Accelerating Porous Materials Design and Development

JAICI Webinar, July / August 2026

Dr Christopher J. Kingsbury

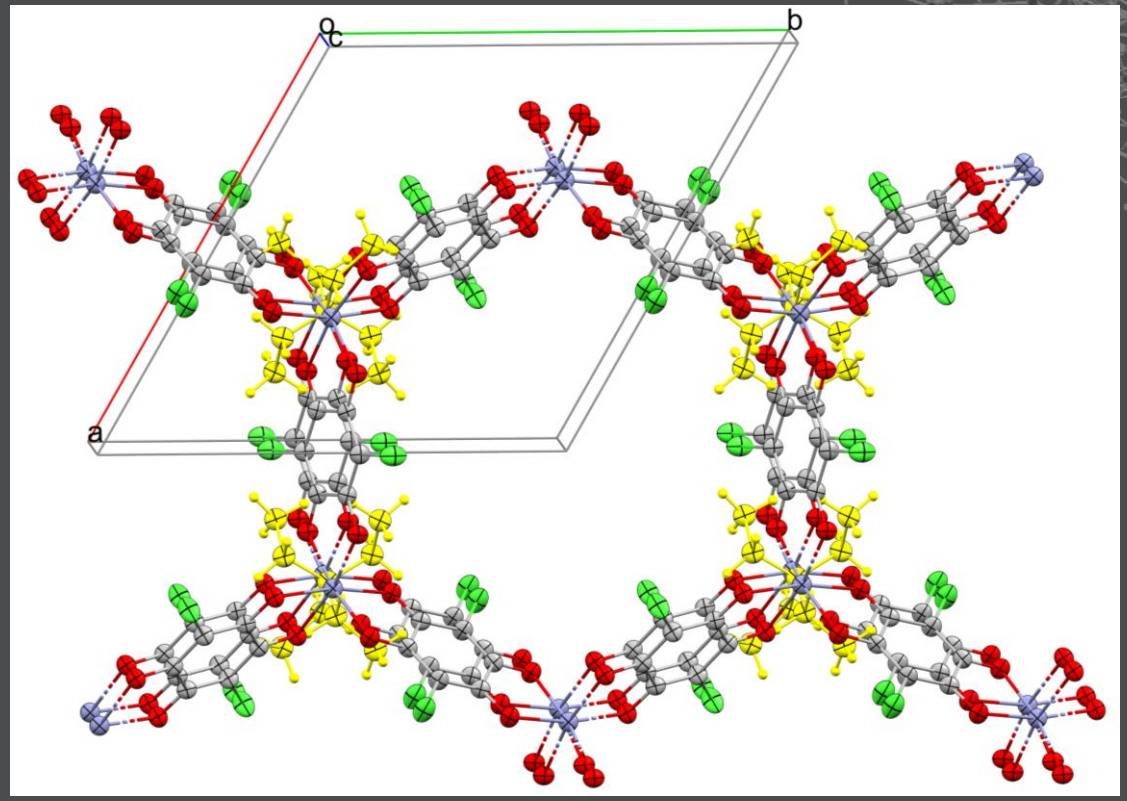
Research Scientist

CCDC

Overview

In this webinar we will discuss:

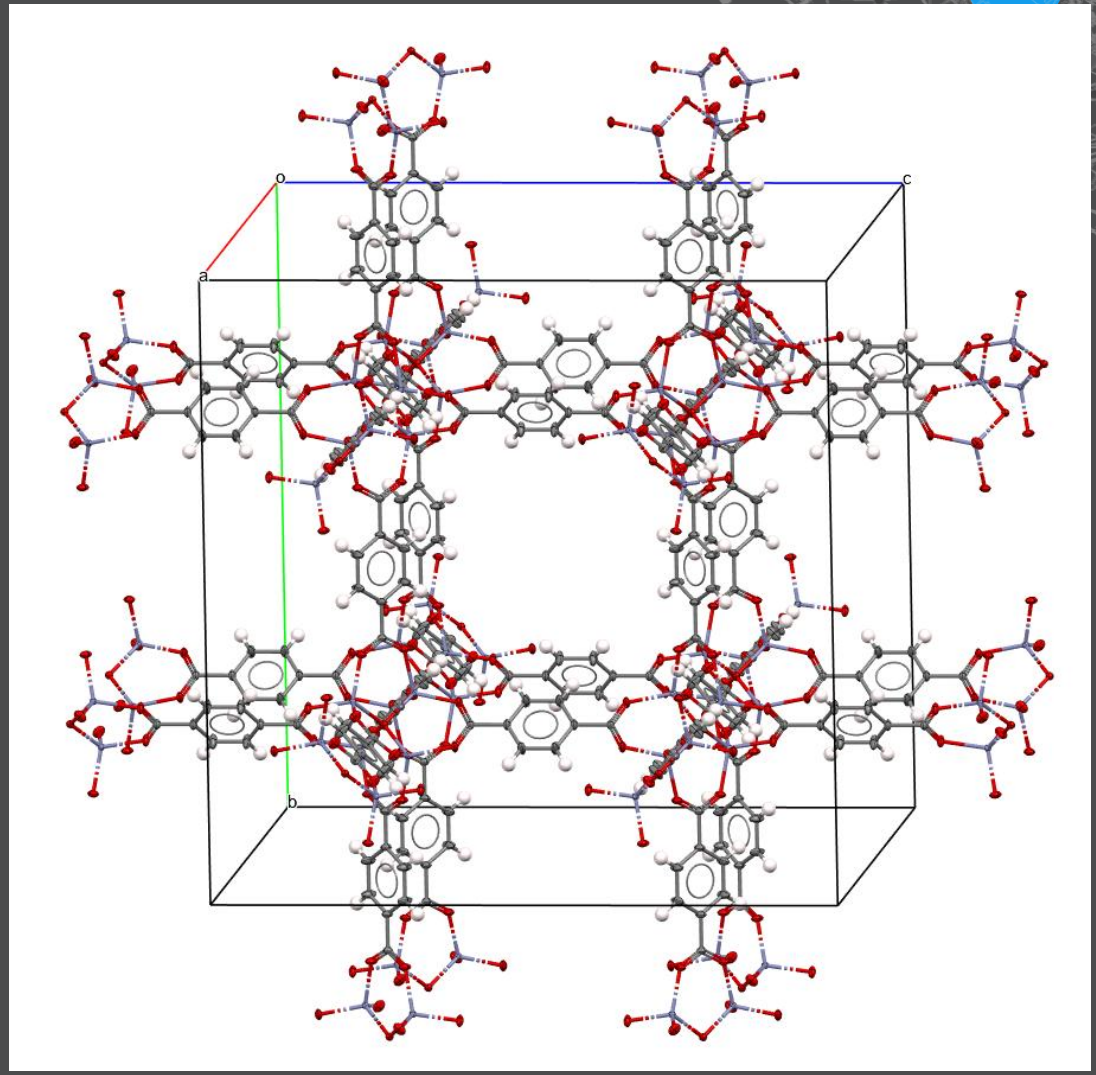
- Latest news, updates and events
- CSD-Frameworks, including how to:
  - Search for structures
  - Visualize 3D structures of MOFs
  - Analyse solvates contained within structures
  - Calculate voids and pores
  - Simulate powder diffraction patterns
  - Compare crystal structures
- Q&A – the floor is yours



CCDC

What is the CSD?

- The CSD is the largest collection of experimental crystal structures
- We also make software to find and investigate these structures
- ccdc.cam.ac.uk/solutions/whats-new/

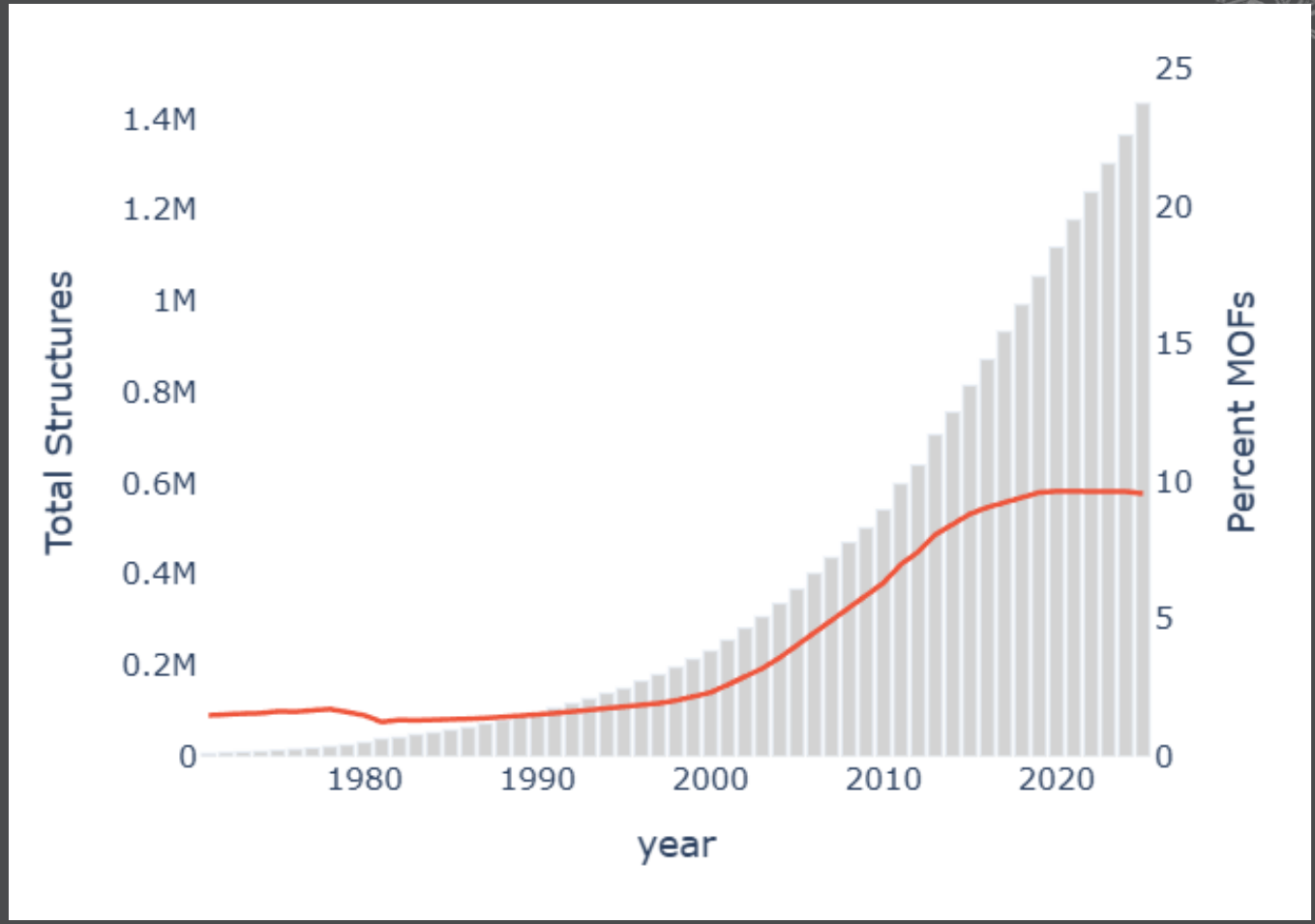


- MOF-5 / “MIBQAR”

CCDC

The Cambridge Structural Database

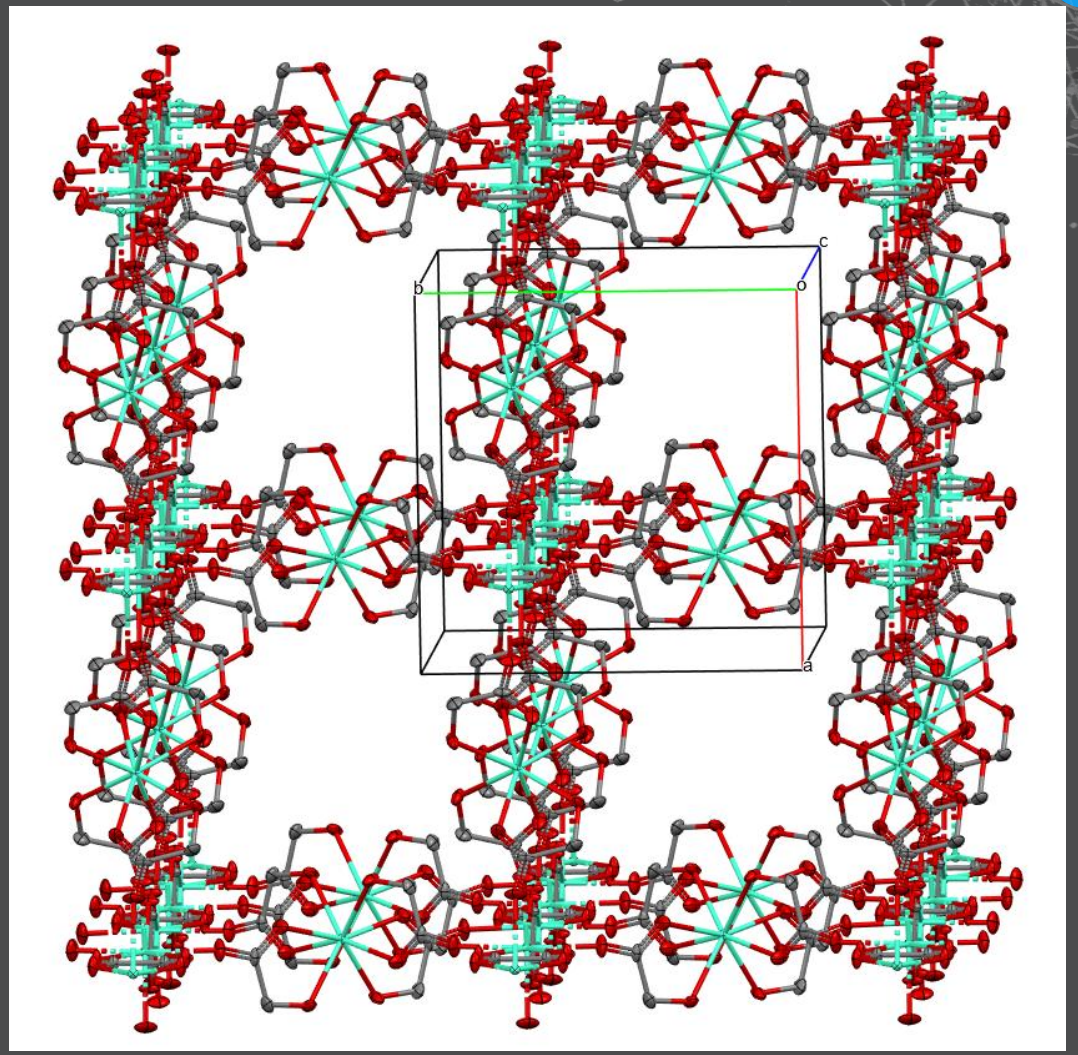
- >1.4 million experimentally determined crystal structures
- About 10% of those structures are MOFs
- Millions of structural parameters can be mined to generate insights



CCDC

Mercury

- Visualisation and analysis software
- Can output high quality images
- Measure distances and angles
- Create planes and centroids
- Run additional tools

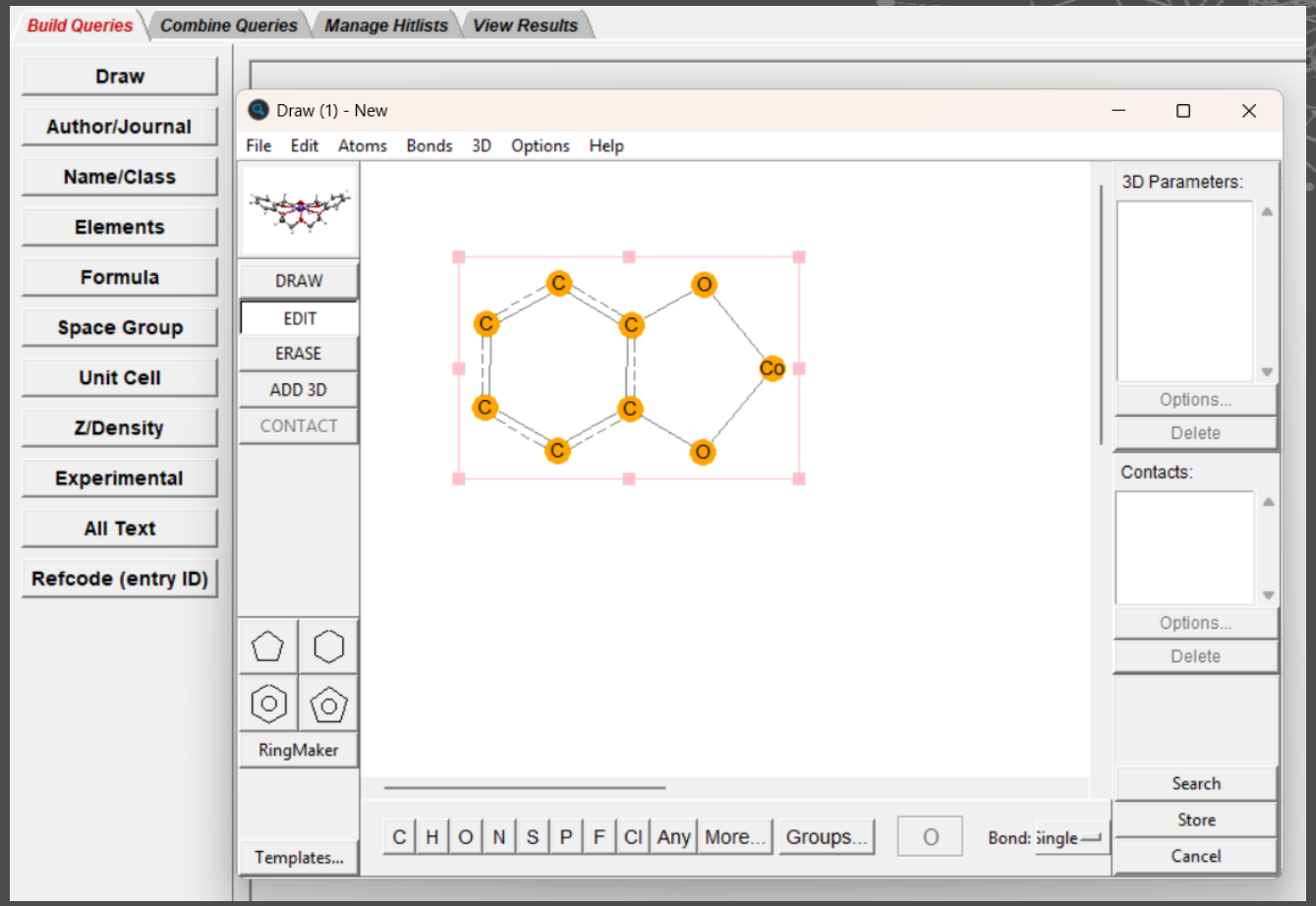


- $Ln_2(C_2O_4)(O_2CCH_2OH)_4$
- “COMFEQ”

CCDC

ConQuest

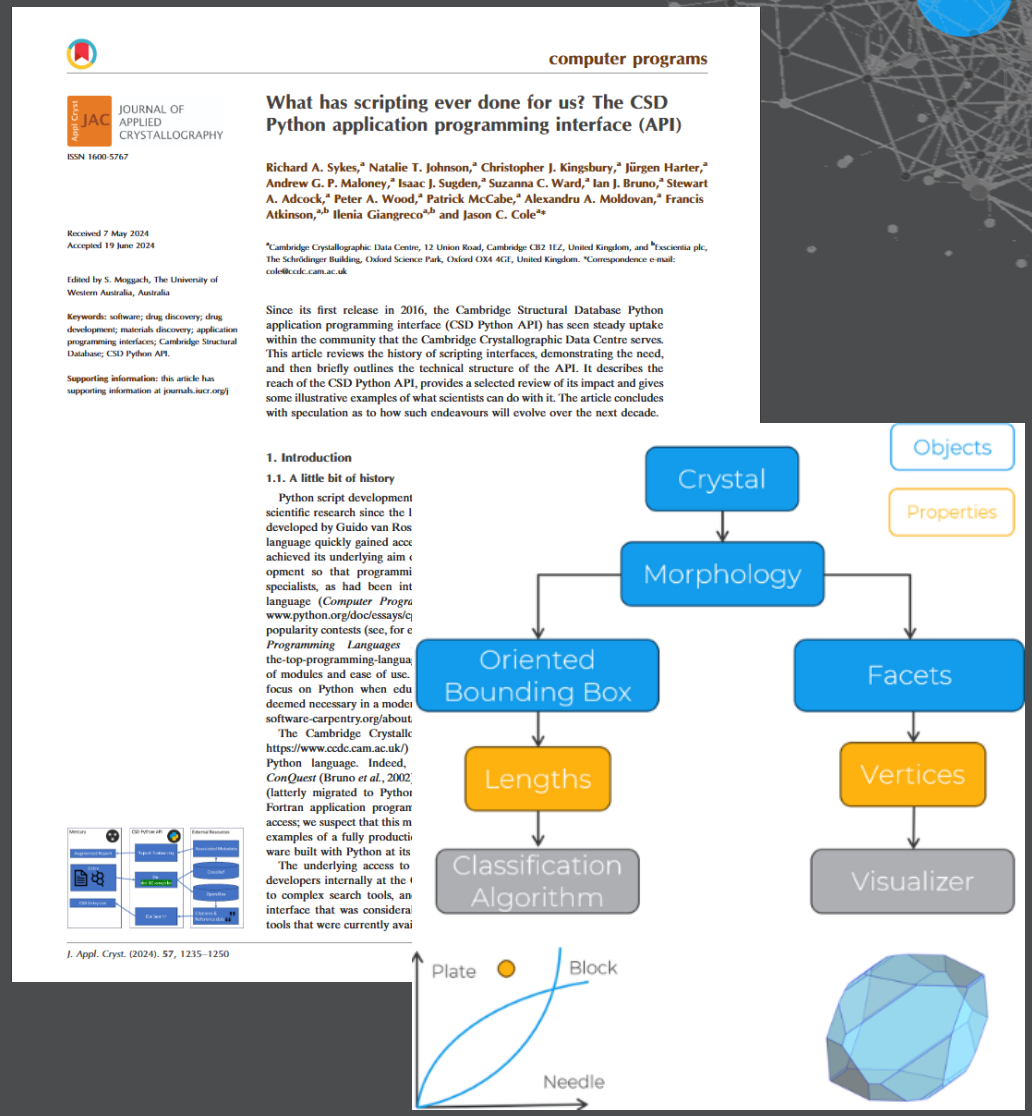
- Search software
- Can find crystals based on molecular structure
- Search text or experimental details
- Define parameters



CCDC

CSD Python API

- Access molecule and crystal data through programs
- Write custom workflows



computer programs

JOURNAL OF APPLIED CRYSTALLOGRAPHY

ISSN 1600-5767

What has scripting ever done for us? The CSD Python application programming interface (API)

Richard A. Sykes,[a] Natalie T. Johnson,[a] Christopher J. Kingsbury,[a] Jürgen Harter,[a] Andrew G. P. Maloney,[a] Isaac J. Sugden,[a] Suzanna C. Ward,[a] Ian J. Bruno,[a] Stewart A. Adcock,[a] Peter A. Wood,[a] Patrick McCabe,[a] Alexandru A. Moldovan,[a] Francis Atkinson,[a,b] Ilenia Giangreco[a,b] and Jason C. Cole[a]*

[a]Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, United Kingdom, and [b]Exscientia plc, The Schrödinger Building, Oxford Science Park, Oxford OX4 4GE, United Kingdom. *Correspondence e-mail: cole@ccdc.cam.ac.uk

Received 7 May 2024
Accepted 19 June 2024

Edited by S. Moggach, The University of Western Australia, Australia

Keywords: software; drug discovery; drug development; materials discovery; application programming interfaces; Cambridge Structural Database; CSD Python API.

Supporting information: this article has supporting information at journals.iucr.org/j

Since its first release in 2016, the Cambridge Structural Database Python application programming interface (CSD Python API) has seen steady uptake within the community that the Cambridge Crystallographic Data Centre serves. This article reviews the history of scripting interfaces, demonstrating the need, and then briefly outlines the technical structure of the API. It describes the reach of the CSD Python API, provides a selected review of its impact and gives some illustrative examples of what scientists can do with it. The article concludes with speculation as to how such endeavours will evolve over the next decade.

1. Introduction

1.1. A little bit of history

Python script development scientific research since the l developed by Guido van Ros language quickly gained acce achieved its underlying aim opment so that programmi specialists, as had been int language (*Computer Progra* www.python.org/doc/essays/c popularity contests (see, for e *Programming Languages* the-top-programming-langua of modules and ease of use. focus on Python when edu deemed necessary in a mode software-carpentry.org/about The Cambridge Crystallo https://www.ccdc.cam.ac.uk/) Python language. Indeed, *ConQuest* (Bruno *et al.*, 2002 (latterly migrated to Pytho Fortran application program access; we suspect that this m examples of a fully producti ware built with Python at its The underlying access to developers internally at the to complex search tools, an interface that was considera tools that were currently avai

J. Appl. Cryst. (2024). **57**, 1235-1250
https://doi.org/10.1107/S1600576724005934

CCDC

Consultancy

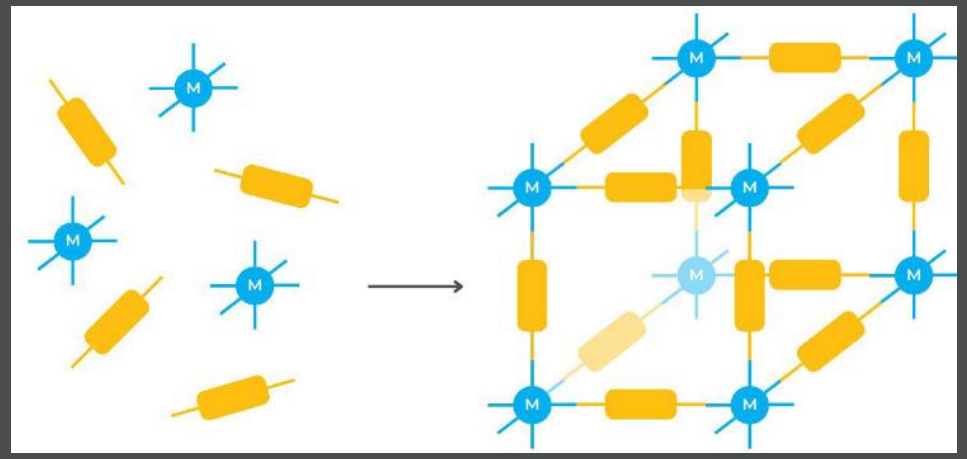
- Any question not answered here
- Scientific Consultancy
- Custom Searching
- Workflows

English: (**hello@ccdc.cam.ac.uk**) for more info
Japanese: **https://www.jaici.or.jp/inquiry/db-service/**

CCDC

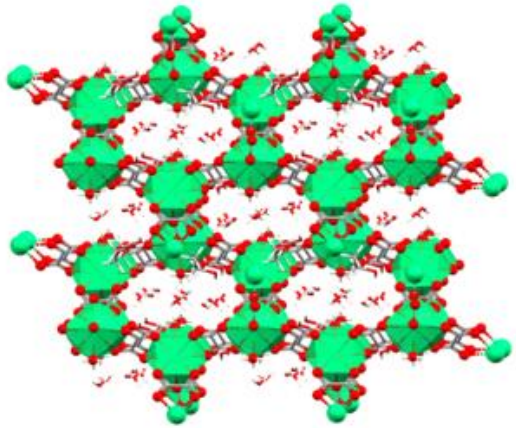
Metal-Organic Frameworks

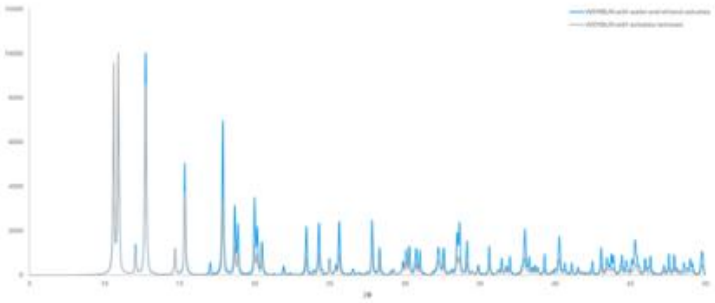
- **“Materials by design”** – components joined together by pre-organised “nodes”
- Can contain voids with tuneable size and different chemistry
- Adsorption changes host and guest properties



CCDC

MOF Analysis Workflow





Search

- Data mining
- Pre-created subsets

- Packing
- Metal polyhedral
- Polymer expansion

Visualize

Analyze

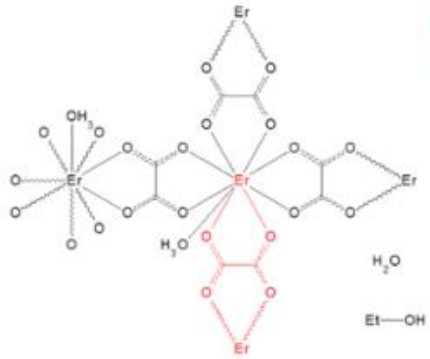
- Void geometry
- Water and solvent volumes

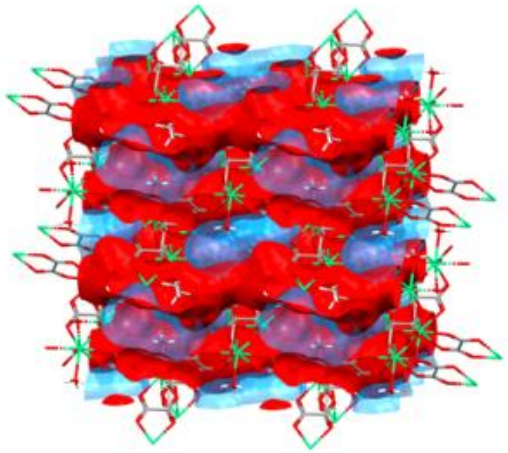
- Calculate from SCXRD structures
- Preferred orientations
- Custom wavelengths

PXRD

API

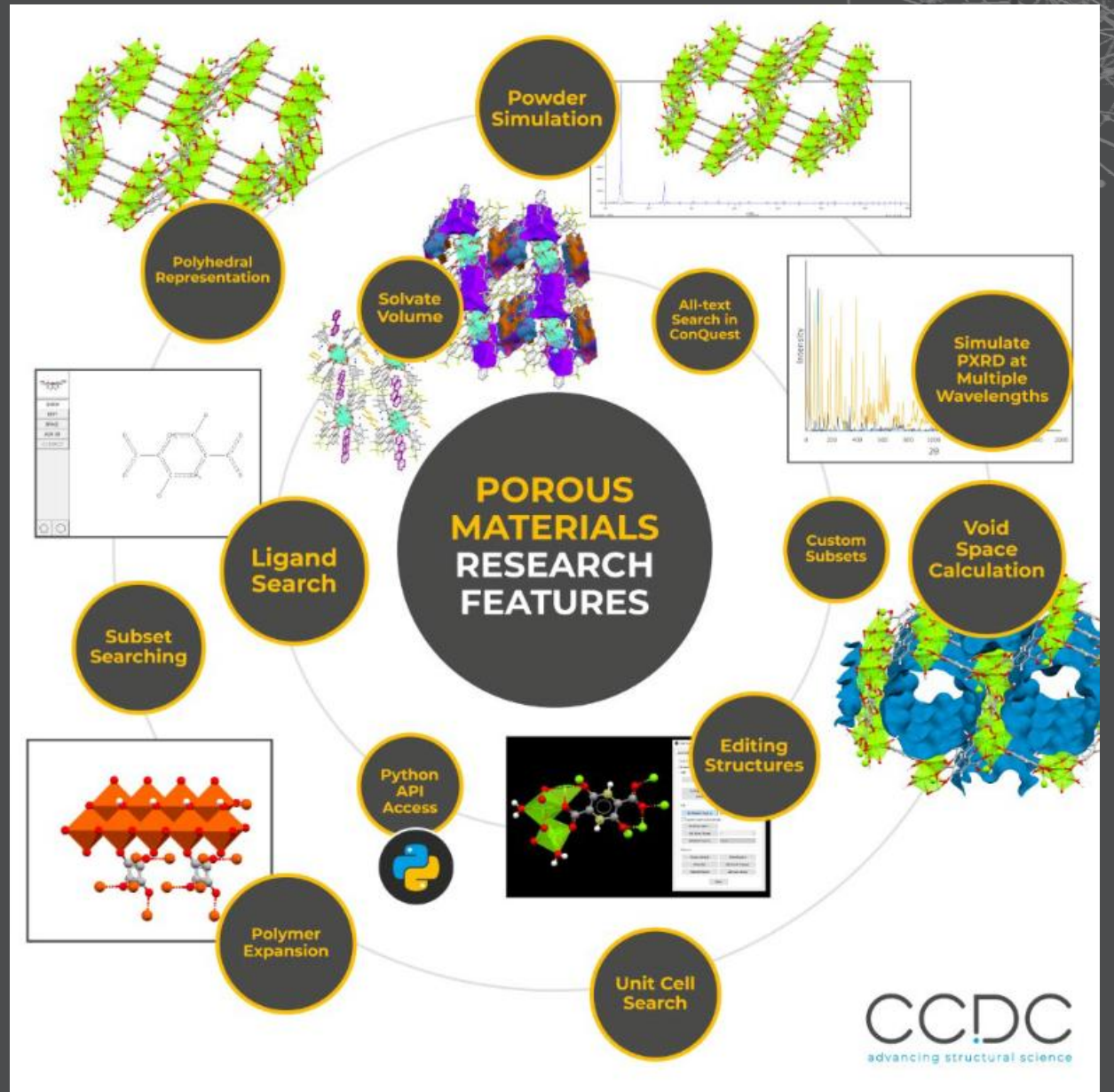
- Access structure data
- Calculate properties
- Integrate into workflows





CSD-Frameworks

- Newly released software package
- Provides features to specifically aid in MOF analysis workflow
- Open to consultation about specific needs



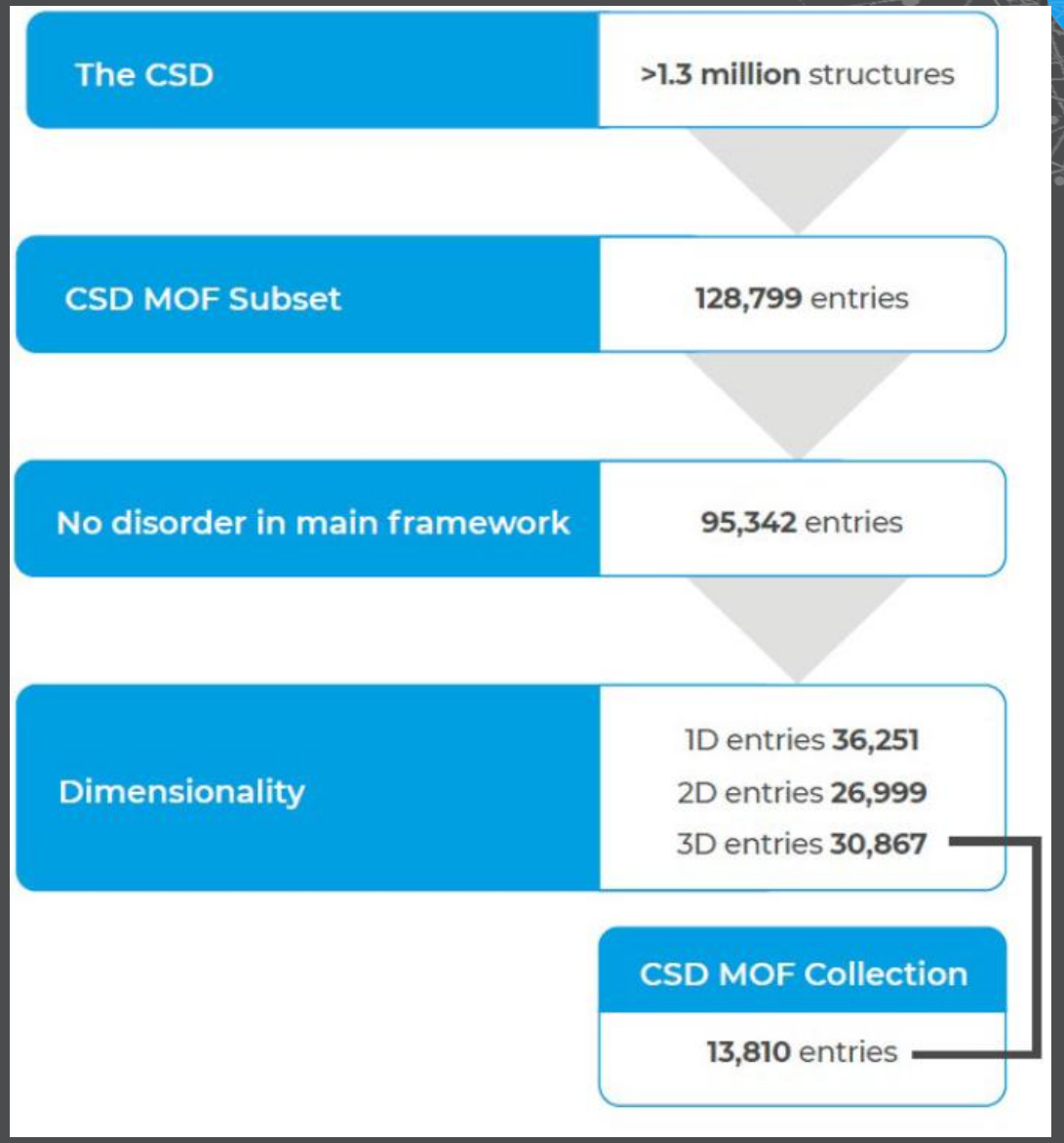
MOF Subset & MOF Collection

A highlighted part of the CSD

MOF-Subset: Limit searches to relevant structures

MOF-Collection: adapted for Computational Chemistry

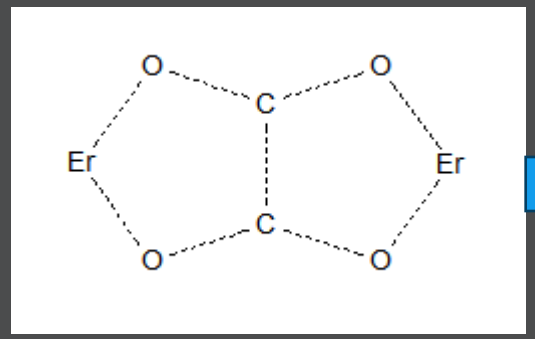
- *Convert to P1*
- *Remove guest molecules*
- *Add missing hydrogens*

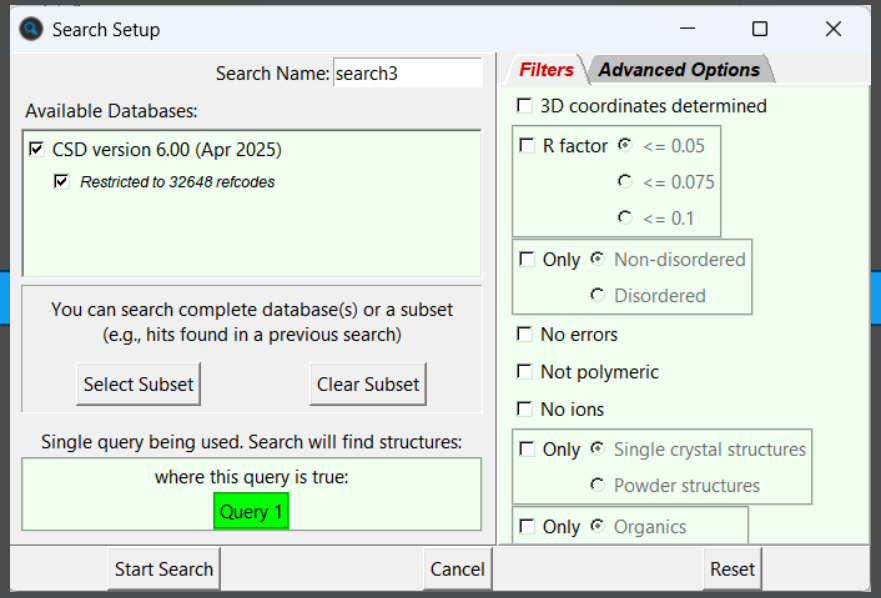


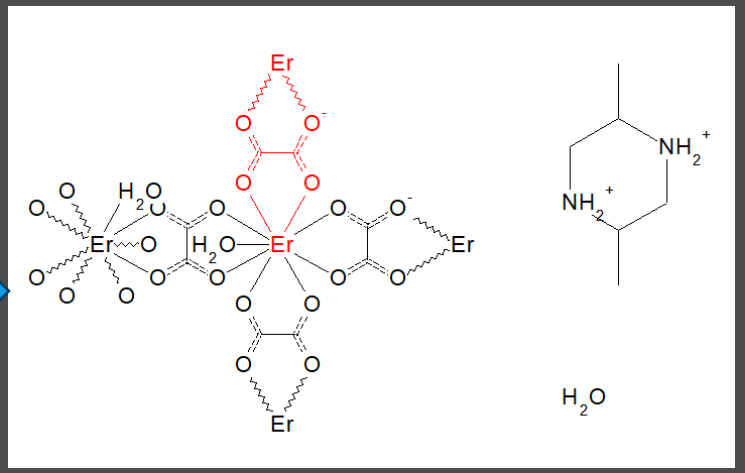
CCDC

How to Search: ConQuest

- Identify similar compounds from other researchers
- Gather structural property data from relevant materials to gain insights

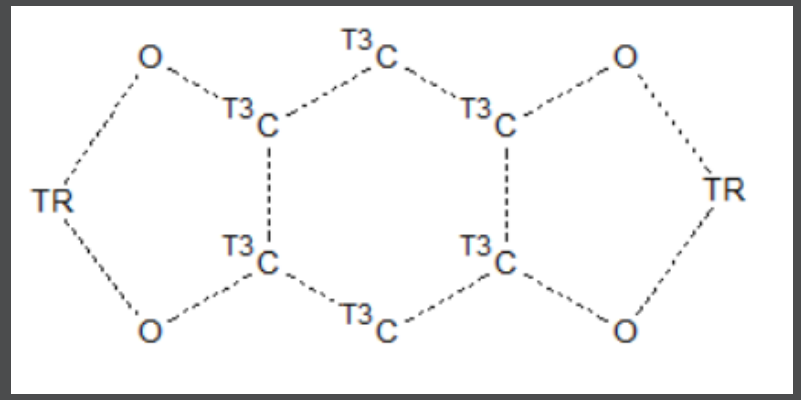




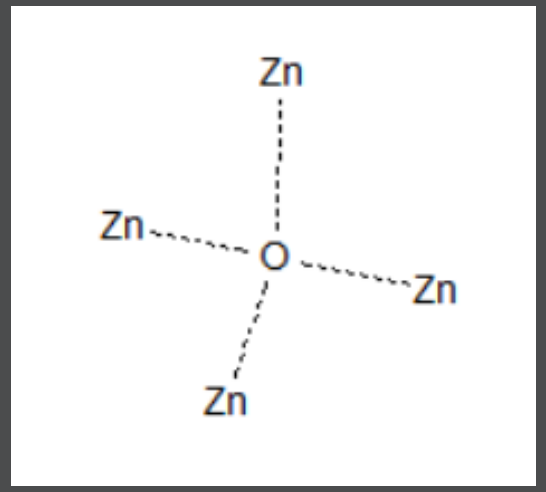


Search types:

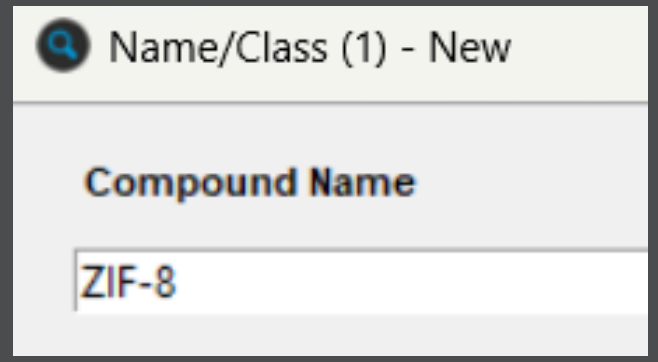
- By Ligand substructure:
- “DHBQ”
- 332 hits in MOF



- Metal Cluster
- Zn_4O (basic zinc acetate)
- 865 hits in MOF



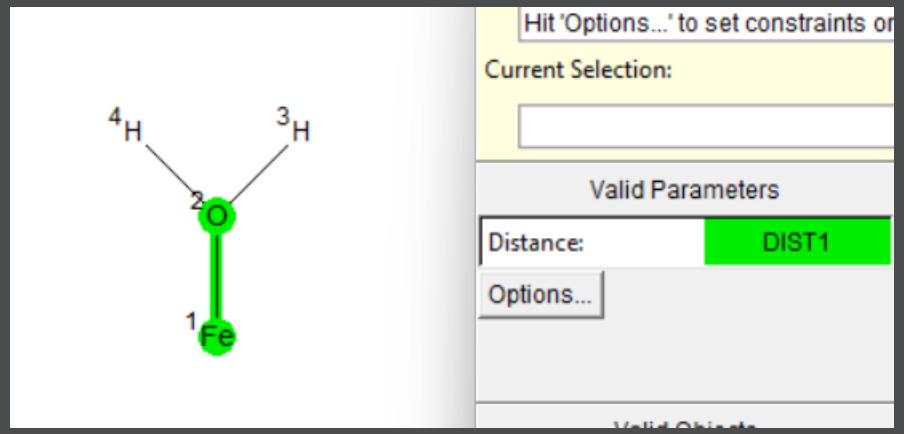
- By name
- “ZIF-8”
- 71 hits in MOF

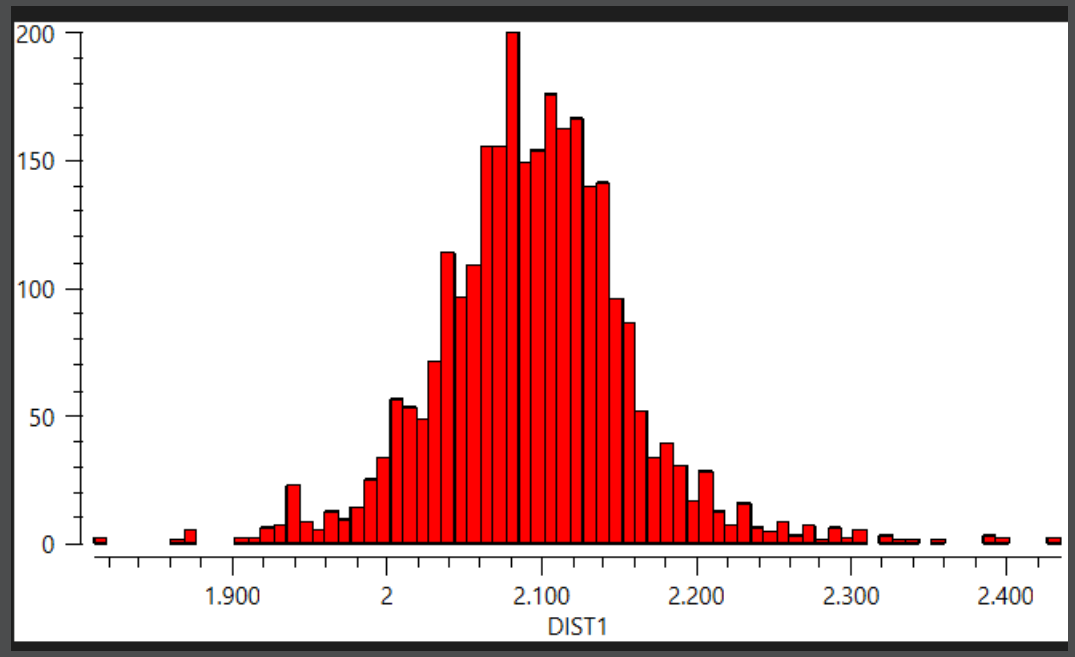


CCDC

Aggregating data:

- Fe – water distance:



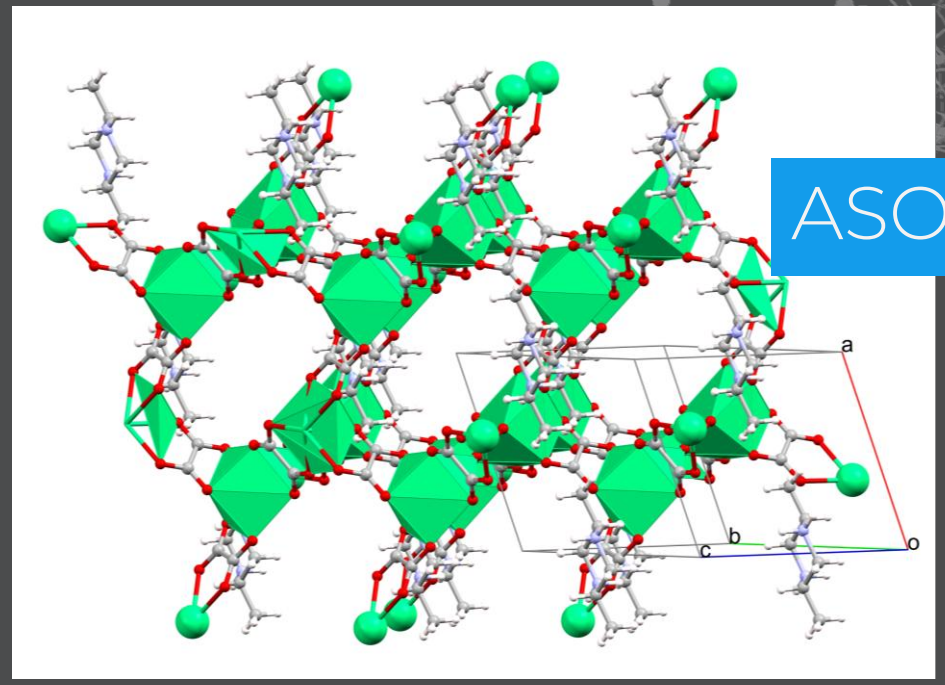


- Can be used with more complex queries

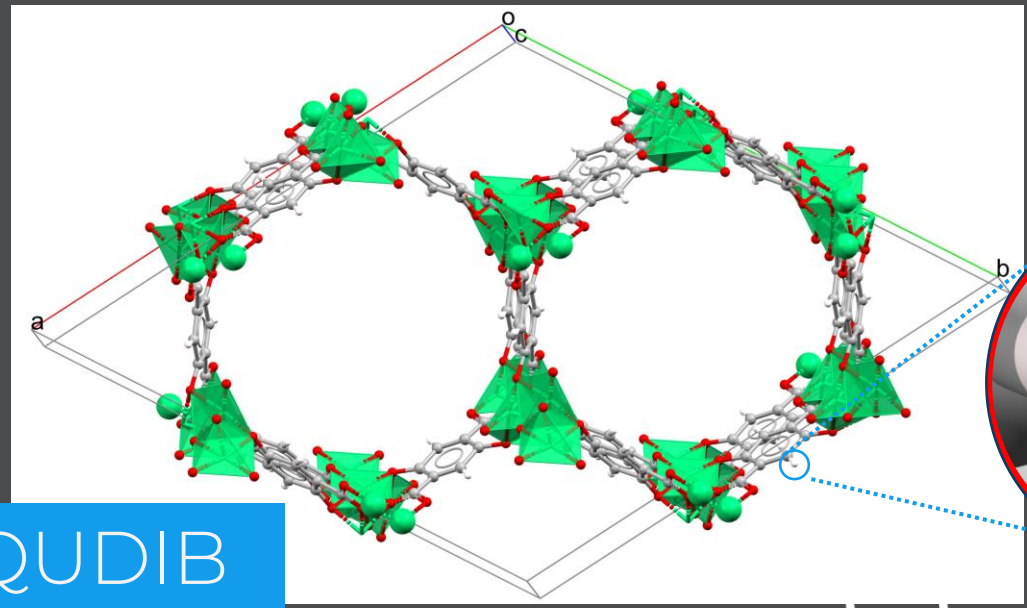
CCDC

Visualization

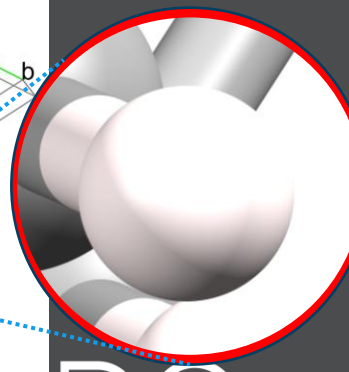
- Generate publication-ready figures and communicate your findings
- Visualize:
  - Crystal structure packing
  - Void space
  - Intermolecular interactions
  - Coordination environments
- High quality images with “File > POV-ray image”



ASONII

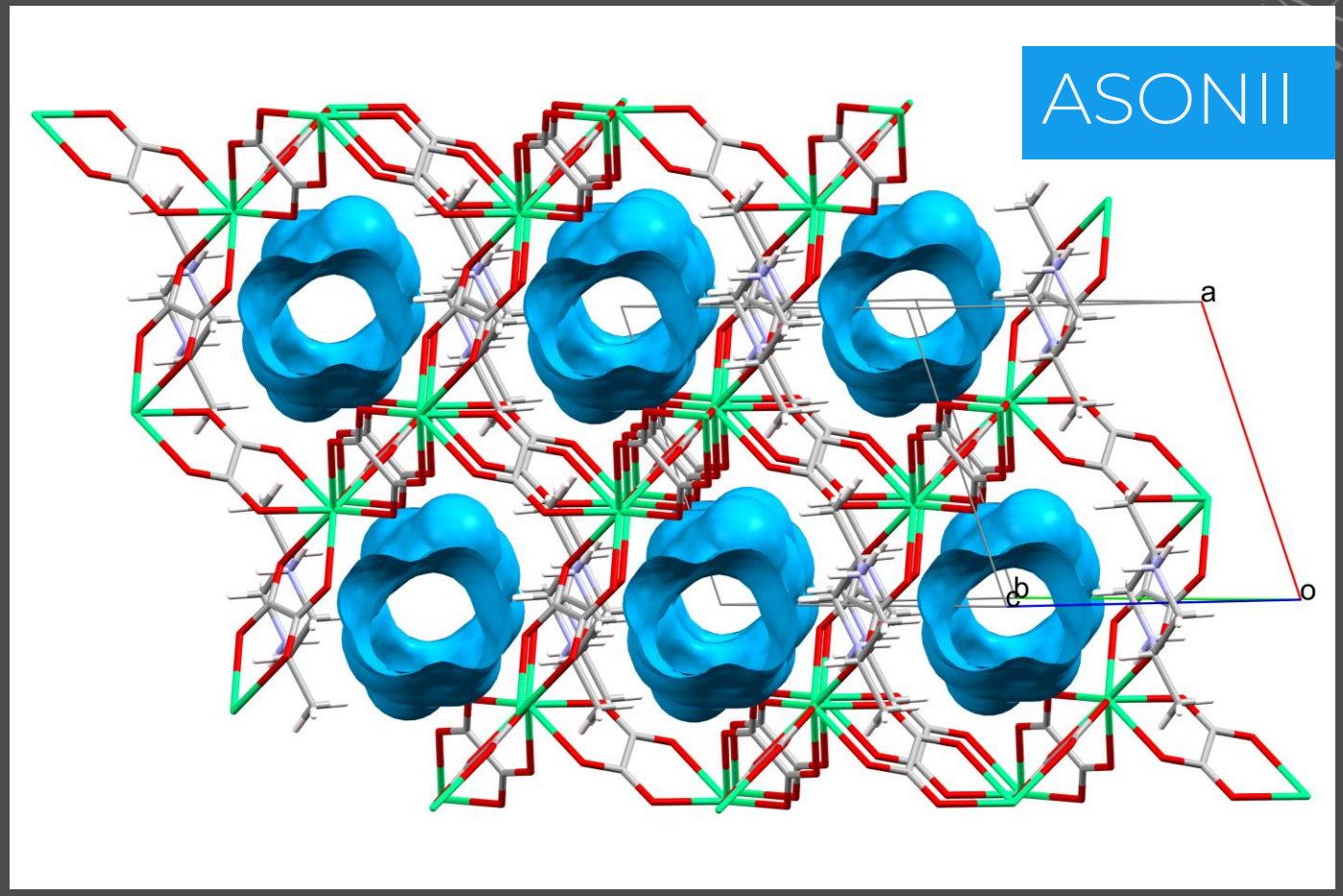


AQUDIB



Void and Pore Analysis

- Visualize voids and calculate their properties
- Visualise with “Display > Voids”
- Analyse with “Calculate > Pore Analyser”
  - total volume (Å^3)
  - pore diameter (Å)
  - percolated dimensions

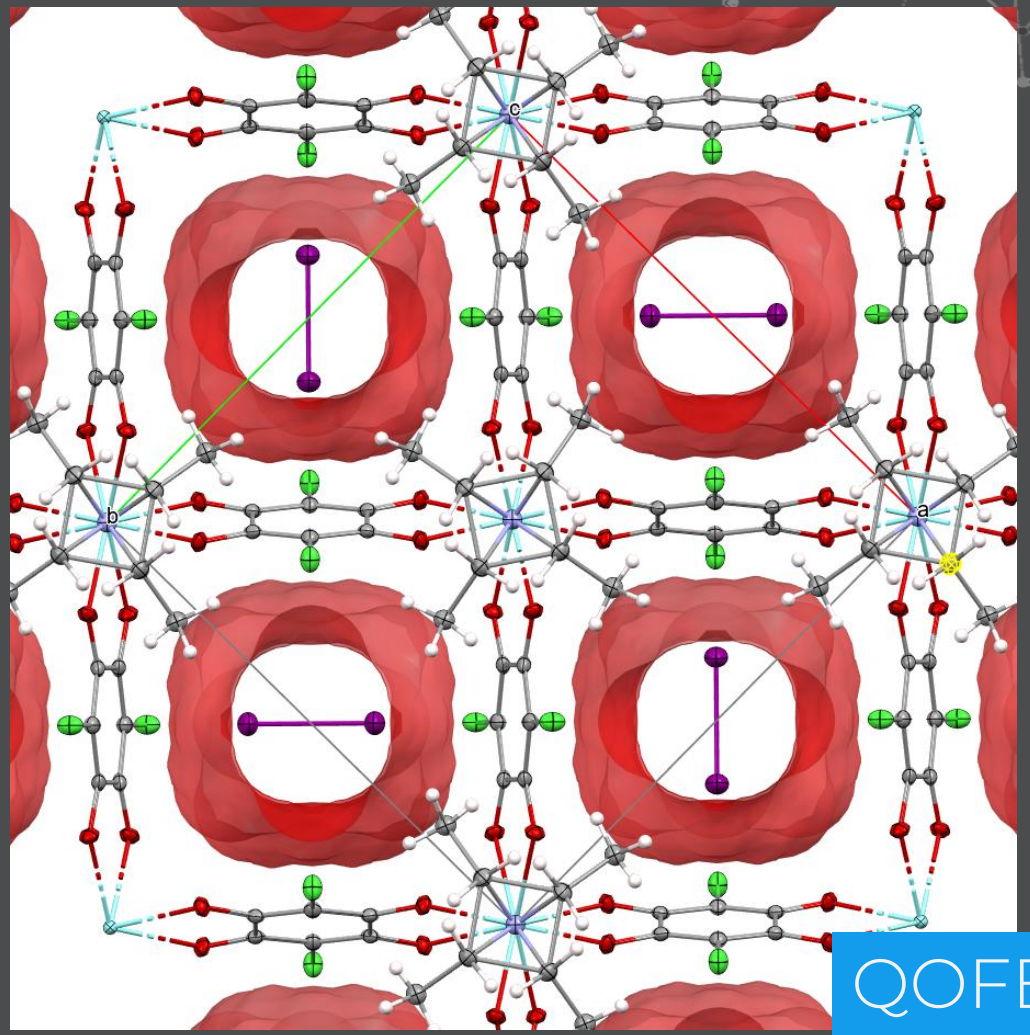


Using “Display > Voids”

CCDC

Guest Analysis

- With “CSD-Materials > Solvate Analyser”
- Visualize guest molecule networks (solvents, hydrates, gases, drug APIs, etc.)
- Calculate volume occupied by these guests
- Compare volume of voids with guest volume to determine how well-filled the network is



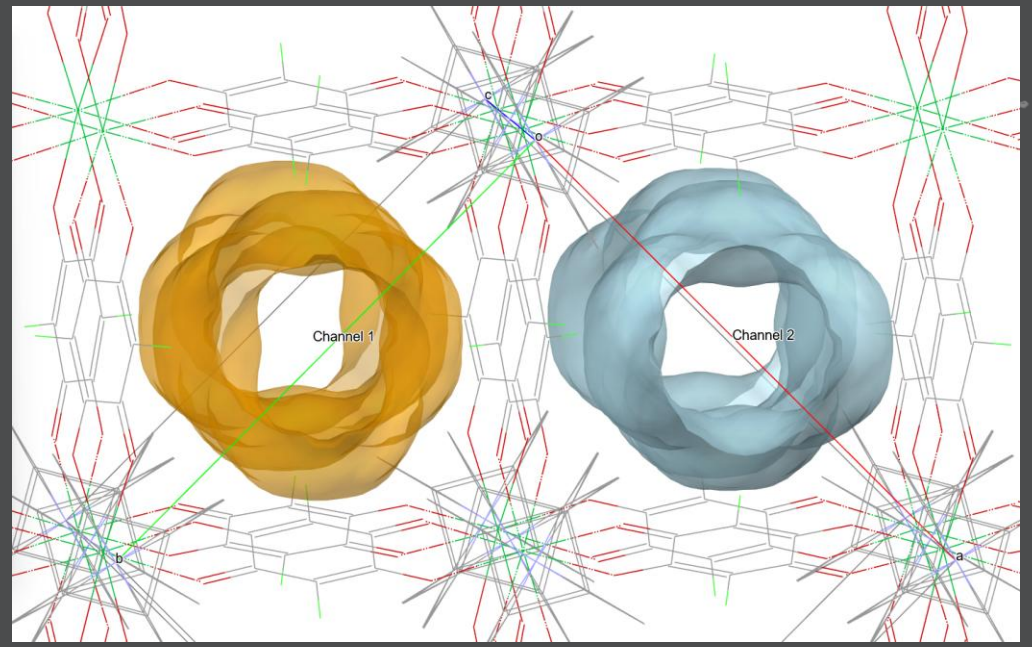
QOFBUL

I_2 in MOF – 86% efficiency

CCDC

Individual Void Analysis

- In “Display > Voids > individual voids”
- Different voids may have different properties



Details of individual voids and colour/style applied

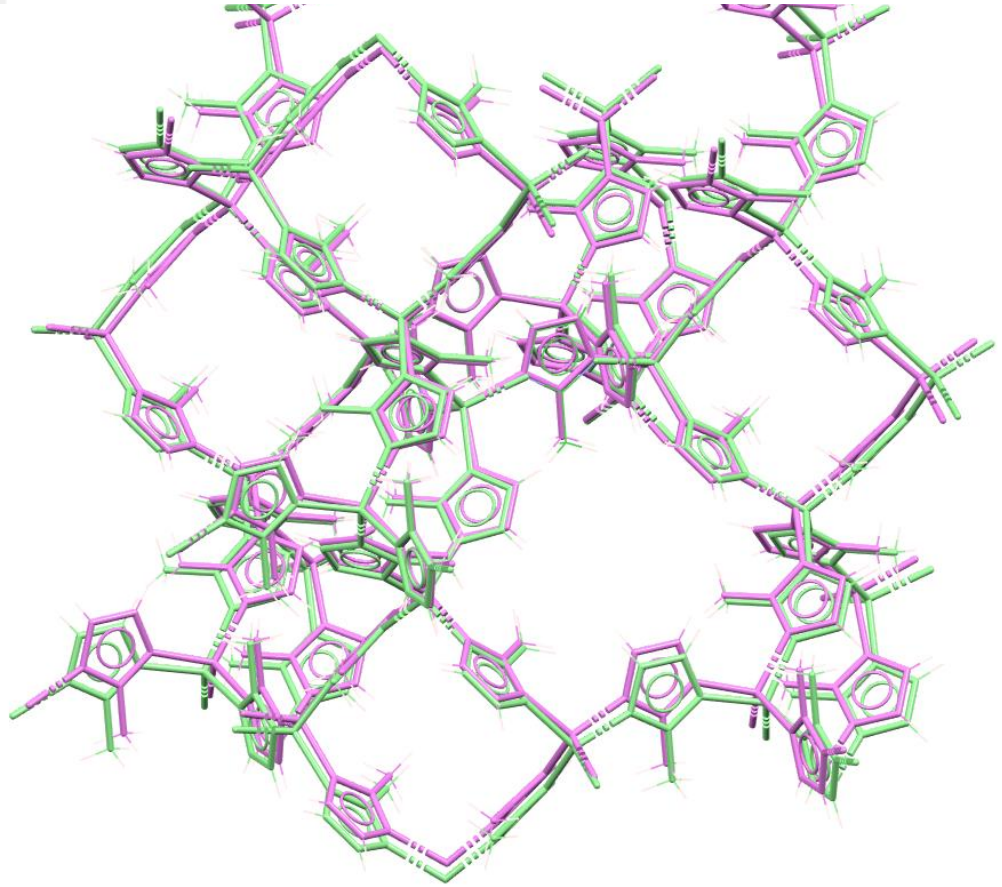
	Label	Volume (Å³)	Volume (% of unit cell)	Outside Colour	Visible	Symmetry ID	Donors nearby	Acceptors nearby
1	Channel 1	339.127	11.390	Orange	☑	1		Cl2, Cl3, Cl4, Cl5, ...
2	Channel 2	339.127	11.390	Light Blue	☑	1		Cl2, Cl3, Cl4, Cl5, ...

CCDC

Structure Comparison

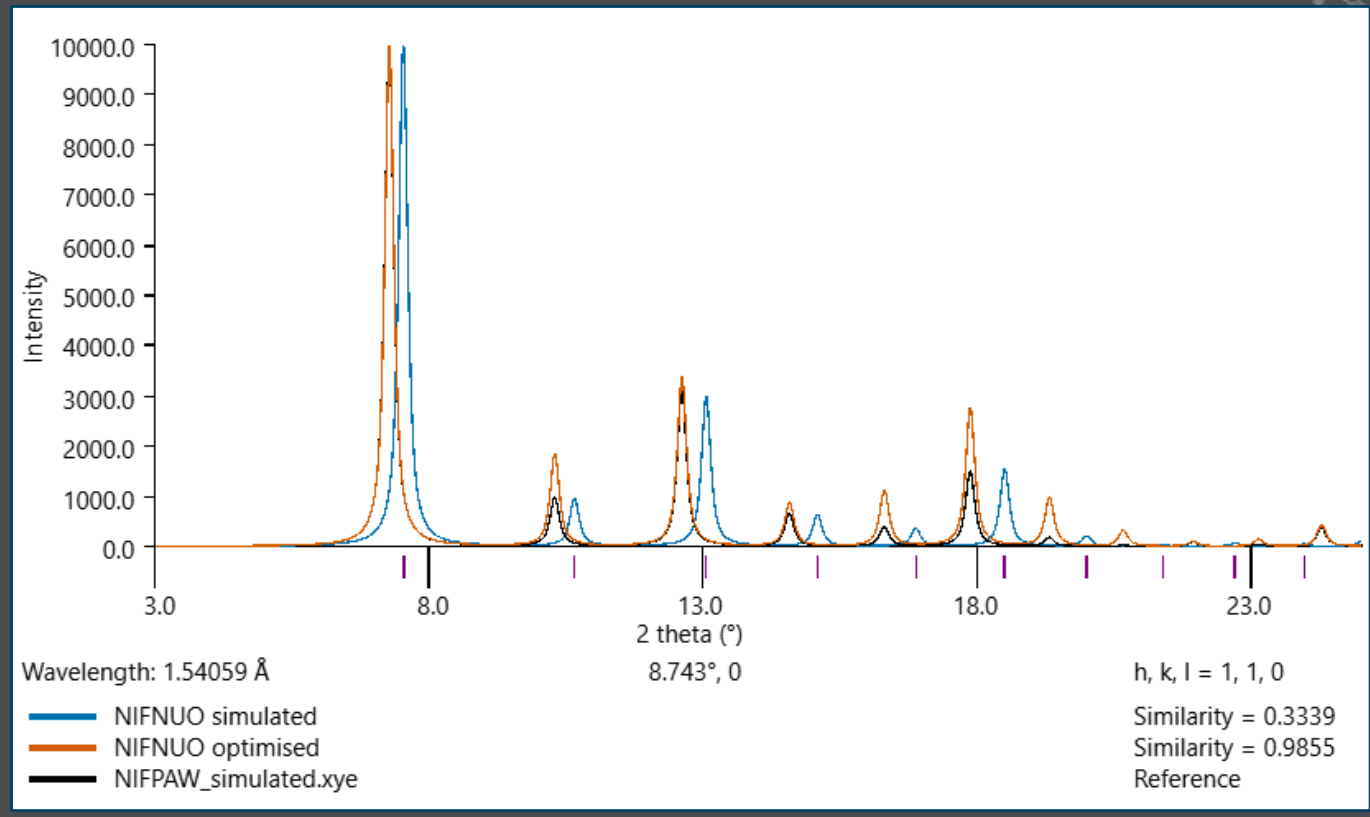
- With “CSD-Materials > Search > Crystal Packing Similarity”
- Frameworks with different:
  - Metals
  - Guest molecules
  - Solvents removed
- Right: ZIF-8 with CH_4 (pink) and Ar (green)

reference	comparison	molecules in com	RMS
NIFNUO	NIFPAW	26 out of 30	0.386



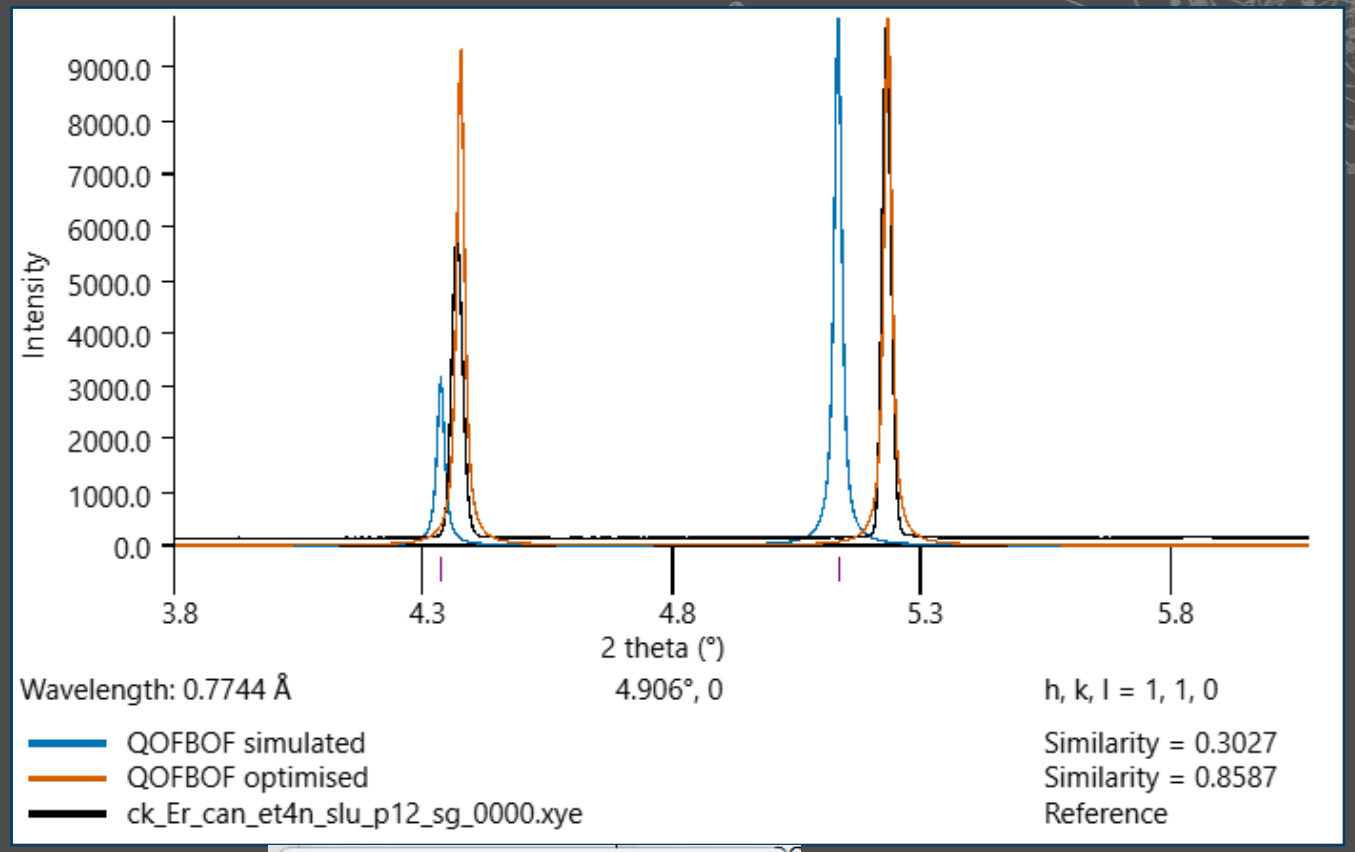
Simulating Powder Diffraction Patterns

- Compare and fit:
  - experimental data
  - calculated from single crystal
- Use “Powder > Customise > Match settings”
- Match increases from 0.334 to 0.985 (out of 1)



Matching Powder Diffraction

- (from my own PhD work)
- Experimental (Black, synchrotron PXRD, metal = Er)
- Simulated (single crystal, Blue, metal = Y)
- Optimised (85% match)
- (same match obtained with M = Er single crystal QOFNOR)

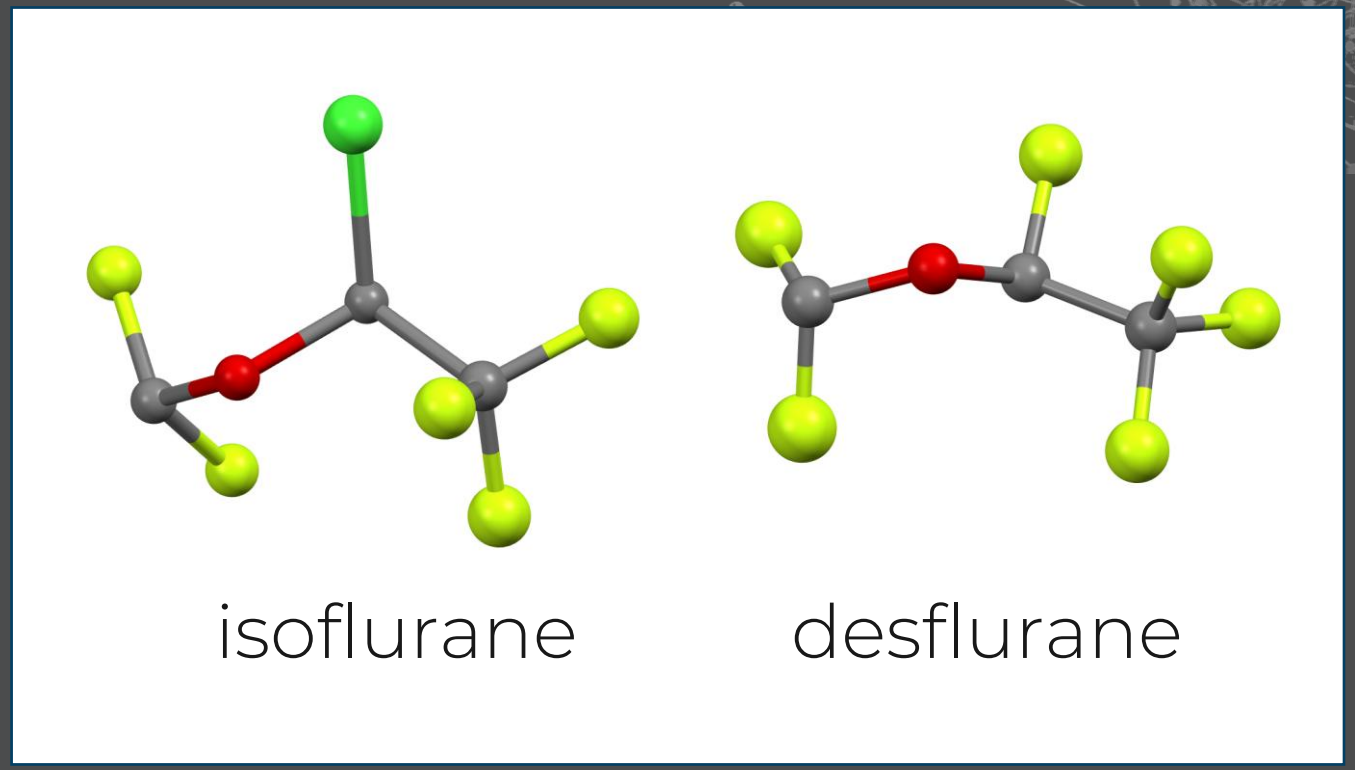


Powder Pattern Optimisation Res...

	QOFBOF	QOFBOF	Unit
a	12.1367	12.1269	Å
b	12.1367	12.1269	Å
c	20.2931	20.4187	Å

Case Study – MOF for volatiles uptake

- Goal: to find a MOF to capture anaesthetic drugs from breath in a hospital
- These are strong greenhouse gases and pollutants
- We need a MOF stable to water and the right size to filter these compounds



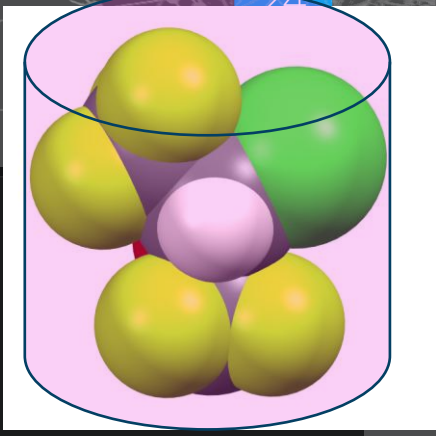
isoflurane desflurane

See: Abrahams et al; *Chem. Eur. J.* **2017**, *23*, 7871. https://doi.org/10.1002/chem.201700389

CCDC

Case Study – MOF for volatiles uptake

- 3D structure (the 3D MOF subset)
- Water stability e.g. [Zr] in ZIF or a “chelate”
- Pore limiting diameter from molecular data
  - See right – requires a PLD of 7.0 Å (isoflurane) or 6.2 (desflurane) Å



```python
import ccdc.io
import numpy as np
import scipy.optimize

reader = ccdc.io.CrystalReader()
mol = reader.molecule("TOBBIV")

# spherical to cartesian unit vector
def s2c(th1, ph1):
    return np.array((np.sin(ph1) * np.cos(th1), np.sin(ph1) * np.sin(th1), np.cos(ph1)))

def minimum_enclosing_cylinder(molecule, atom_lookup = "Alvarez"):
    positions = [np.array(atom.coordinates) for atom in molecule.atoms]
    radii = [atom.vdw_radius for atom in molecule.atoms]
    geometric_centre = np.mean(positions, axis = 0)
    p0 = [0.77, 0.77, 0, 0, 0] # theta, phi, x, y, z
    def fitfunc(pars):
        vector = s2c(pars[0], pars[1])
        places = [pos - geometric_centre - pars[2:] for pos in positions]
        offsets = [np.linalg.norm(place - (vector * (place @ vector))) +  radius
                   for place, radius in zip(places, radii)]
        return max(offsets)
    return scipy.optimize.basinhopping(fitfunc, p0, niter = 1000)

minimum_enclosing_cylinder(mol)
```

✓ 35.5s

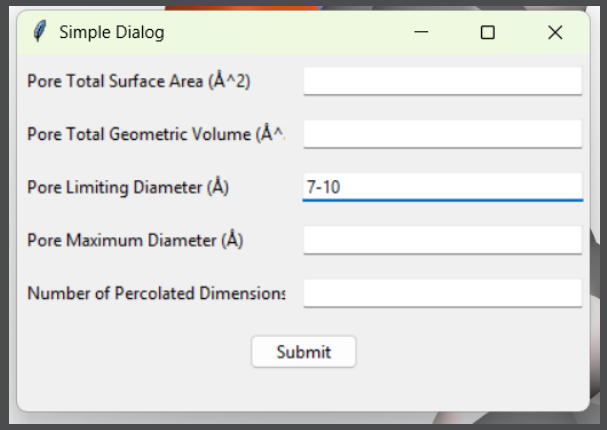
```
                    message: ['requested number of basinhopping iterations completed successfully']
                    success: False
                        fun: 3.511702722968617
                          x: [ 7.148e-01  2.100e+00 -6.079e-02 -1.040e-01
                               -7.314e-02]
```

CCDC

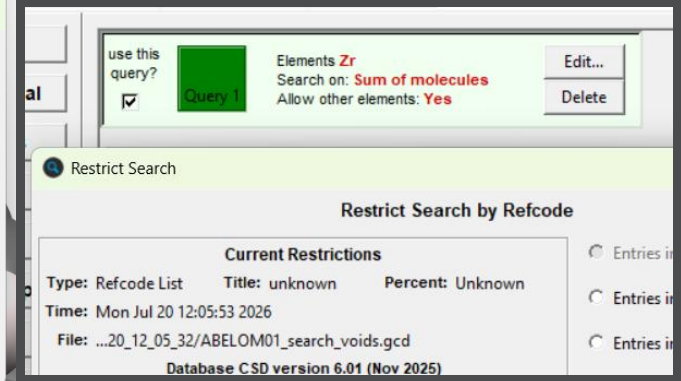
Case Study – MOF for volatiles uptake

XUPVIR – Huang et al.
Inorg. Chem. Front. (2025) 12 (9): 3531–3544.
"Encapsulation of toxic liquid molecules…"

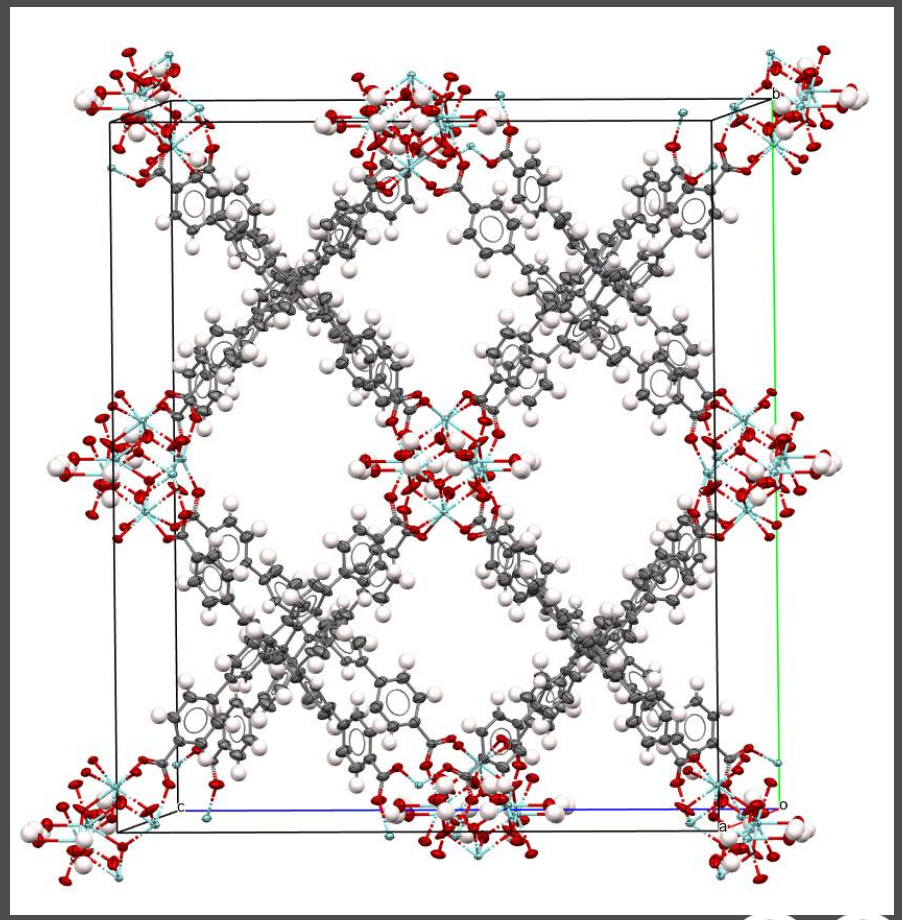
- Combine these lists and search in ConQuest
- (37 hits)



- Search Prototypes menu



- ConQuest filtering

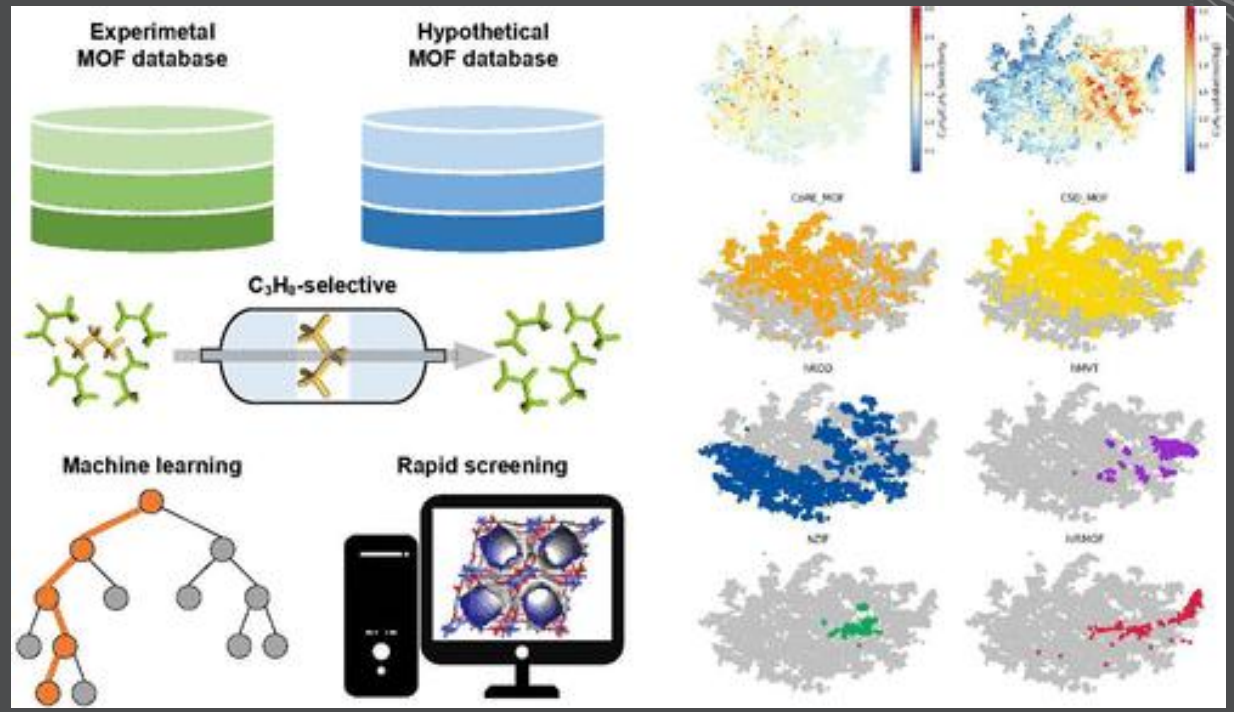


CCDC

How is CSD data used in academia?

Finding good materials for an industrially relevant separation

- From Tang et al.
- Using CSD MOFs to predict binding of C_3H_6 and C_3H_8 (with DFT)
- Use this information to design better MOFs



Tang et al. *ACS Appl. Mater. Interfaces* 2021, 13, 45, 53454–53467 https://doi.org/10.1021/acsami.1c13786

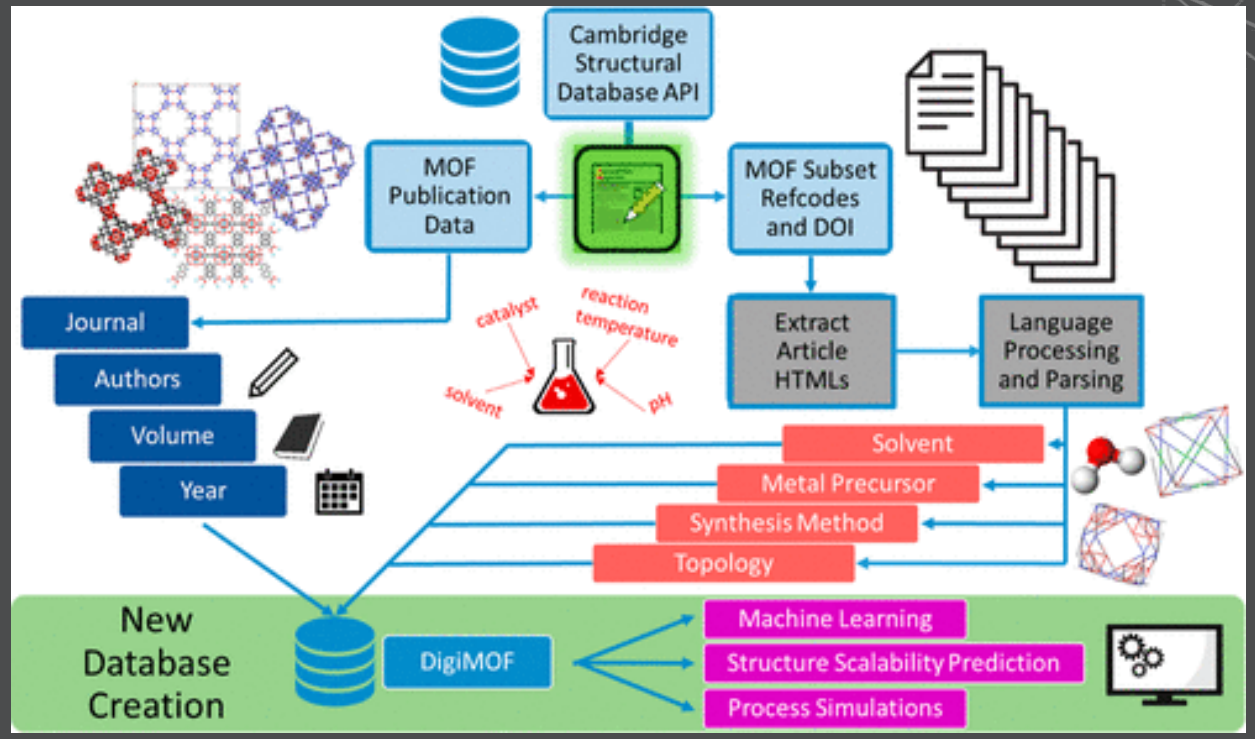
CCDC

How is CSD data used in academia?

Finding synthesis conditions for a MOF from structure

Glasby et al.

- MOF data was used to find structure components
- Chemistry Literature used to find how these are made
- DigiMOF can predict how to make a new MOF



Glasby et al. *Chem. Mater.* 2023, 35, 11, 4510–4524, https://doi.org/10.1021/acs.chemmater.3c00788

CCDC

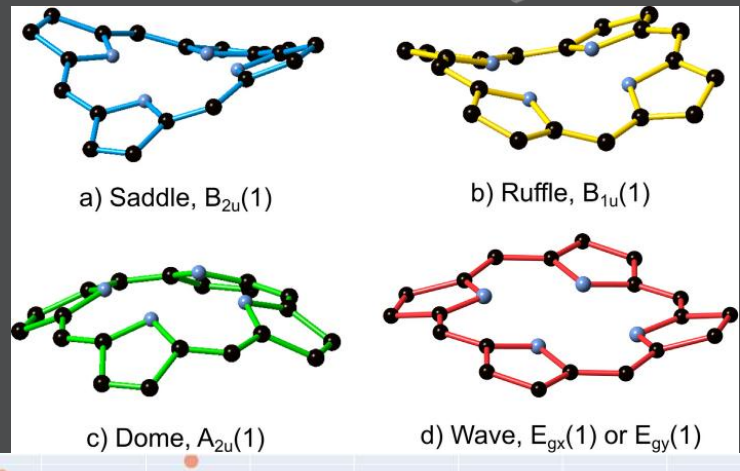
How is CSD data used in academia?

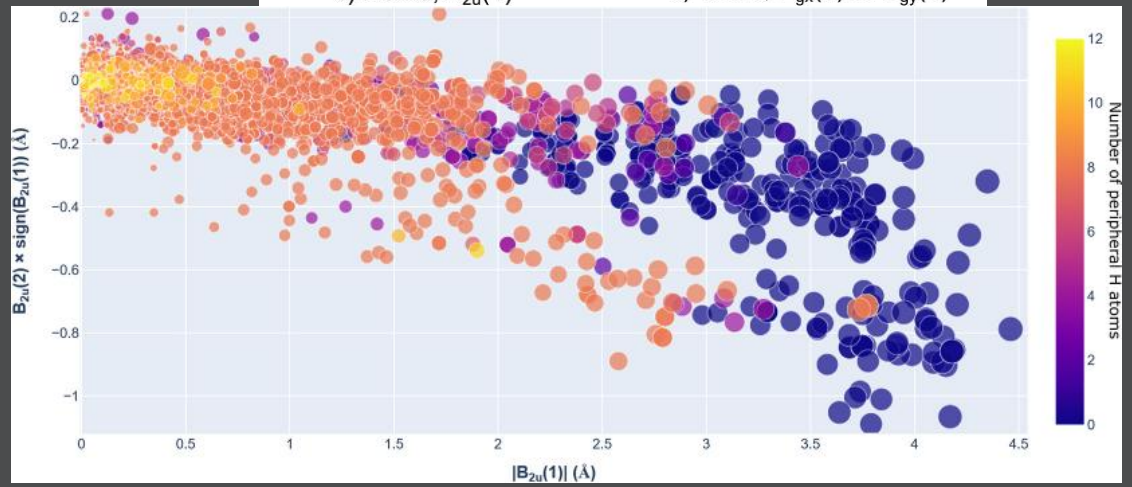
Conformation of macrocyclic ligands

Kingsbury and Senge

Crystal structure data was aggregated to find deformation pathways

- Structure activity relationships
- Used to make “chiral porphyrins”





Kingsbury and Senge, Coordination Chemistry Reviews, 2021 213760 https://doi.org/10.1016/j.ccr.2020.213760

CCDC

Future directions

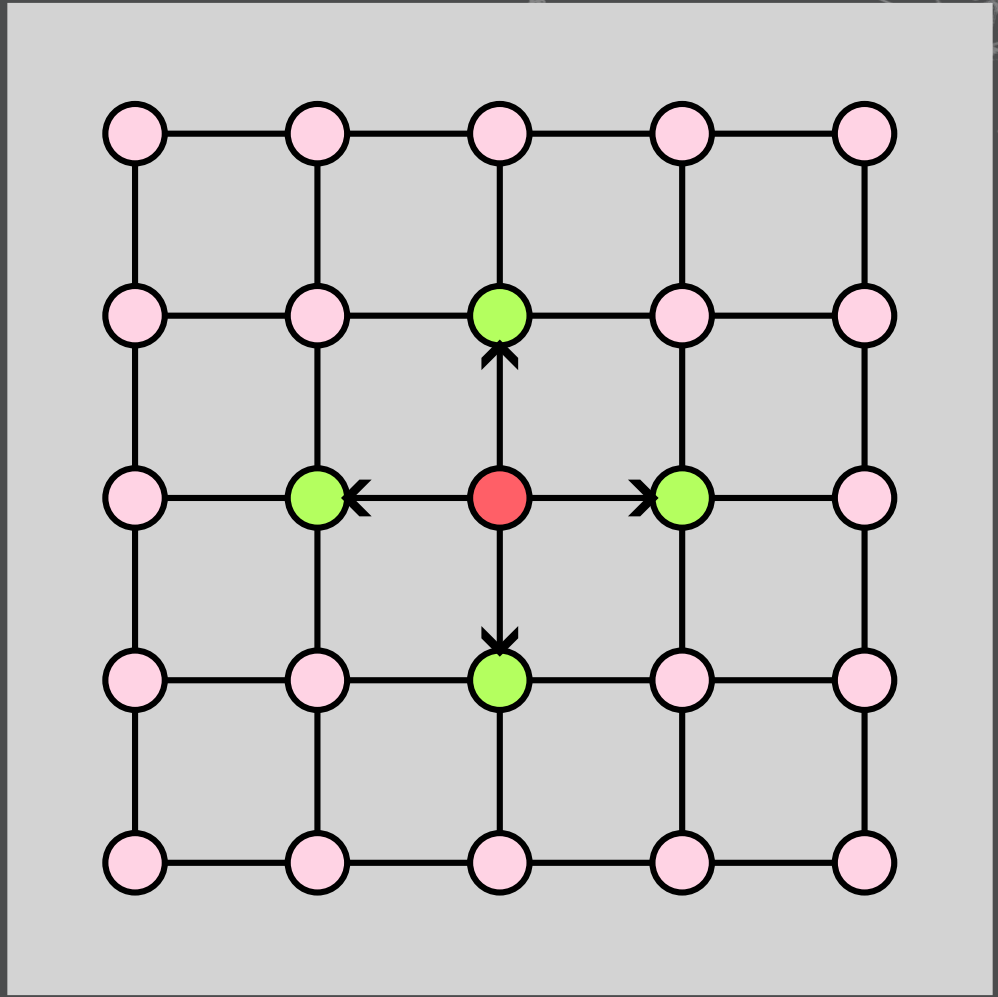
- Topology (searchable in 2027.1)
- Void search (in 2026.2)
- Internal surfaces (TBD)
- MOFid (TBD)

Please contact if you'd like access to tools in development

CCDC

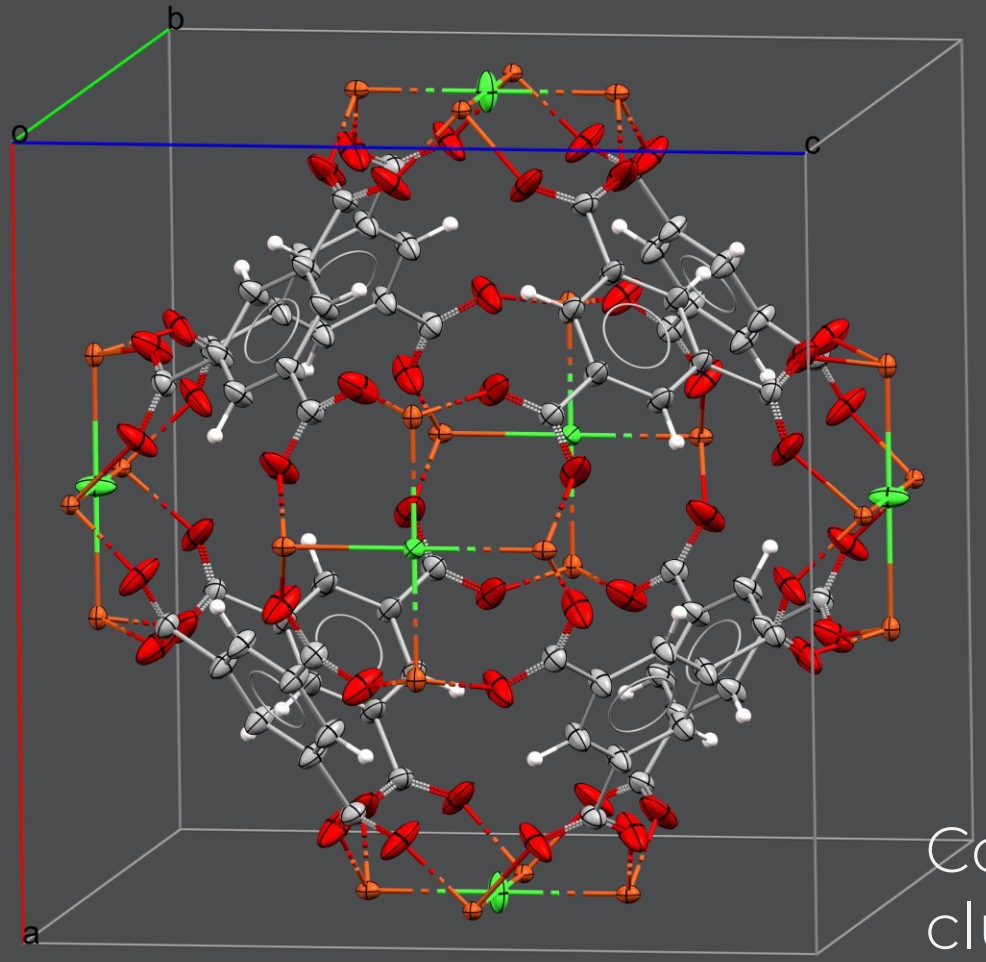
Topology in MOFs

- An abstract representation of structure – each circle represents an atom or a few atoms
- “Connection” is more important than connectivity
- Can be a code **“sql”** or a string **“$4(4^4 \cdot 8_2^2)$”**



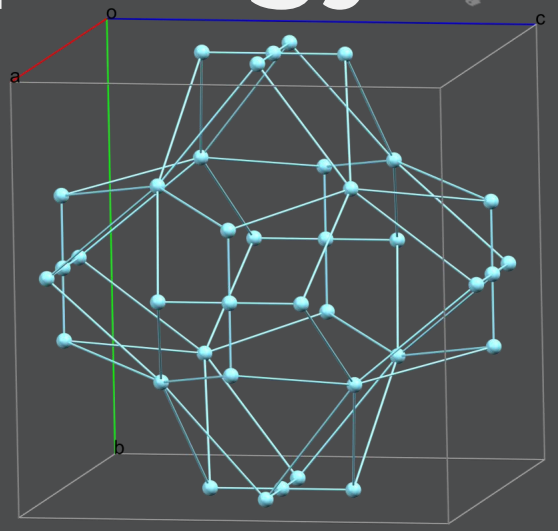
CCDC

MOF prototypes - Topology



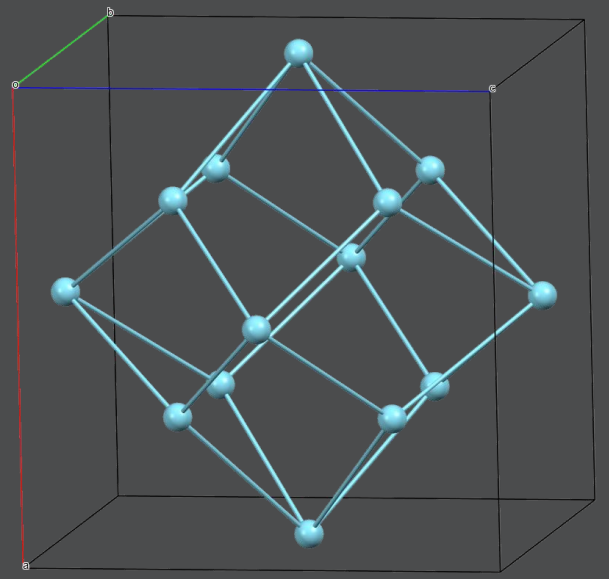
ABEMIF Crystal structure from CSD

Metals as nodes



Using “M”
nets/lqtxbmlv
CCDC ident.
not in RCSR

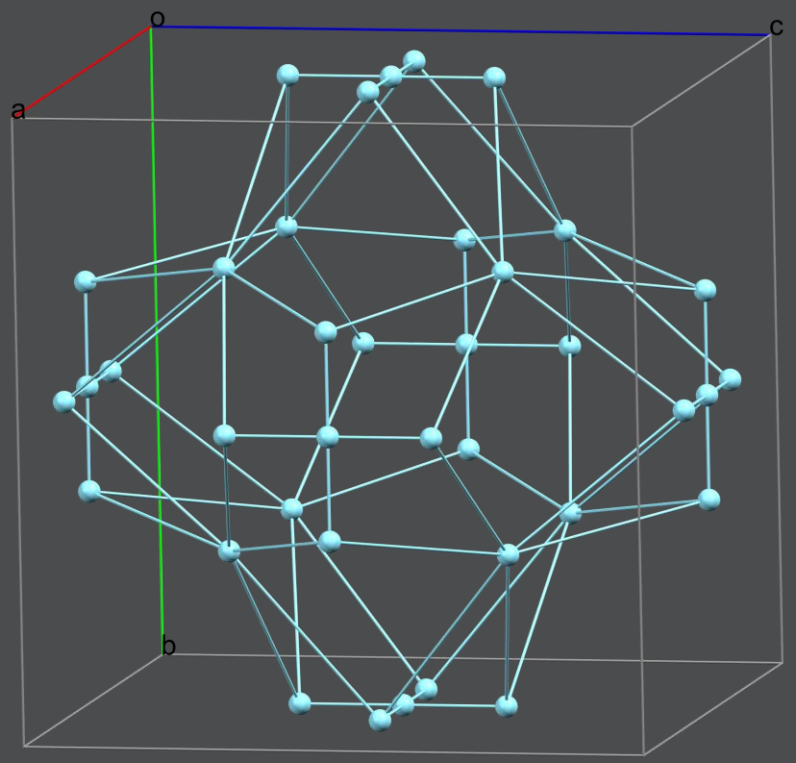
Combine clusters



Using “M_4Cl”
nets/the
(RCSR ident.)
nets/asptuxtf

CCDC

MOF prototypes - Topology



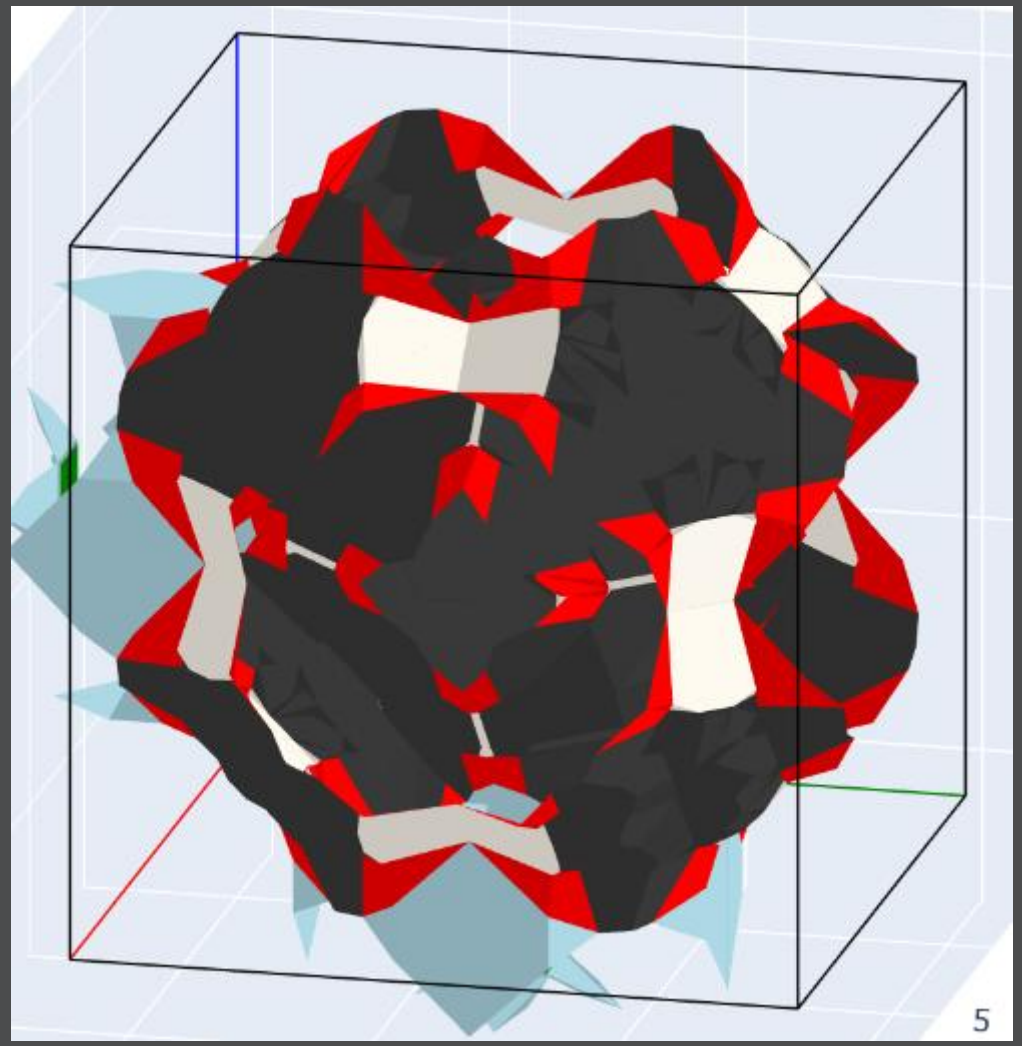
Topology representation: “nets/lqtxbmlv”

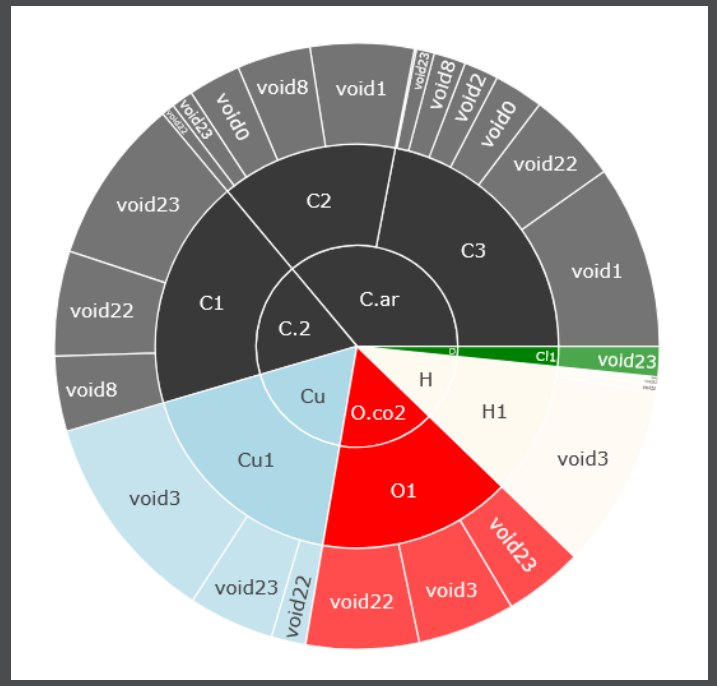
- Find IRMOFs – 8 unique structures with different metals or ligands
- Confirm Dimensionality (3D)

Metal-Organic Framework

Vertex symbol	$4(4_2^4.6_4^2)^3 + 5(4^8.8_{26}^2)^{12} + 6(4^3.4_2^3.6^6.8_{12}^3)^8$
Assignment	not found
Dimensionality	3D
cs10	2262
Assignment (w.cs10)	None
Colouring	2
Is Planar	False
FLTO Dimensionality	0
Interpenetration	1
Unique Identifier	nets/lqtxbmlv
CSD Summary	0 matches out of 1157 values

MOF prototypes - Surfaces

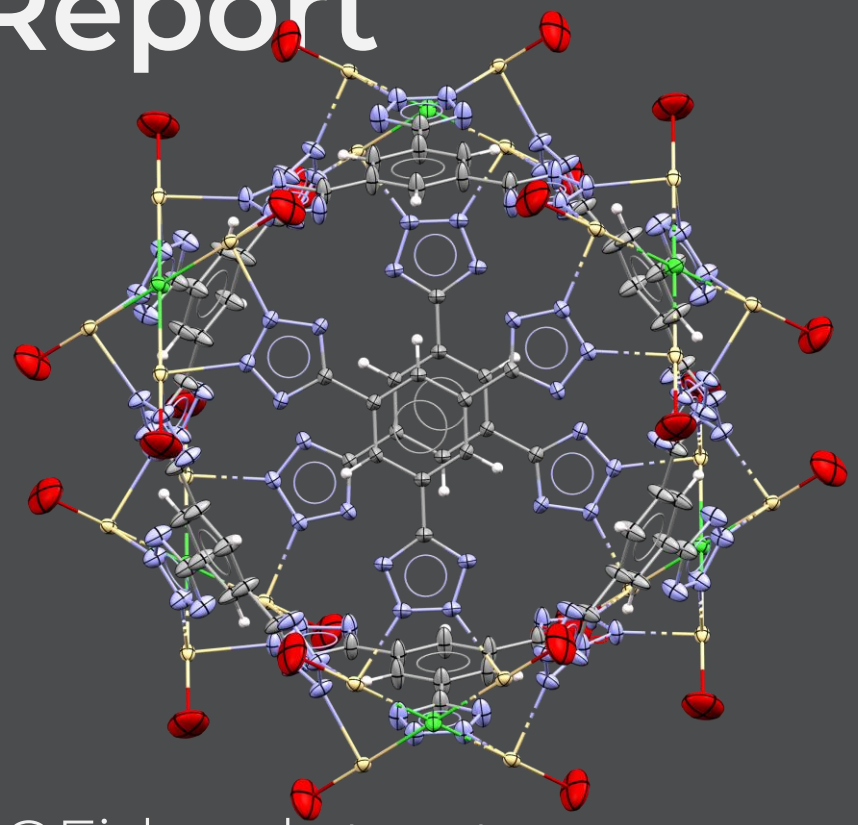




- Find the composition of the internal surface
- Correlate adsorption properties with surface area and atom types

CCDC

MOF prototypes - Report



- MOFid and structure composition
- Similarity search
- Void analysis
- Searchable Database

MOF report: BETHIT

CCDC advancing structural science

Topology

Metal-Organic Framework (Combine Clusters)

Vertex symbol	$12(4^8.6_2^4.8_4^8.8_6^4.10_{16}^4.\infty^{36})^3 + 3(4^3)^8$ + terminal^{12}
Assignment	not found
Dimensionality	3D
cs10	2197
Assignment (w.cs10)	None
Colouring	2
Is Planar	False
FLTO Dimensionality	None
Unique Identifier	nets/bddcqzmw
CSD Summary	0 matches out of 1157

Metal-Organic Framework (Combine Clusters) Frame

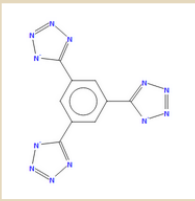
Vertex symbol	$3(4^3)^8 + 8(4^8.6_2^4.8_4^8.8_6^4.10_{16}^4)^3$
Assignment	nets/the
Dimensionality	3D
cs10	1306
Assignment (w.cs10)	nets/the
Colouring	2
Is Planar	False
FLTO Dimensionality	0
Unique Identifier	nets/asptuxtf

Metal-Organic Frame Clusters) Simple

Vertex symbol	$3(4^3)^8 + 8(4^8.6_2^4.8_4^8.8_$
Assignment	nets/the
Dimensionality	3D
cs10	1306
Assignment (w.cs10)	nets/the
Colouring	2
Is Planar	False
FLTO Dimensionality	0
Unique Identifier	nets/asptuxtf
CSD Summary	0 matches out of 1157

Ligands

CID 176639841



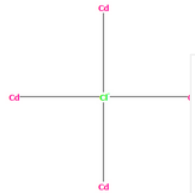
smiles: c1c(cc(cc1C1=NN=N[N-]1)C1=NN=N[N-]1)C1=NN=N[N-]1
connectivity: 6
denticity: 6

Water

smiles: O
connectivity: 1
denticity: 1
charge: 0
csd_matches: 255 / 633
csd_refcodes: ALIFUY, AQIVIF, AVUFED01
rotatable_bonds: 0
adjacent_rotatable_bonds: 0
mean_compressibility_percent: n/a

Metals / Clusters

Cd4 Cl1 1-



display_formula: Cd_4Cl
smiles: [Cd][Cl-]([Cd])([Cd])[Cd]
atoms: ['Cd19', 'Cd20', 'Cd14', 'Cd13',
connectivity: 20
csd_matches: 0 / 633
csd_refcodes:

MOFid (CCDC version)

MOFID: Cd4Cl.O.[Cl-].[N-2]
[N-2].c1c[cc(cc1C1=NN=N[N-]1)C1=NN=N[N-]1)C1=NN=N[N-]1
MOFidCSD.the.cat0.BETHIT
MOFkey:
Cd4Cl.LODGQTPDPQHRSS.UXMOGIXOTMNDOH.VEXZGXHMUGYJMC.XLYOFNC

Similarity Results

	ABEMIF	AWEKAR	APUZIV01	APUZIV	BASJUB
refcode	ABEMIF	AWEKAR	APUZIV01	APUZIV	BASJUB
common_features	2	2	2	2	1
scaled_by_freq	1.213201	0.275823	0.275823	0.275823	0.213201
same	{(topology, nets/asptuxtf), (ligand, [Cl-])}	{(ligand, O), (ligand, [Cl-])}	{(ligand, O), (ligand, [Cl-])}	{(ligand, O), (ligand, [Cl-])}	{(ligand, [Cl-])}

Voids (using molecule.void_volume / PoreAnalyser)

(minus disconnected molecules and volatile ligands)

Void Volume: 67.3 %
Volume in unit cell: 4888.0 ($Å^3$)
Removed Volatile Ligands: ['Water', 'Water', 'Water', 'Water', 'Water', 'Water', 'Water', 'Water', 'Water', 'Water', 'Water']
Removed Neutral Spectators: []

Network Accessible Volume: 5428.8 ($Å^3$)
Network Accessible Surface Area: 912.22 ($Å^2$)
Network Accessible Surface Area per mass: 1489.2 (m^2g^{-1})
Total Geometric Volume: 5429.18 ($Å^3$)
Total Geometric Area: 1186.83 ($Å^2$)
Max Pore Diameter: 16.16 (Å)
Pore Limiting Diameter: 10.36 (Å)
Number of Percolated Dimensions: 3

Calculated by the CSD MOF Report Tool. (Readme)
Dr. Chris Kingsbury (ckingsbury@ccdc.cam.ac.uk) (2025)

Demonstration

Dr Christopher J. Kingsbury, Ph.D.

Functional Materials Scientist

- How to:
- Load a MOF structure
- Generate a high-quality image
- Compare two PXRD Patterns

CCDC