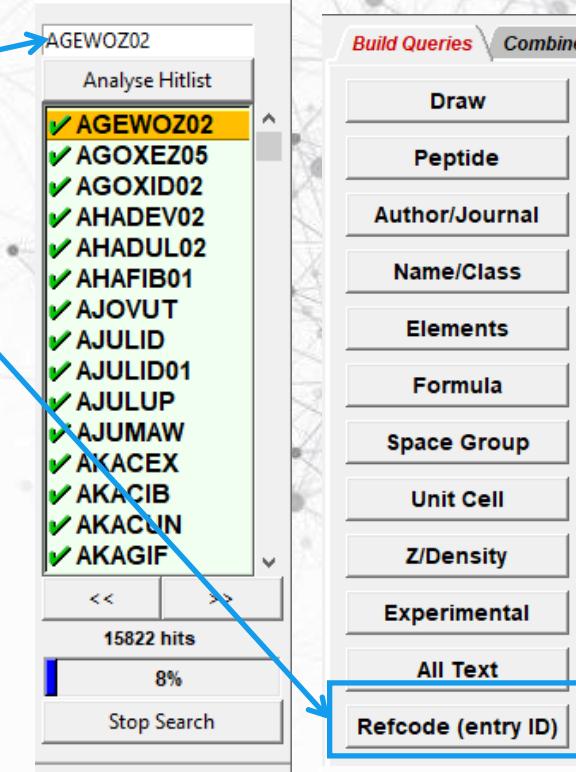


You can use the box to jump to a recode in your search results

You can search the entire CSD for a Recode too

What is AGEWOZ02?

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



AGEWOZ02

Analyse Hitlist

Refcode
AGEWOZ02
AGOXEZ05
AGOXID02
AHADEV02
AHADUL02
AHAFIB01
AJOVUT
AJULID
AJULID01
AJULUP
AJUMAW
AKACEX
AKACIB
AKACUN
AKAGIF

<< >>

15822 hits

8%

Stop Search

All Text

Refcode (entry ID)

## 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
- 
- Some other recognisable refcodes are: KITTEN, DISNEY, GAUTAM, NELSON, GLYCIN
  - Do you have a favourite refcode? You can use the chat box to tell us yours!

**Link to the publication**

**Link to deposited CIF file**

CCDC ConQuest (1) : search4 [Search]

File Edit Options View Databases Results Help

Build Queries Combine Queries Manage Hitlists View Results

All Text Author/Journal

Refcode: AGEWOZ02 CSD version 5.42 updates (Feb 2021)

Author(s)	Shanshan Zhang
Reference	J.Mol.Struct. (2021), 1230, 129871
Publication DOI	<a href="https://doi.org/10.1016/j.molstruc.2021.129871">10.1016/j.molstruc.2021.129871</a>
Deposition	CCDC <a href="#">2026390</a>
Formula	(C <sub>1</sub> H <sub>16</sub> CdN <sub>10</sub> O <sub>6</sub> ) <sub>n</sub>
Compound	catena-[bis( $\mu$ -1,1'-methylenebis(1H-imidazole))-bis(nitroato)m] <sub>n</sub>
Spacegroup	Name: P21/c Number: 14
Cell	a: 15.191(<1) b: 9.159(<1) c: 14.6 alpha: 90.00 beta: 110.32(<1) gamma: 90.0 Volume: 1916.125
Reduced Cell	a: 9.159 b: 14.687 c: 15.191 alpha: 110.32 beta: 90.00 gamma: 90.00 Volume: 1916.125
Molecular Volume	479.031
Chemical Units	1
Z, Z'	Z: 4.0 Z': 1.0
R-Factor (%)	2.61
Temperature (K)	Room Temp. (283-303)
Density	CCDC: 1.847 Author: 1.847
Intensity Meas	diffractometer

Detach 100% Stop Search

AGEWOZ02 Analyse Hitlist

- ✓ AGEWOZ02
- ✓ AGOXEZ05
- ✓ AGOXID02
- ✓ AHADEV02
- ✓ AHADUL02
- ✓ AHAFIB01
- ✓ AJOVUT
- ✓ AJOWAA
- ✓ AJOWII
- ✓ AJOWOO
- ✓ AJOWUU
- ✓ AJOXAB
- ✓ AJOXOP
- ✓ AJOXUV
- ✓ AJUCIU
- ✓ AJUCUG
- ✓ AJUFAP
- ✓ AJUFET
- ✓ AJUFIX
- ✓ AJULID
- ✓ AJULID01
- ✓ AJULUP
- ✓ AJUMAW
- ✓ AKACEX
- ✓ AKACIB
- ✓ AKACUN
- ✓ AKAGIF

206086 hits 100% Stop Search

Curated 2D chemical diagram

CCDC ConQuest (1) : search4 [Search]

File Edit Options View Databases Results Help

Build Queries Combine Queries Manage Hitlists View Results

All Text Author/Journal Chemical Crystal Experimental

Diagram

3D Visualiser CSD Internals Search Overview

Refcode: AGEWOZ02 CSD version 5.42 updates (Feb 2021)

Diagram window showing a 2D chemical structure of a catena polymer. The structure features a central cadmium atom (Cd) coordinated to four imidazole rings. Each imidazole ring is substituted with nitro groups (NO<sub>2</sub>). A 'Use-as-Query Options' dialog is open, containing settings for 'Hydrogens' (checkbox for 'Include hydrogen atoms' checked) and 'Chemical Units (molecules, ions, etc.)' (radio buttons for 'Biggest chemical unit only' (selected), 'Make each chemical unit a separate query', and 'Include all chemical units in a single query').

Use-as-Query Options

Hydrogens:

Include hydrogen atoms

Chemical Units (molecules, ions, etc.):

Biggest chemical unit only

Make each chemical unit a separate query

Include all chemical units in a single query

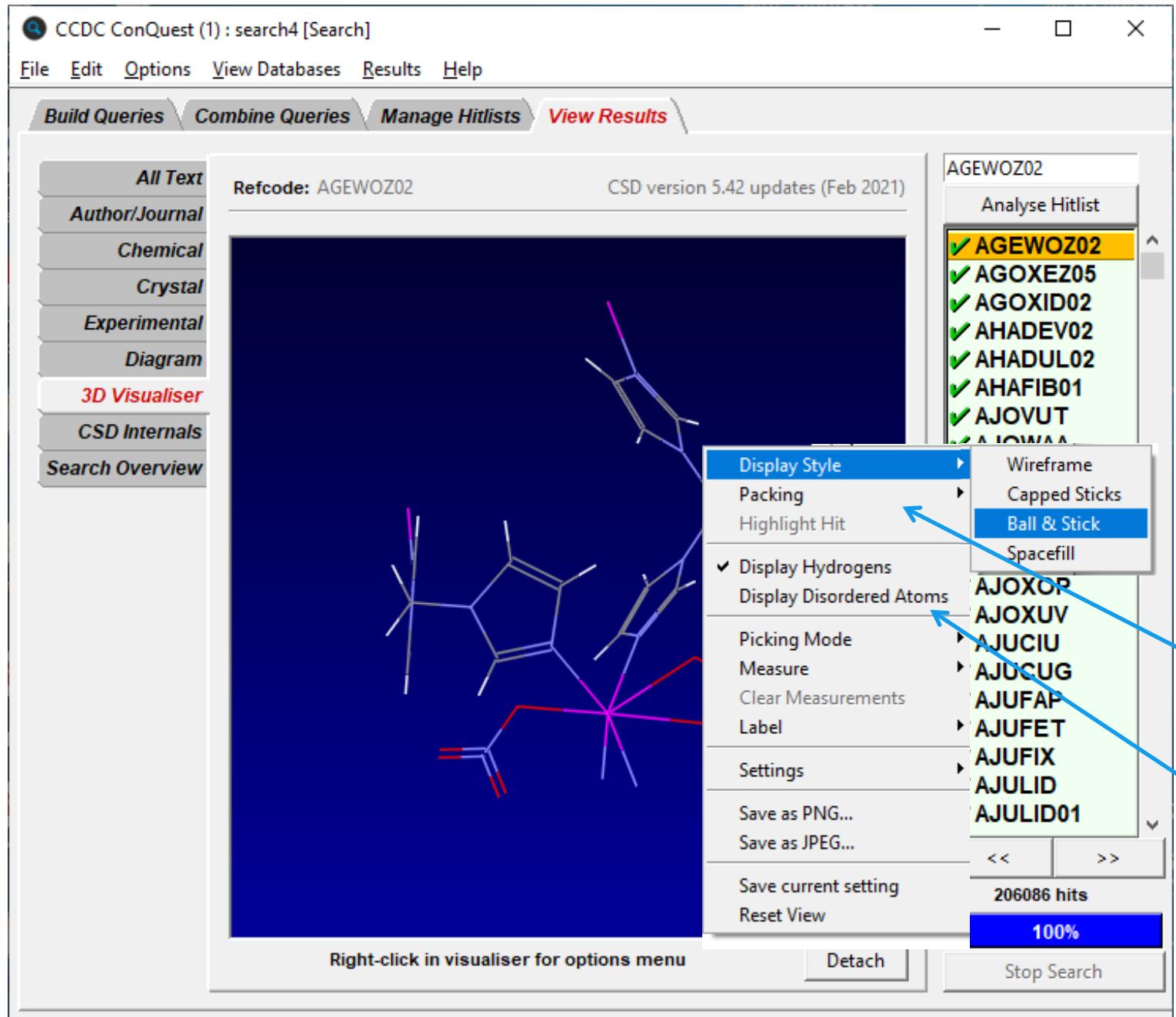
Cancel OK

Show terminal carbons Use as Query... Detach

AGEWOZ02 Analyse Hitlist

- ✓ AGEWOZ02
- ✓ AGOXEZ05
- ✓ AGOXID02
- ✓ AHADEV02
- ✓ AHADUL02
- ✓ AHAFIB01
- ✓ AJOVUT
- ✓ AJOWAA
- ✓ AJOWII
- ✓ AJOWOO
- ✓ AJOWUU
- ✓ AJOXAB
- ✓ AJOXOP
- ✓ AJOXUV
- ✓ AJUCIU
- ✓ AJUCUG
- ✓ AJUFAP
- ✓ AJUFET
- ✓ AJUFIX
- ✓ AJULID
- ✓ AJULID01

206086 hits 100% Stop Search



Left clicking each tab will display different information

Right clicking on 3D visualiser enables you to change the display style

Packing > Molecule or unit cell

Display or hide disordered atoms

# ConQuest – more text/numeric searches

**Name/Class (1) - New**

Compound Name Contains:

catena-

Ignore non-alphabetic characters,  
e.g. "butadiene" will match "buta-1,3-diene"

Find exact word,  
e.g. "hydrazine" will not match "acetylhydrazine"

Add Replace Delete

CSD contains IUPAC style names with CCDC conventions as well as more common names such as MOF names when known – these may not be comprehensive

Use catena- to search for metal-organic polymers

Search Store Cancel Reset

**Z/Density (1) - New**

No. of "Molecules" per Unit Cell (Z) =

No. of "Molecules" per Asymmetric Unit (Z') =

No. of Atoms with 3D Coordinates =

Calculated Density (g/cm<sup>3</sup>) =

No. of Chemical Units (molecules, ions, etc.) in Entry =

Search Store Cancel Reset

*Tip:* use a combined search to find structures with catena- or catena( and this will avoid finding catenanes

**All Text (1) - New**

**Text Search Required Fields**

New Box

Either select from list or enter in box(es) below

cylindrical  
decompose  
deliquescent  
diamagnetic  
diamond  
efflorescent  
electron  
explosive  
fawn  
fluorescent  
green  
grey  
habit

The search will find words starting with what is entered in the boxes.

If two or more words are typed into the same box the search will be for the exact phrase specified. To find entries containing two or more words that need not be adjacent, use the New Box button and type the required words into separate input boxes.

Search Store Cancel Reset

# ConQuest – Formula search

**Build Queries** **Combine**

- Draw
- Peptide
- Author/Journal
- Name/Class
- Elements
- Formula
- Space Group
- Unit Cell
- Z/Density
- Experimental
- All Text
- Refcode (entry ID)

**Formula (1) - New**

**Formula**

**C<sub>10</sub>H<sub>10</sub>Fe**

Type in formula, e.g. C<sub>6</sub>H<sub>12</sub>O<sub>6</sub>  
You may specify an inexact formula, e.g. Ca1-3 O>4  
finds entries with 1,2 or 3 Calciums and more than 4 Oxygens.

or **Select from Table**

**Formula applies to**

an individual molecule  
 all molecules in structure added together  
 Other atoms allowed in molecule/structure

**Search** **Store** **Cancel** **Reset**

**Select Formula Elements**

Select required elements

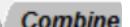
1A	H	1M	2M	3M	4M	8A												
HD	D	2A																
Li	Be																	
Na	Mg	3B	4B	5B	6B	7B	8B	8Z	1B	2B	3A	4A	5A	6A	7A	He		
1R	K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
TR	Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
3R	Cs	Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
	Fr	Ra	Ac															
	LN	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu			
	AN	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr			

Select atom counts

C	=	10
H	=	10
Fe	=	1
	Range	

**Cancel** **Done**

# ConQuest – Elements search

**Build Queries**  **Combine**

**Draw**

**Peptide**

**Author/Journal**

**Name/Class**

**Elements**

**Formula**

**Space Group**

**Unit Cell**

**Z/Density**

**Experimental**

**All Text**

**Refcode (entry ID)**

 Elements (1) - New - □ ×

**Elements Required to be Present**

Type in elements, e.g. C H Se

or **Select from Table**  

Elements must be in

same molecule

same crystal structure

Other elements allowed in molecule/structure

**Heaviest Permitted Element in Formula Unit**

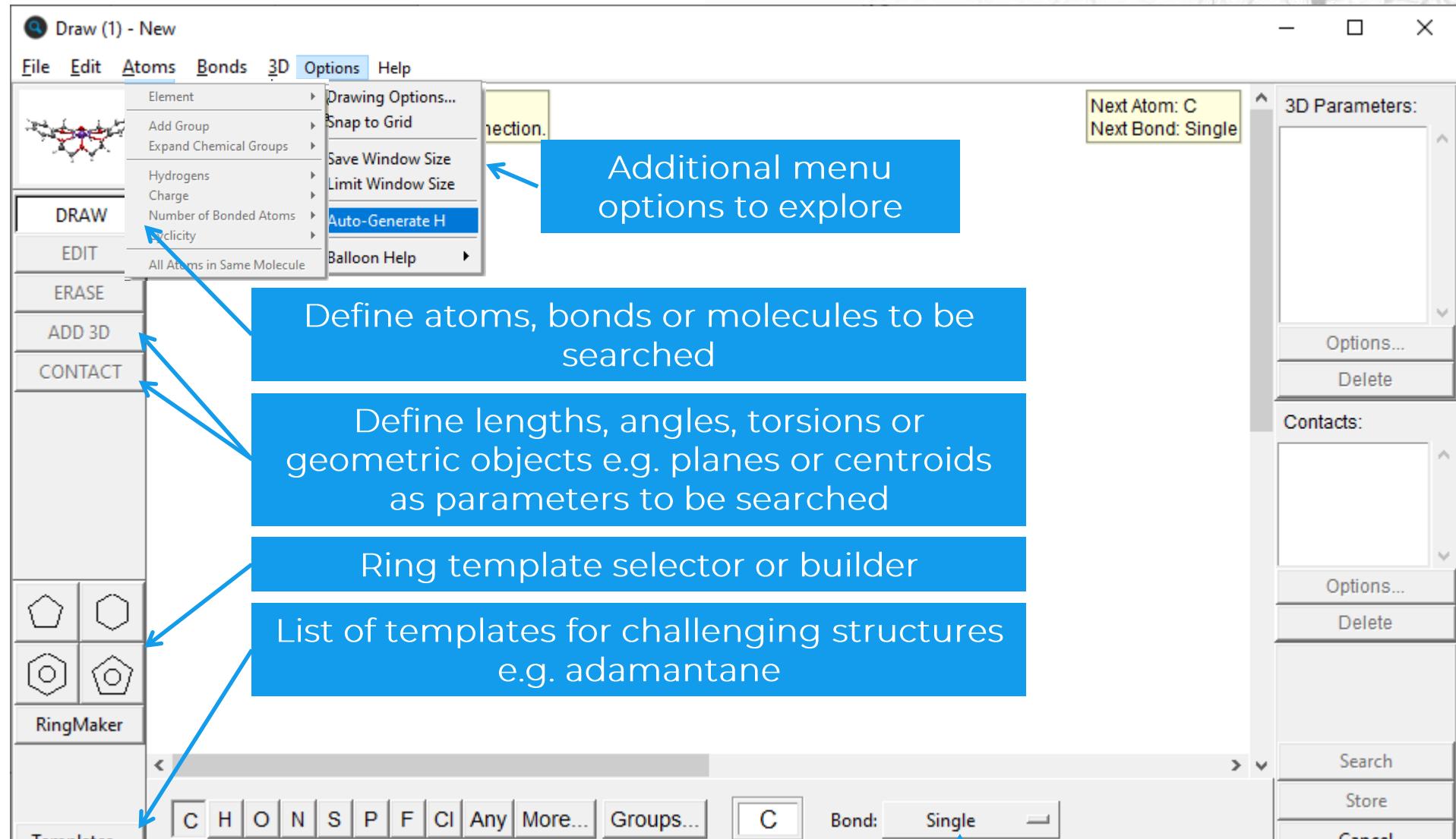
-- Not Set -- Select from Table

**Search** **Store** **Cancel** **Reset**

# Draw

Build Queries Combine

- Draw
- Peptide
- Author/Journal
- Name/Class
- Elements
- Formula
- Space Group
- Unit Cell
- Z/Density
- Experimental
- All Text
- Refcode (entry ID)

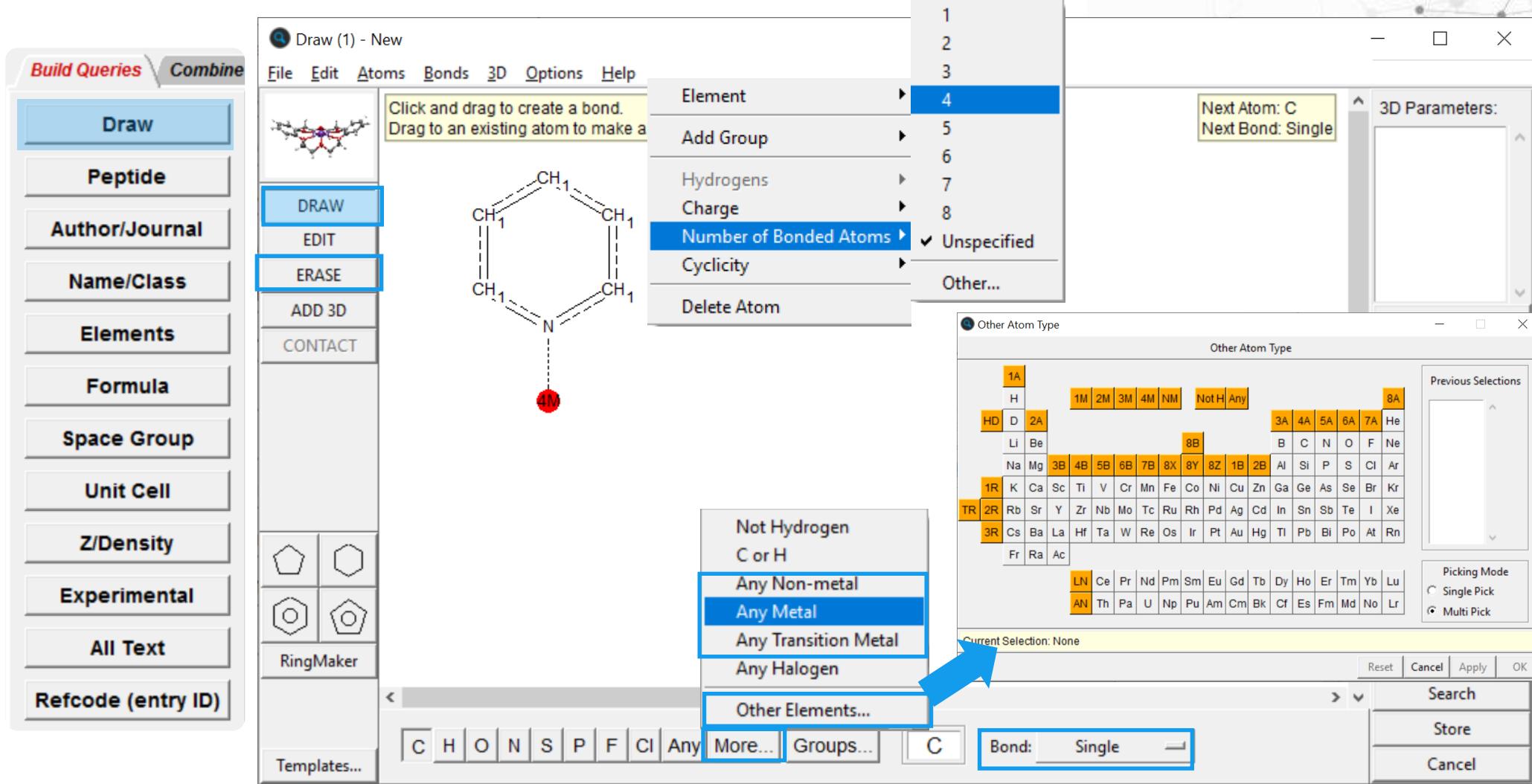


Select specific or general atom types/functional groups

Select bond type

CCDC

# ConQuest – Draw/Structure search



Left click in space  
to add an atom



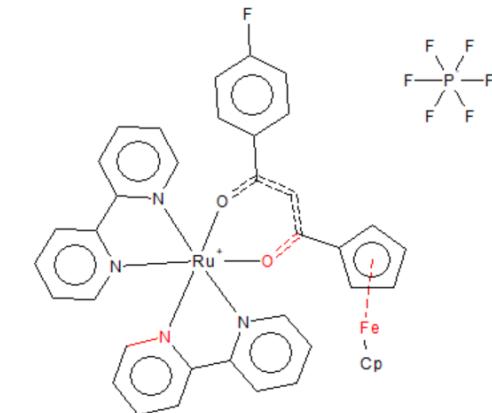
Left click and drag  
from an atom to  
attach an atom



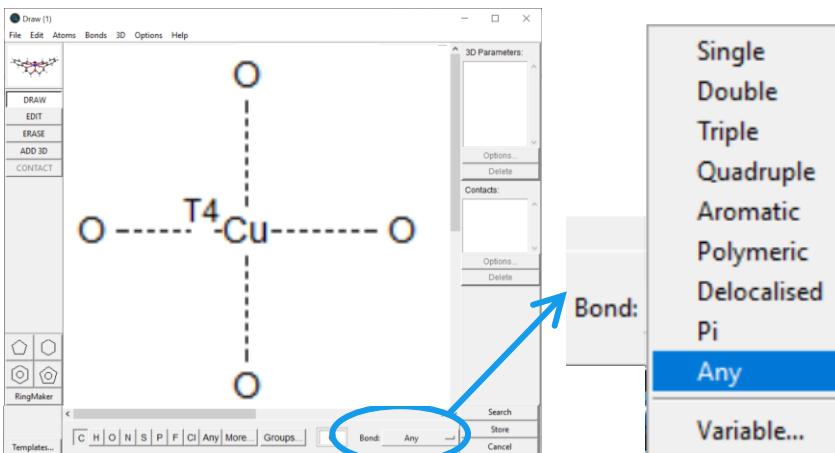
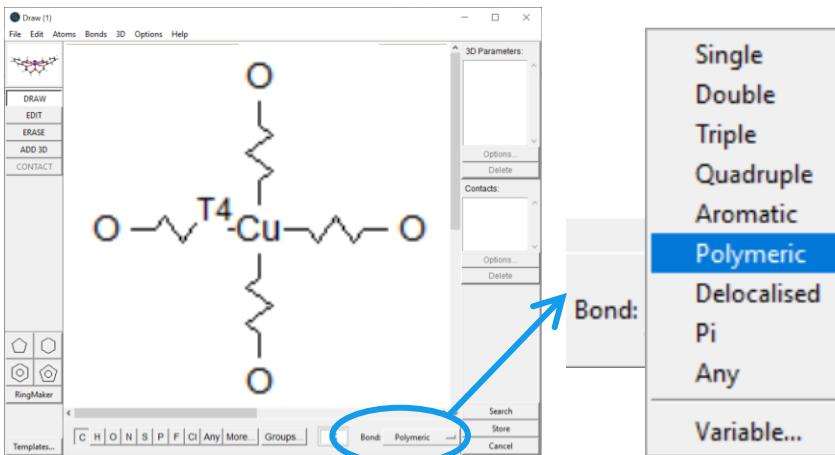
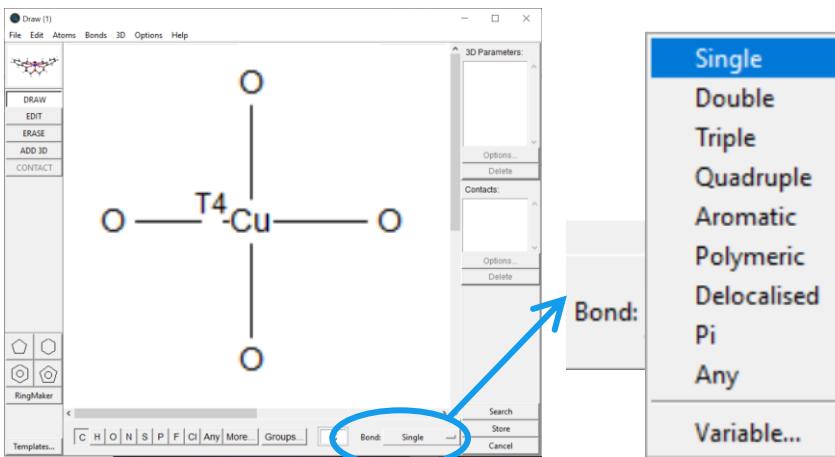
Right click on an  
atom or bond to  
edit or add extra  
properties to it

# Some useful CCDC drawing conventions

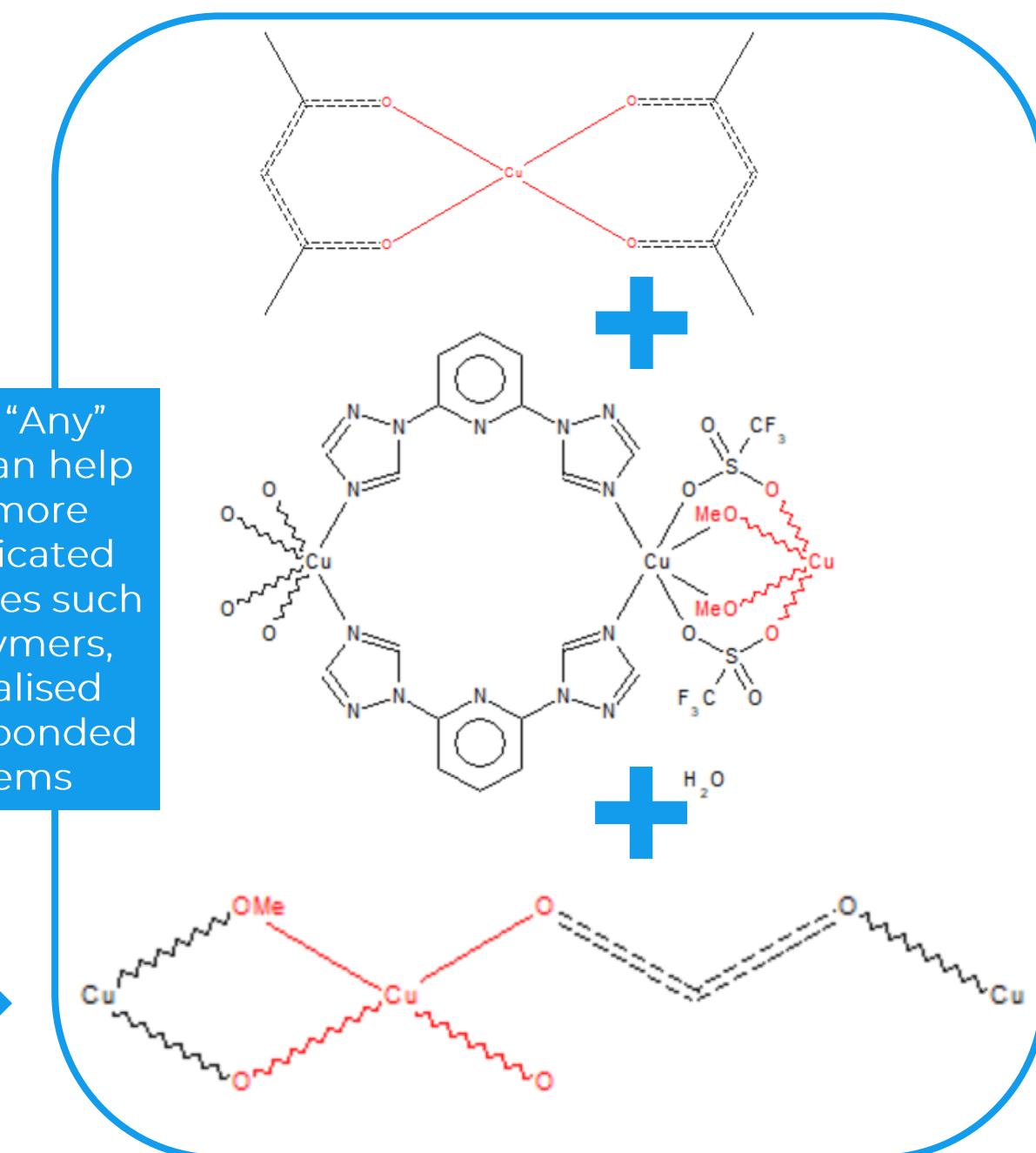
- When appropriate the CSD uses:
  - Aromatic bonds
  - Delocalised bonds
  - Pi-bonds
  - Polymeric bonds – depicted polymeric unit may vary
- CSD 2D chemical representations usually include:
  - The CCDC representation of the structure
  - Missing H atoms/other atoms
  - Squeezed or undetermined parts of the structure
  - Charge – this is the property of an atom rather than an entity but is only used for charge balance purposes



# Substructure found



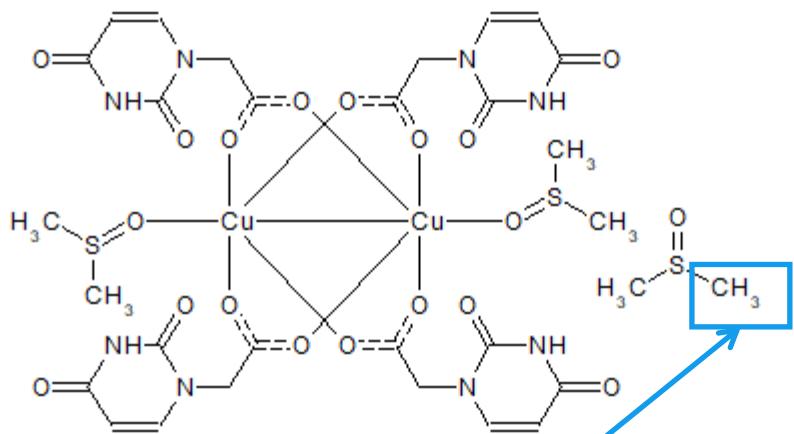
Using "Any" bond can help find more complicated structures such as polymers, delocalised and pi-bonded systems



All Text  
Author/Journal  
Chemical  
Crystal  
Experimental  
Diagram  
3D Visualiser  
CSD Internals  
Search Overview

Refcode: TUWMOP

CSD version 5.42 (November 2020)

 Show terminal carbons

Use as Query...

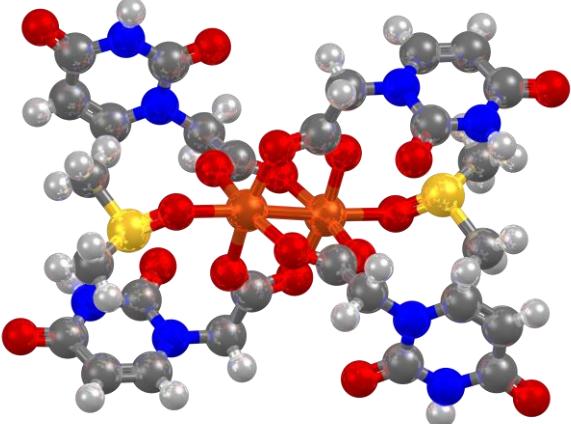
Detach

TUWMOP
Analyse Hitlist
<b>TUWMOP</b>
TUWMOQ
TUWMUT
TUWMUU
TUWMUV
TUWMUW
TUWNAA
TUWNAB
TUWNAC
TUWNAD
TUWNEE
TUWNEF
TUWNEG
TUWNEH
TUWNII
TUWNIJ
TUWNIK
TUWNIL
TUWNOO
TUWNOQ
TUWNUU
TUWNUV
TUWNUW
TUWNIIIX

&lt;&lt; &gt;&gt;

1097619 hits

100%



Could we use the CSD to explore the nature of copper carboxylate paddle structures and then go on to explore other metals?

Use-as-Query Options

Hydrogens:

Include hydrogen atoms

Chemical Units (molecules, ions, etc.):

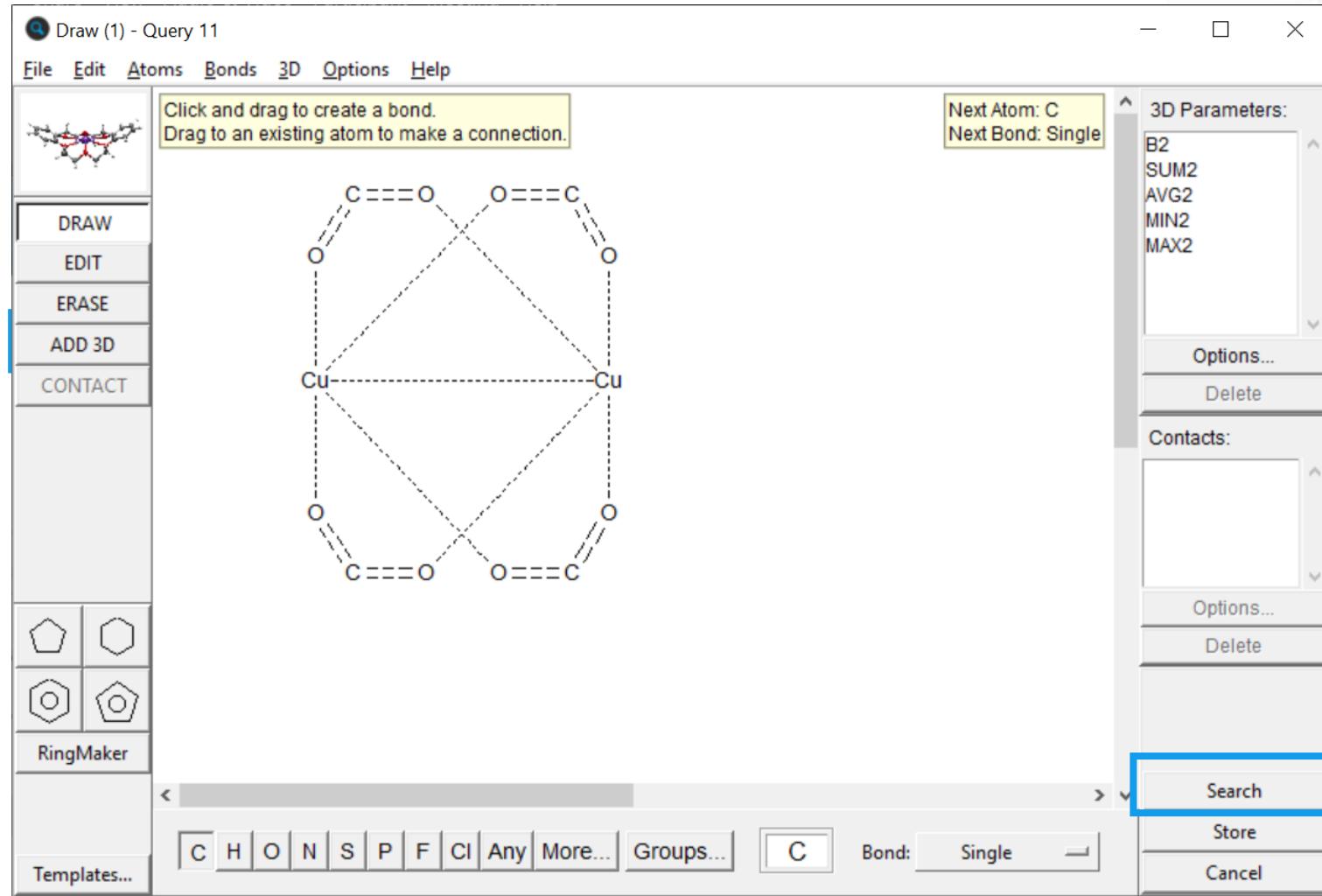
Biggest chemical unit only

Make each chemical unit a separate query

Include all chemical units in a single query

Cancel OK

# 3D searching

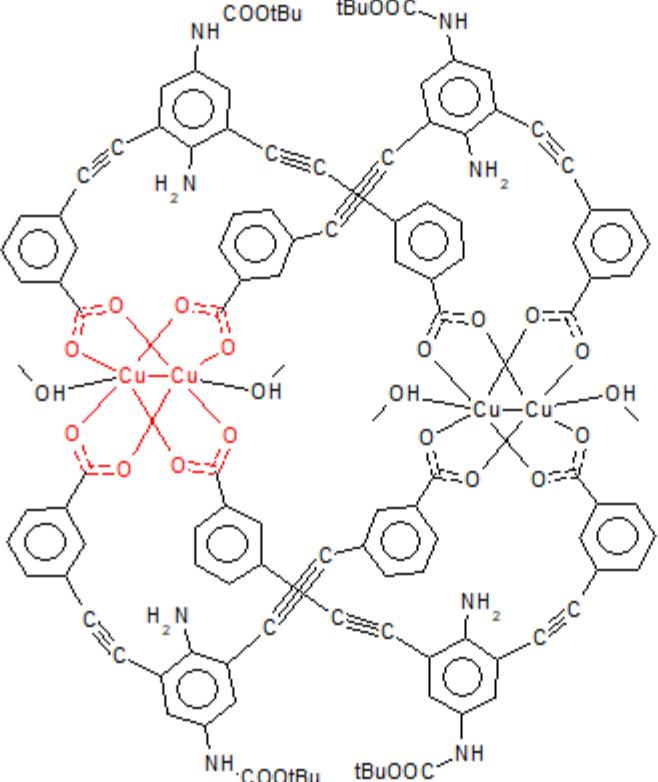


1. Draw your substructure
2. Click Add 3D
3. Select parts of the structure you are interested in
4. Define the parameters you want to analyse
5. Search or store your query
6. View the results
7. Analyse the hit list

All Text  
Author/Journal  
Chemical  
Crystal  
Experimental  
Diagram  
3D Visualiser  
CSD Internals  
Search Overview

Refcode: JUZTOQ

CSD version 5.42 updates (Feb 2021)



Parameters  
B2  
2.603  
SUM2  
2.603  
AVG2  
2.603  
MIN2  
2.603  
MAX2  
2.603

JUZTOQ

Analyse Hitlist

- CUYYUT
- JUZTOQ
- JUZTUW
- LOYYED
- LUYVEJ
- OKUSAR
- PYCUAC
- UJOKUC
- UJUNOF
- ABUXU
- ACACC
- ACACC
- ACACC
- ACAPC
- ACGLC
- ACNICU
- ACURC
- ACURC
- ACURC
- ACURC
- AGUTA
- AHEDO
- AMACU

Visualise Structures

Analyse Data

Select All

Deselect All

Invert Selection

Search highlighted in 2D

Parameters displayed

Analyse in Mercury

Select the items you wish to include from the choices below

File type: Mercury data file (.c2m)

 Include Defined Parameters

## Crystal data

- R-factor
- Space Gp. Symbol
- Space Gp. Number
- No. of Coordinates
- Z Value
- Z Prime
- Study Temp.
- Calc. Density

## Cell data

- a
- b
- c
- Alpha
- Beta
- Gamma
- Cell Volume
- Reduced Cell a
- Reduced Cell b
- Reduced Cell c
- Reduced Cell Volume
- Reduced Cell Alpha
- Reduced Cell Beta
- Reduced Cell Gamma

## Other

- Publication Year
- Unique Chemical Units
- Multiplier Sum
- Compound Name

829 h

0%

Reset Cancel Analyse in Mercury

Multiple Hits: Show    of 2 Show Parameters Show terminal carbons

Use as Query...

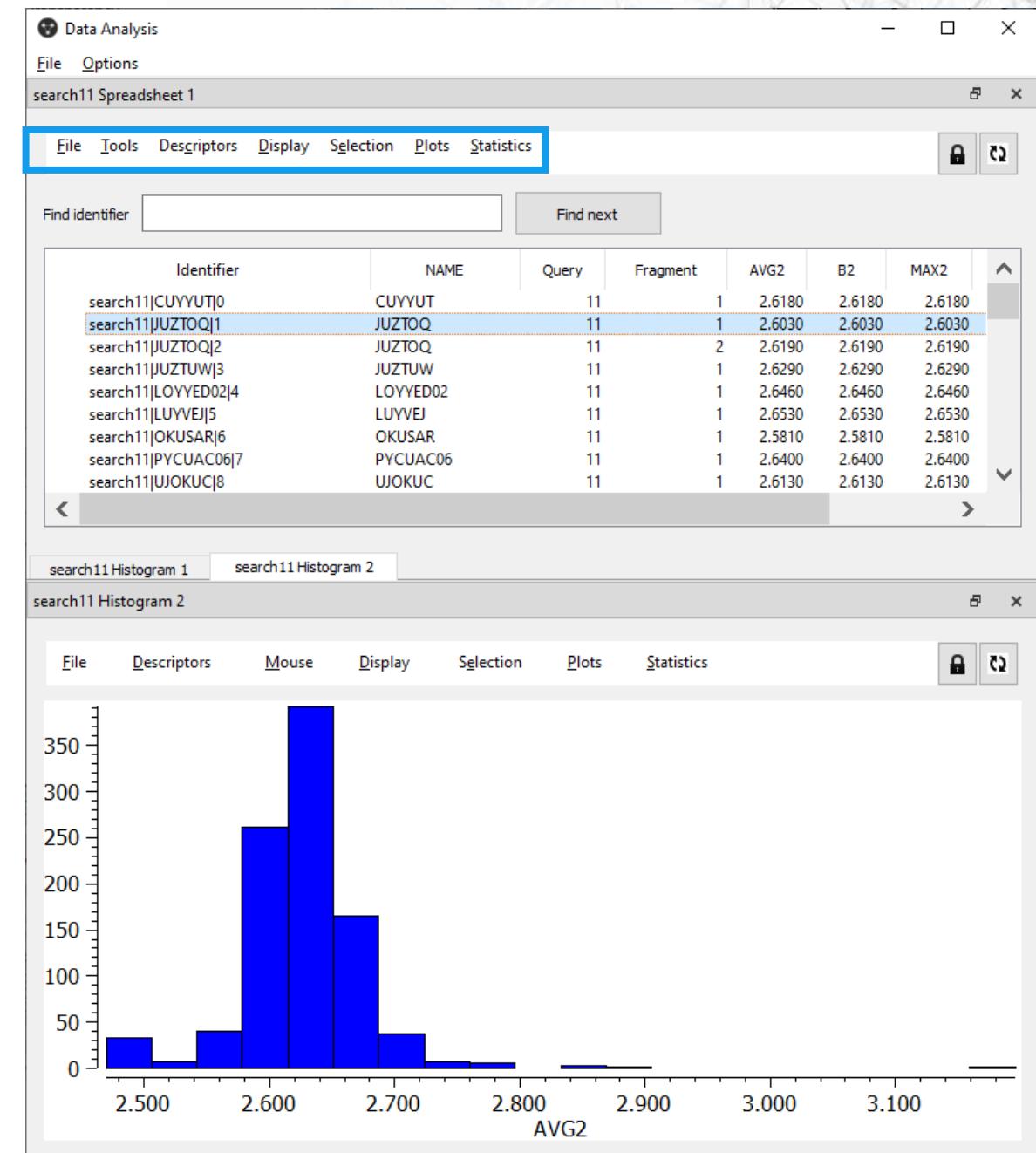
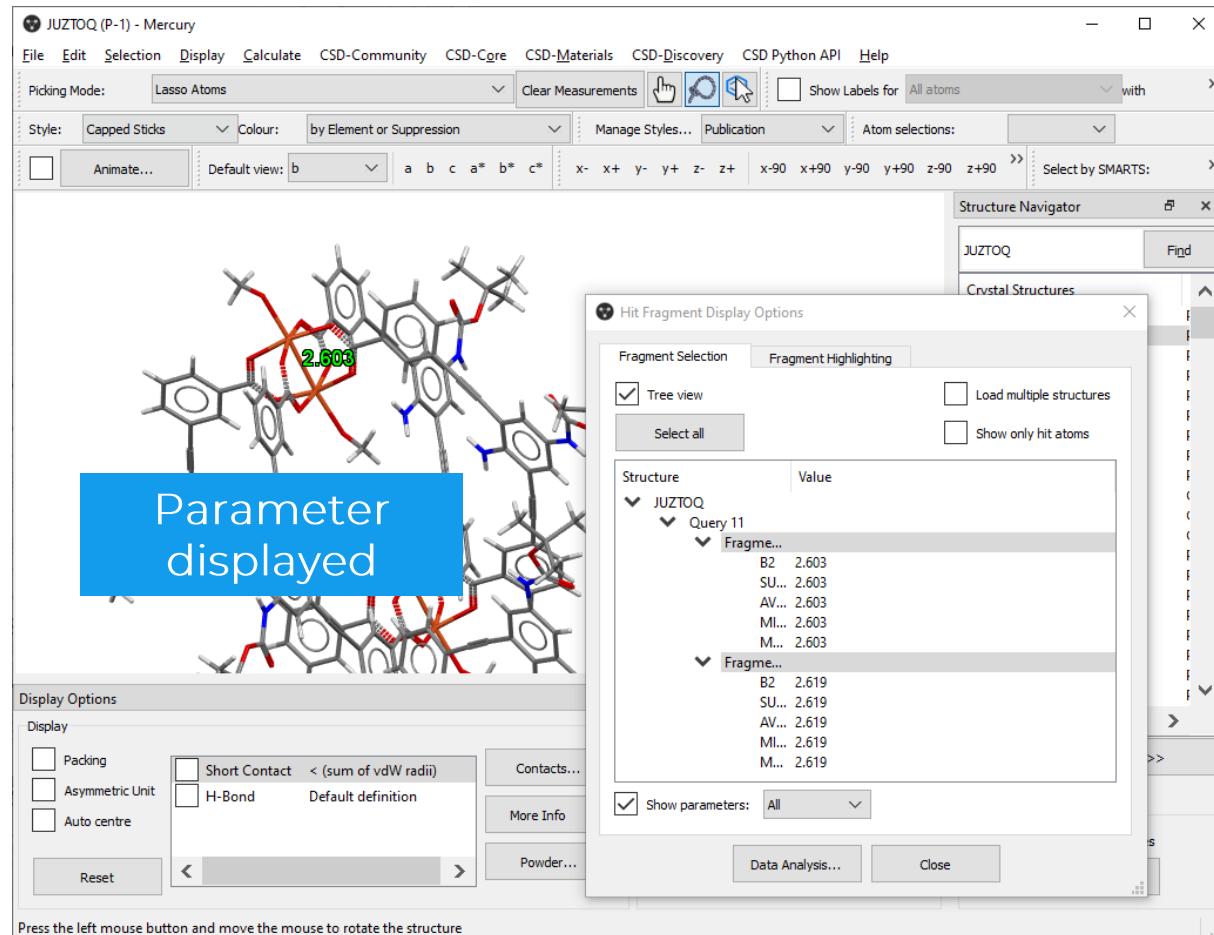
Detach

100%

Stop Search

CCDC

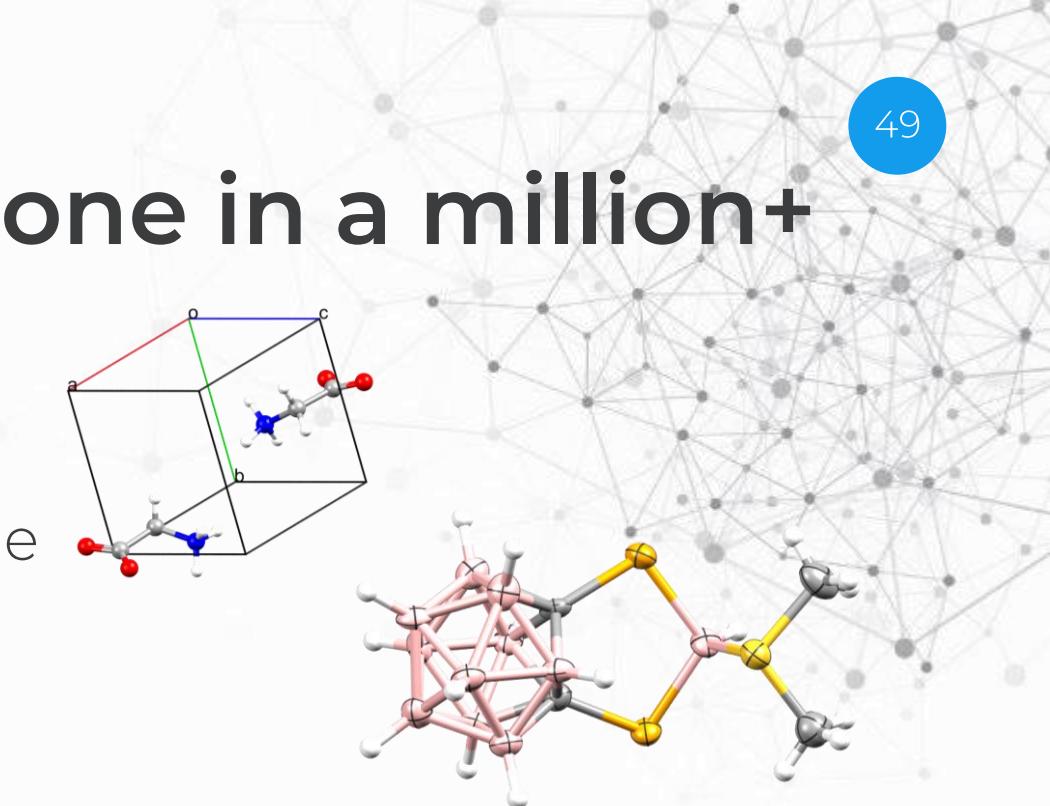
# Data analysis



# CSD Subsets: Helping find one in a million+

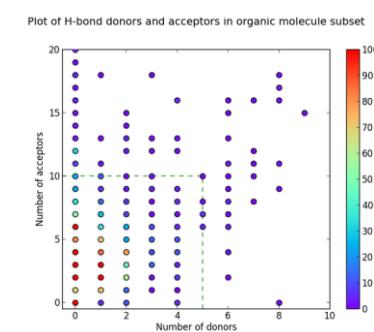
Why provide subsets?

- The CSD contains a huge range of compounds; subsets allow easy access to the most relevant structures of interest
- Metal-organics in particular can be hard to name and draw consistently
- Gives CSD users the benefit of analysis from in-house and external experts
- Convenient starting point for analysis using CSD or 3<sup>rd</sup>-party tools

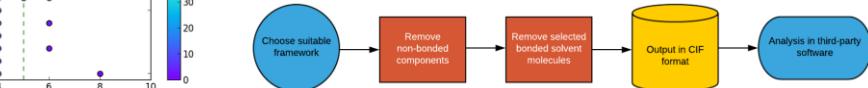


DRUGBANK

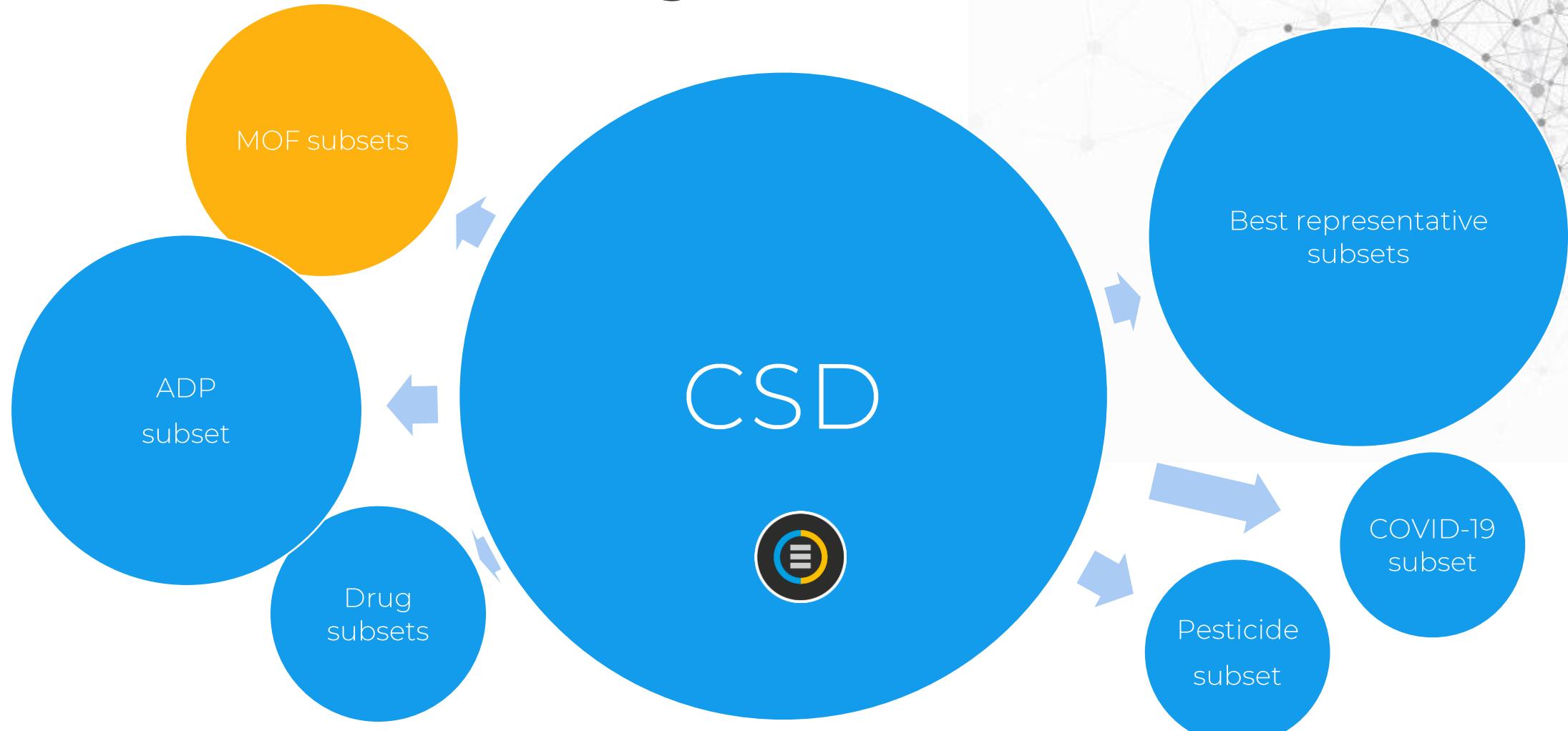
PPDB



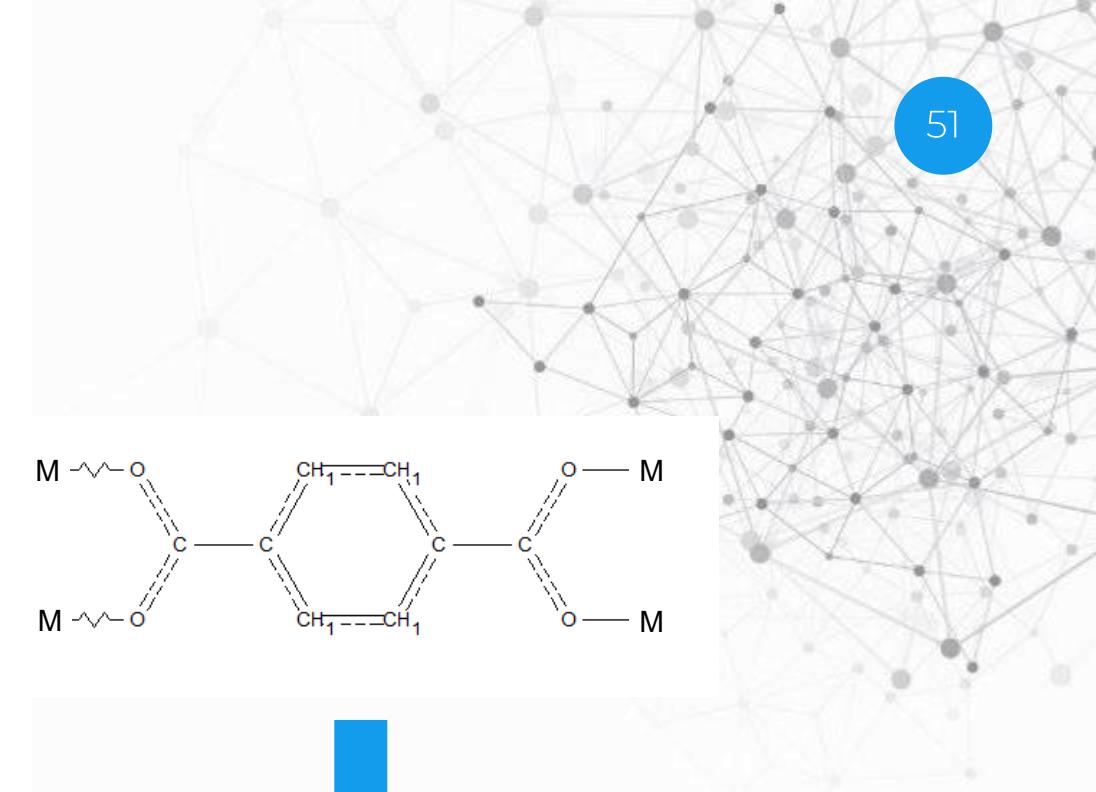
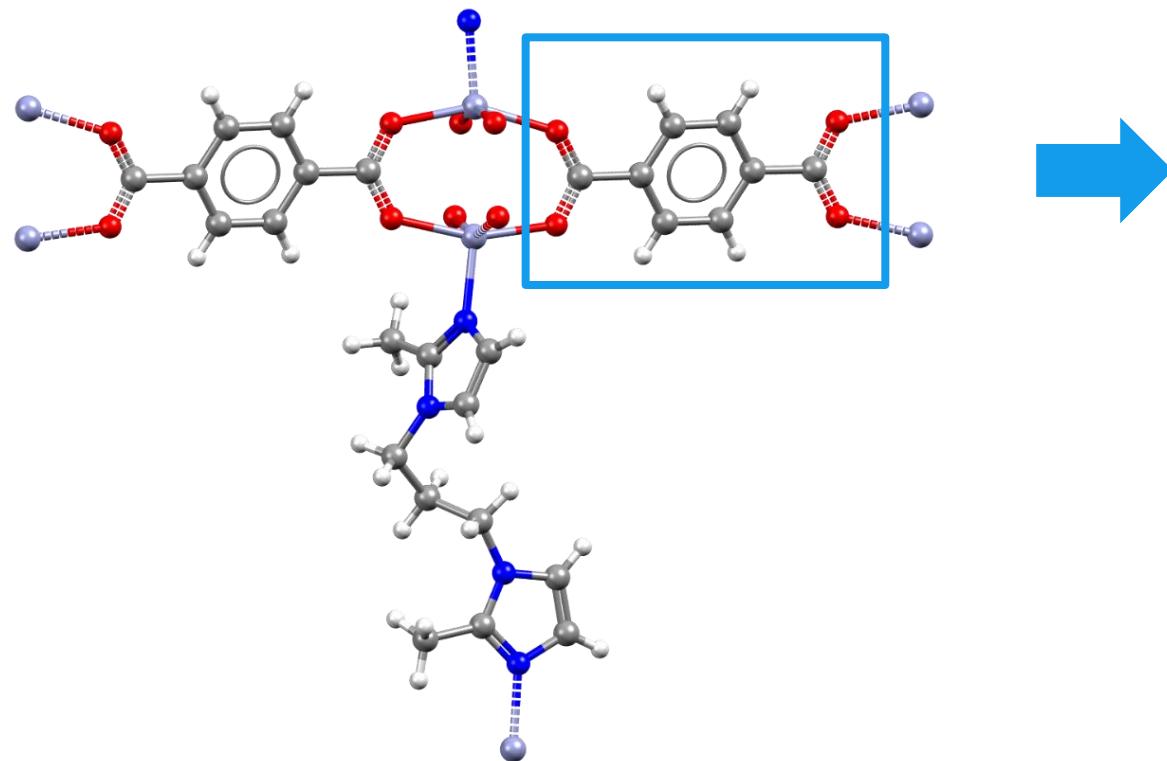
CCDC



# CSD subsets – targeted datasets

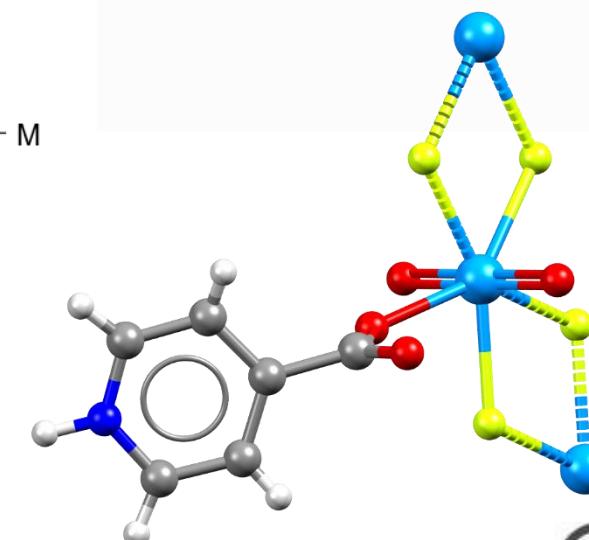
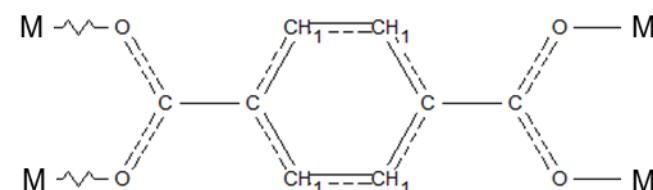
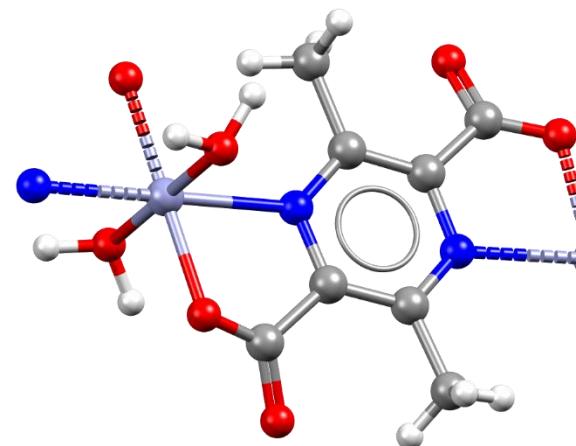
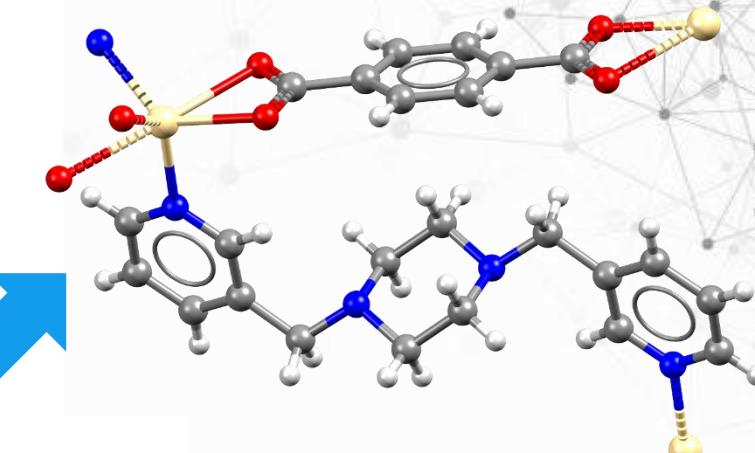
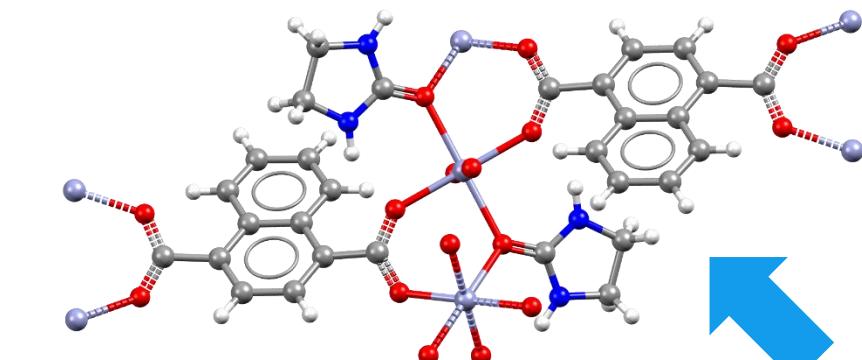


# CSD MOF subset creation



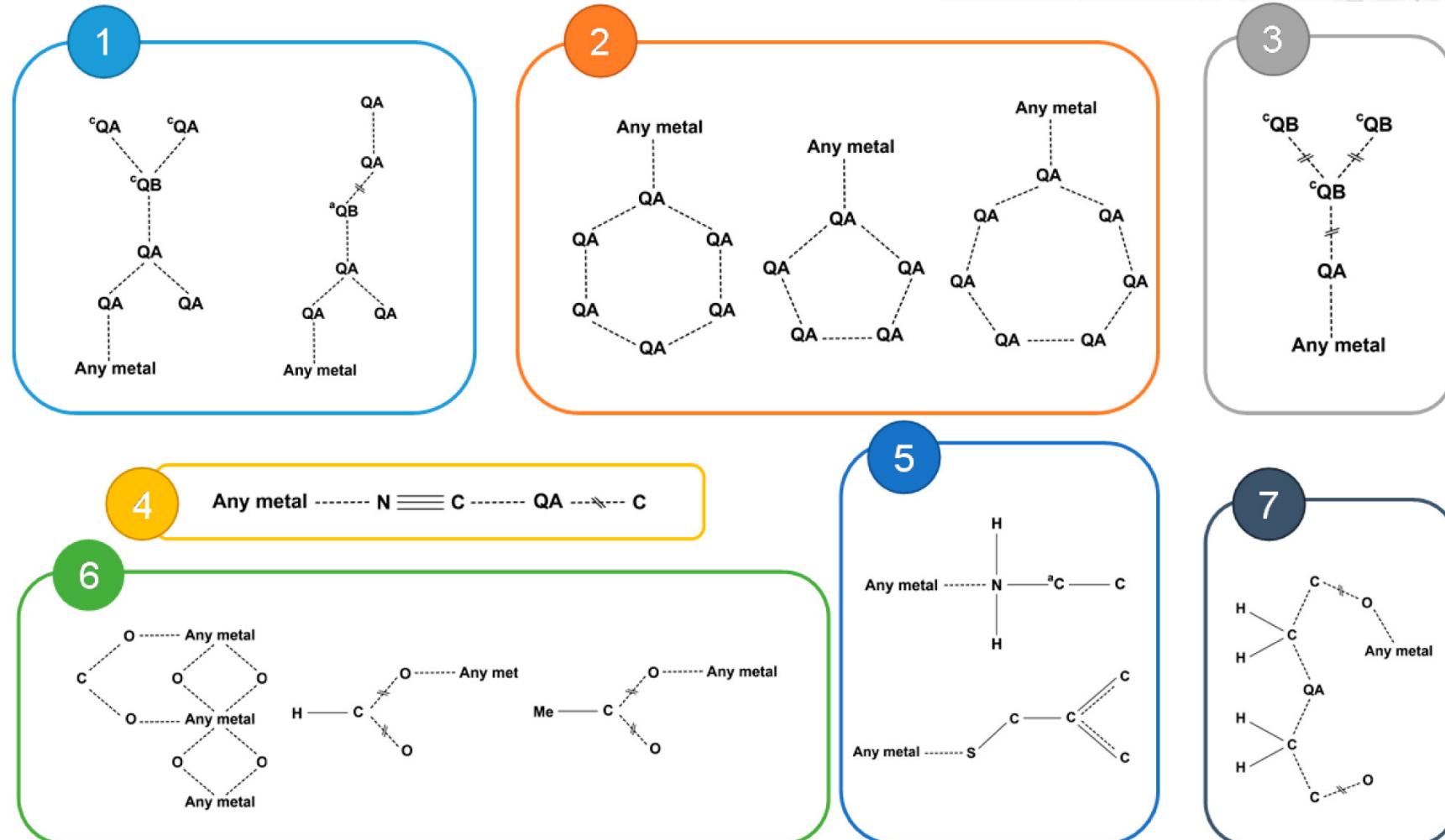
CCDC

# CSD MOF subset creation



CCDC

# CSD MOF subset creation – Draw Queries



# Using the subsets

CCDC ConQuest (1) : CSD\_COVID-19\_subset [Refcode List] - CSD\_COVID-19\_subset.gcd

File Edit Options View Databases Results Help

**Build Queries**

All Text Author/Journal Chemical Crystal Experimental Diagram 3D Visualiser CSD Internals Search Overview

Entries in CSD version 5.41 updates Entries in CSD version 5.41 (November 2019)

Lists in CSD version 5.41 (November 2019) >

Available Databases...

Author(s) Reference Publication DOI Deposition Formula Compound Synonym CCDC Class Spacegroup Cell Reduced Cell

Author(s) Reference Acta Crystallogr., Sect. E: Struct. Rep. Online DOI 10.1107/S1600536807020361 Deposition CCDC 651371; IUCr FL212 Formula C<sub>22</sub>H<sub>29</sub>F O<sub>5</sub> Compound 9α-Fluoro-16α-methyl-11,20-dione Synonym (11β,16α)-9-Fluoro-11,17,21-trihydroxy-16-methylpregna-1,4-dien-3,20-dione; Decadron; Dexamethasone; DrugBank: DB00443; PDB Chemical Component code: DEX

CCDC Class Steroids

Name: P212121 Number: 19

Type: Refcode List Title: unknown Percent: Unknown

Time: Thu Feb 18 11:16:17 2021

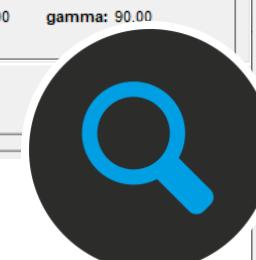
File: ...subsets\Non-disordered\_MOF\_subset.gcd

Cell:

a:	10.364(1)	b:	16.157(1)	c:	23.206(1)
alpha:	90.00	beta:	90.00	gamma:	90.00
Volume:	3885.871				

Reduced Cell:

a:	10.364	b:	16.157	c:	23.206
alpha:	90.00	beta:	90.00	gamma:	90.00
Volume:	3885.871				



Search Name: search4

Available Databases:  Show Updates separately  CSD version 5.42 (November 2020) + 1 update  CSD version 5.42 (November 2020)

**Filters Advanced Options**

3D coordinates determined  
 R factor <= 0.05  
 <= 0.075  
 <= 0.1  
 Only  Non-disordered  
 Disordered  
 No errors  
 Not polymeric  
 No ions  
 Only  Single crystal structures  
 Powder structures  
 Only  Organics  
 Organometallic

You can search complete database(s) or a subset (e.g., hits found in a previous search)

Select Subset Clear Subset

Single query being used. Search will find structures:  
where this query is true:  
Query 5

Start Search Cancel Reset

Restrict Search

Restrict Search by Refcode

Current Restrictions

Type: Refcode List Title: unknown Percent: Unknown

Time: Thu Feb 18 11:16:17 2021

File: ...subsets\Non-disordered\_MOF\_subset.gcd

Database CSD version 5.42 updates (Feb 2021)  
1421 Refcodes

Database CSD version 5.42 (November 2020)  
80303 Refcodes

Database CSD version 5.42 (November 2020)  
0 Refcodes  
Entries not found  
0 Refcodes

CSD Version... Select to restrict based on CSD version

OK Clear Cancel

CCDC



## 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 gsd 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 re
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
        setattr(self, format + '_header')()
        for e in self.rdr:
            setattr(self, format + '_line')(e)
        setattr(self, format + '_footer')()

    def csv_header(self):
        data = '/'.join([
            'Identifier',
            'Title',
            'Literature Ref',
            'Formula',
            'Compound Name',
            'Synonyms'
        ])
        ...

```

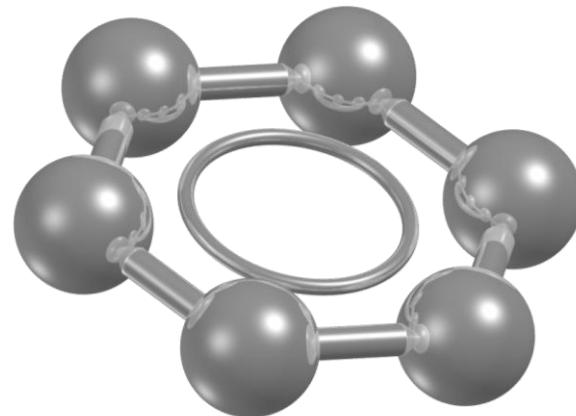
# ***Explore More: More advanced uses***

- Not enough time to explore all the insights that you can get from metal-organic structures in the CSD
- But here are a selection of more advanced tips and tricks and case studies

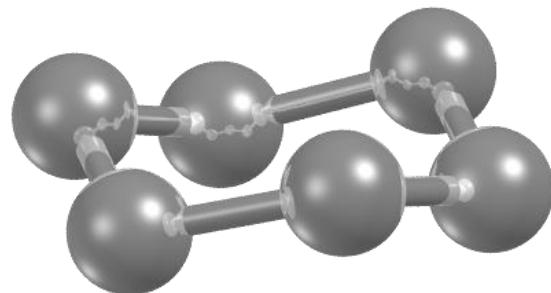
The session has resumed recording

CCDC

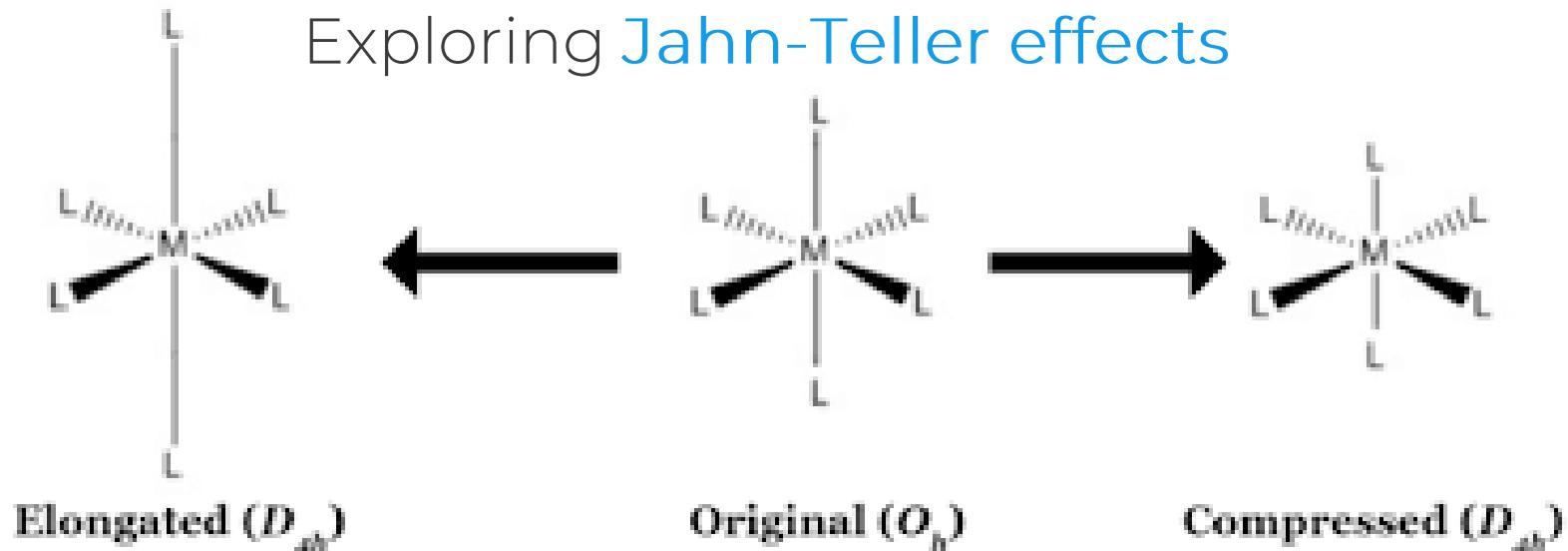
# Exploring chemical concepts



BENZEN07

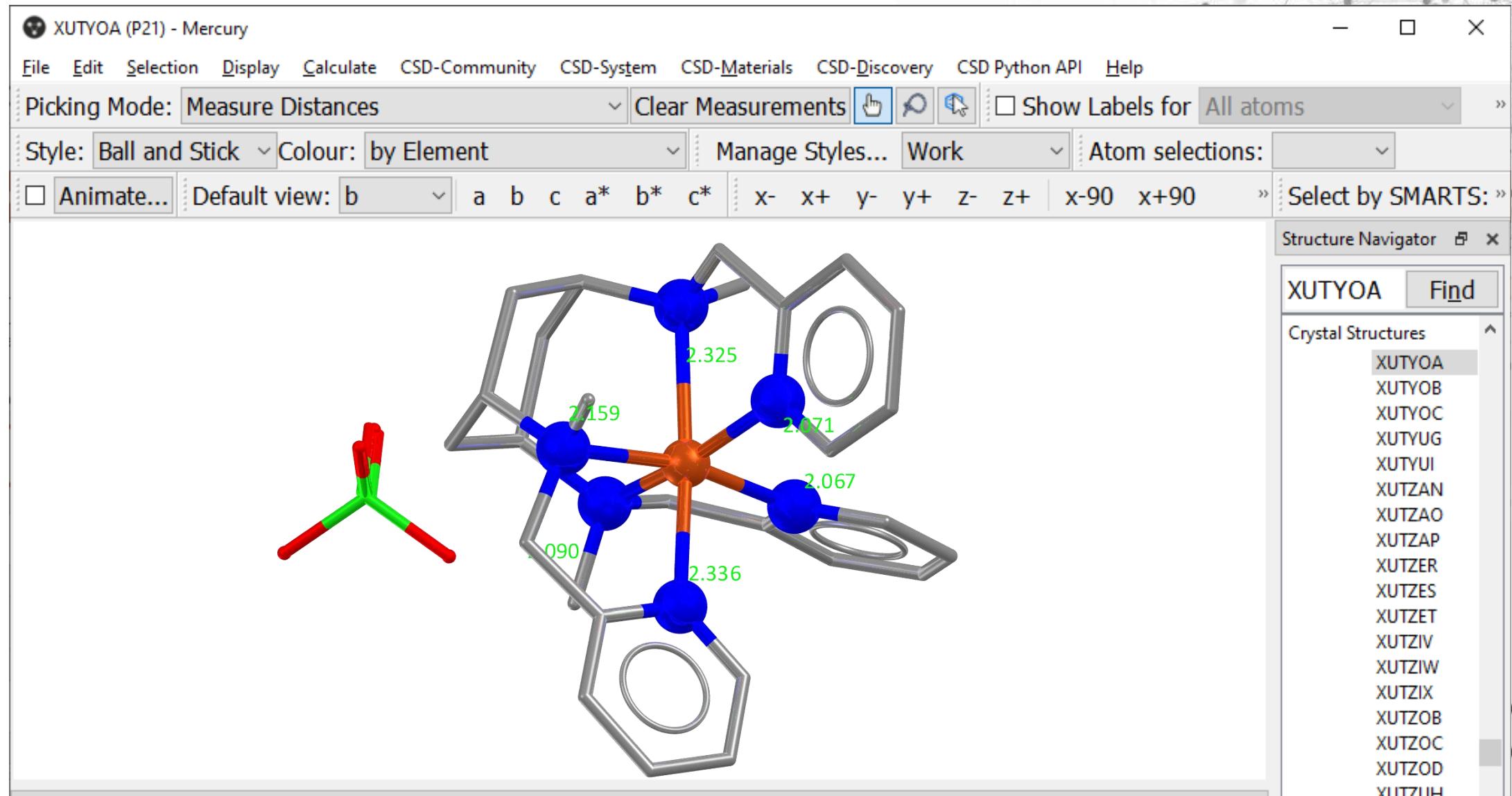


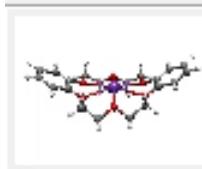
FUHVEM



The Jahn–Teller effect is most often encountered in octahedral complexes of the transition metals.<sup>[3]</sup> The phenomenon is very common in six-coordinate copper(II) complexes.

# Using Mercury to see chemical concepts





Click and drag to create a bond.  
Drag to an existing atom to make a connection.

Next Atom: C  
Next Bond: Single

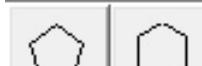
DRAW

EDIT

ERASE

ADD 3D

CONTACT



RingMaker

Templates...

3D Parameters:

Options...

Delete

Contacts:

Options...

Delete

Search

Store

Cancel

C H O N S P F Cl Any More... Groups...

C

Bond: Single

All Text

Author/Journal

Chemical

Crystal

Experimental

Diagram

3D Visualiser

CSD Internals

Search Overview

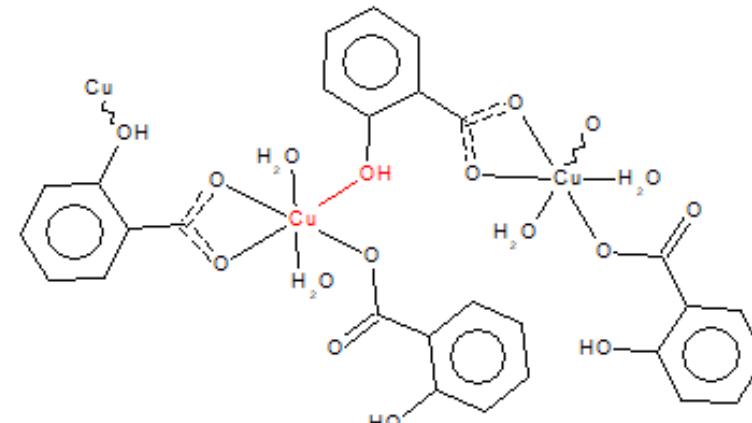
Refcode: ASALCU03

CSD version 5.41 updates (Mar 2020)

Parameters

DIST1

2.237

Multiple Hits: Show    of 11 Show terminal carbons

Use as Query

ASALCU03

Analyse Hitlist

- ASALCU03
- ASALCU05
- BEYRAY07
- BOVNEH01
- BOWNOT
- BOZXEW
- BUBMOD
- BUCBIN
- BUFZUA

Visualise Structures

Analyse Data

Select All  
Deselect All  
Invert Selection

## Analyse in Mercury

Select the items you wish to include from the choices below

File type: Mercury data file (.c2m)

 Include Defined Parameters

## Crystal data

 R-factor       Space Gp. Symbol       Space Gp. Number       No. of Coordinates Z Value       Z Prime       Study Temp.       Calc. Density

## Cell data

<input type="checkbox"/> a	<input type="checkbox"/> b	<input type="checkbox"/> c
<input type="checkbox"/> Alpha	<input type="checkbox"/> Beta	<input type="checkbox"/> Gamma
<input type="checkbox"/> Reduced Cell a	<input type="checkbox"/> Reduced Cell b	<input type="checkbox"/> Reduced Cell c
<input type="checkbox"/> Reduced Cell Alpha	<input type="checkbox"/> Reduced Cell Beta	<input type="checkbox"/> Reduced Cell Gamma

## Other

 Publication Year       Unique Chemical Units       Multiplier Sum       Compound Name

0%

Reset

Cancel

Analyse in Mercury

**Data Analysis**

File Options

search2 Spreadsheet 1

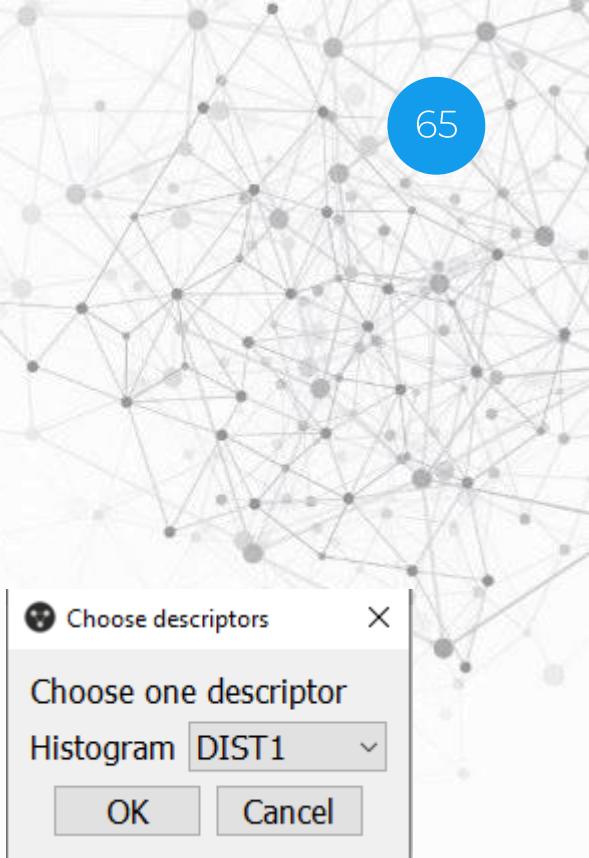
File Tools Descriptors Display Selection Plots Statistics

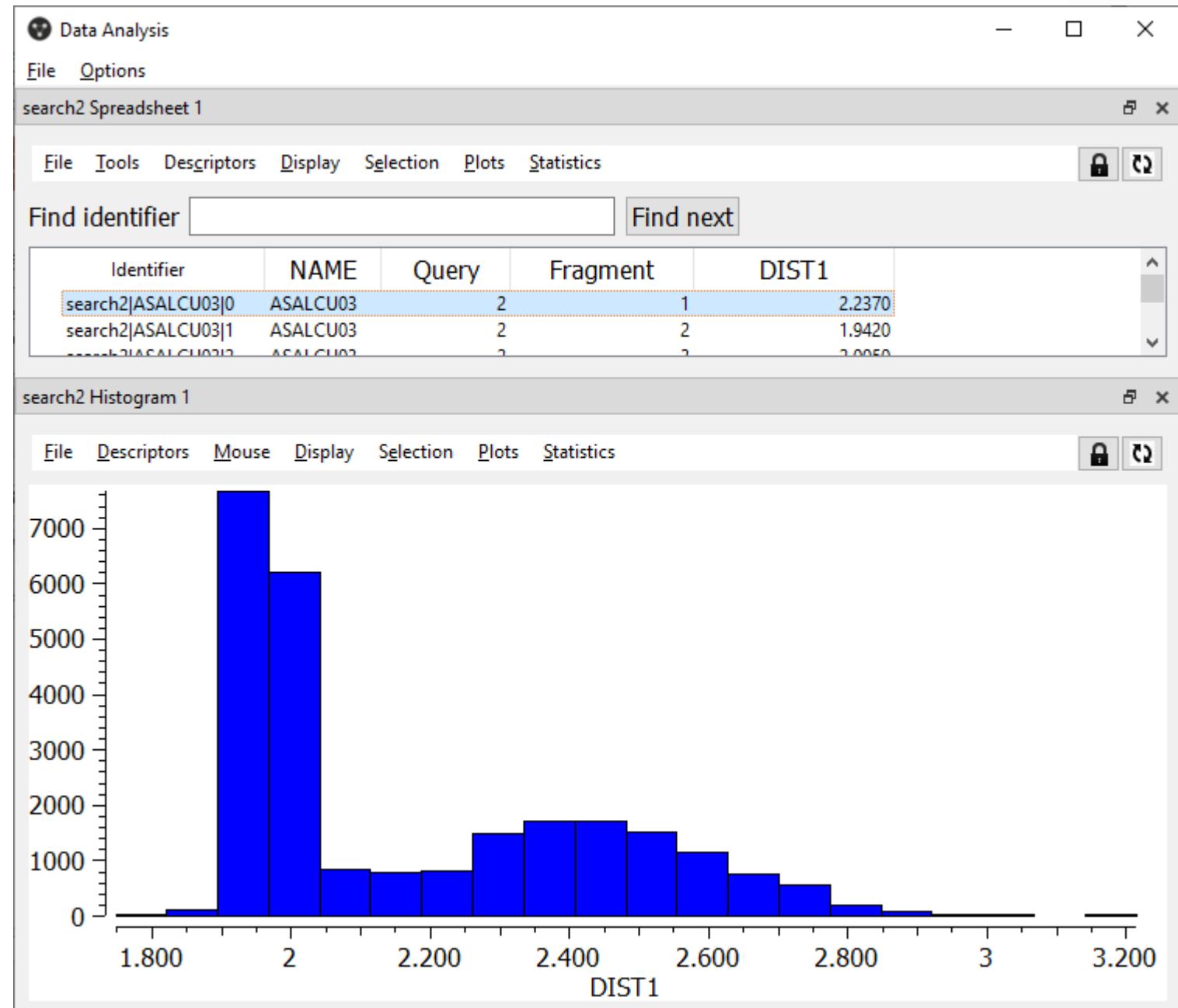
Find identifier

Identifier	NAME	Qu	DIST1
search2 ASALCU03 0	ASALCU03		2.2370
search2 ASALCU03 1	ASALCU03		1.9420
search2 ASALCU03 2	ASALCU03		2.0050
search2 ASALCU03 3	ASALCU03		2.7280
search2 ASALCU03 4	ASALCU03		1.9550
search2 ASALCU03 5	ASALCU03		1.9760
search2 ASALCU03 6	ASALCU03		1.9420
search2 ASALCU03 7	ASALCU03		2.0080
search2 ASALCU03 8	ASALCU03		2.7310
search2 ASALCU03 9	ASALCU03		1.9370
search2 ASALCU03 10	ASALCU03		1.9610
search2 ASALCU05 11	ASALCU05		1.9440
search2 ASALCU05 12	ASALCU05		1.9690
search2 ASALCU05 13	ASALCU05		1.9320
search2 ASALCU05 14	ASALCU05		1.9630
search2 ASALCU05 15	ASALCU05		1.9220
search2 ASALCU05 16	ASALCU05		1.9480
search2 ASALCU05 17	ASALCU05		2.2460
search2 ASALCU05 18	ASALCU05		2.2610
search2 BEYRAY07 19	BEYRAY07		1.9890

Histogram  
Polar histogram  
Scatterplot  
Polar scatterplot  
Heat plot

next

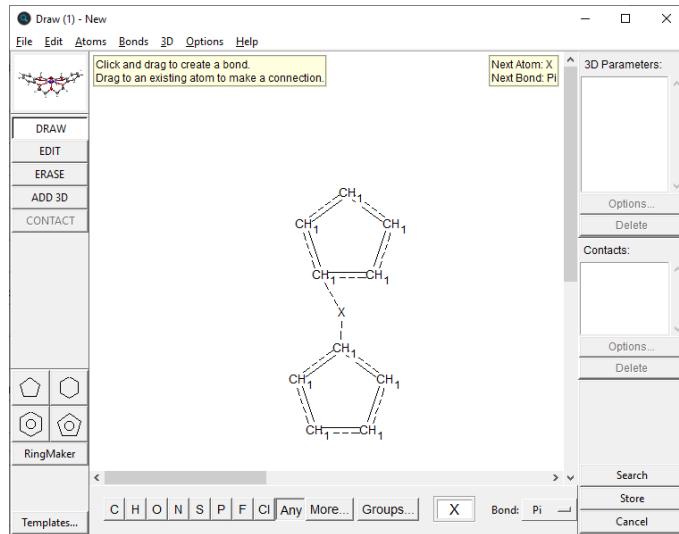




You could use this method to see what other metal atoms exhibit Jahn-Teller distortions

# Catalysts

- Searching for Journal
  - Organometallics
  - Other catalysis based journals
- All-text search
- Similarity search on results
- Structure/Draw searches



The screenshot shows the CCDC ConQuest software interface. It displays two search queries: 'Text Search Catalyst' (Query 4) and 'Text Search catalysis' (Query 5). Both queries have a 'must have' condition. The 'must have at least one of' section contains 'Query 4' and 'Query 5'. There are 'Edit...', 'Delete', and 'Search' buttons for each query.

The screenshot shows the CCDC ConQuest software interface displaying search results for the term 'catalyst'. The results list includes:  
Cat.Sci.Tech. [2011-2019]  
Catal.Lett [2001-2018]  
Catal.Today [2004-2019]  
Catalysis Communications [2008-2019]  
Catalysts [2013-2019]  
Below the list are filters for 'I, 1.2 etc.', 'Page (212,6-A etc.)', and 'Year (1998, 2001 etc.)'. The right side of the interface shows a detailed view of a hit, including its Refcode: ATEDAG, Formula: C<sub>36</sub>H<sub>48</sub>N<sub>10</sub>O<sub>2</sub>Y<sub>2</sub>, Name: bis(μ-2-((pyrrol-2-ylmethylene)amino)-N-(2-((pyrrol-2-ylmethylene)amino)ethyl)ethanaminobis(tetrahydrofuran)-di-yttrium(III)), Synonyms: FEPHAK, FEYJUR, FEYKAY, GEJGEJ, GELKEP, GELLIU, HOHYIM01, HOHYOS01, ICUQED, ICUQED01, IFOWIK, IKORIK, IKORIK01, IKORIK03, IKORIK04, JECMUA, IFCNAH, and a note about being air-sensitive and moisture-sensitive.

