The Cambridge Structural Database

The Cambridge Crystallographic Data Centre (CCDC)

Email: hello@ccdc.cam.ac.uk Website: www.ccdc.cam.ac.uk X: ccdc_cambridge Facebook: ccdc.cambridge YouTube: CCDCCambridge

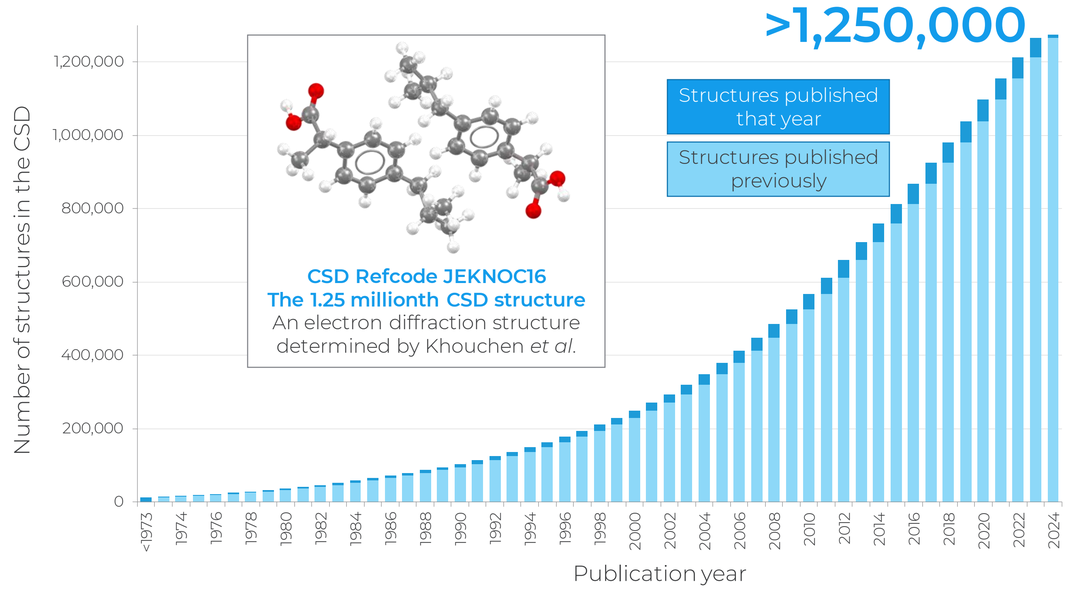
The CSD

The Cambridge Structural Database (CSD) is a curated database of over 1.25 million small molecule organic and metal-organic experimental crystal structures. The data is shared by scientists worldwide and every single entry is enhanced and validated by the CCDC to aid the discoverability and re-use of the data and to help others gain new insights from this valuable resource.

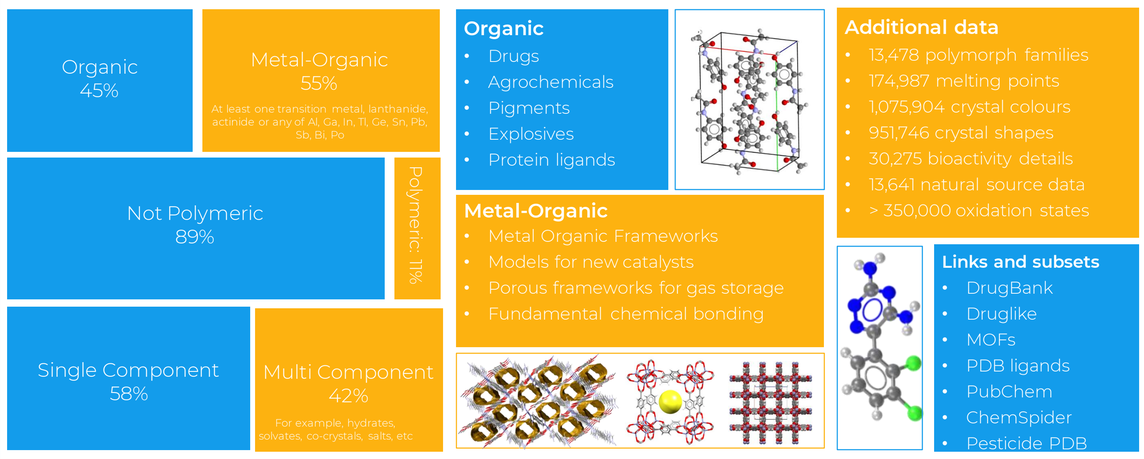
The use of the CSD has grown significantly since it was established in 1965 and today it is used globally to advance scientific research into pharmaceuticals, agrochemicals, functional materials, MOFs, and more in both industrial and academic research. It is also relied on by universities worldwide to help teach chemistry, crystallography, data science and more.

CSD in Numbers

- **> 1.25 million** structures
  - >94M 3D coordinates
- > **28 million** bond lengths
  - >2M unique distributions
- **> 40 million** valence angles
  - >3M unique distributions
- **> 14 million** torsion angles
  - >800K unique distributions
- **> 2 million** rings
  - >400K unique distributions
- **>60K** new structures a year
  - One added every **2 minutes**
- >195K existing CSD entries improved in 2023



The Structures

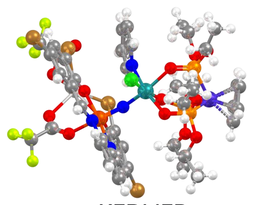


Organic
- Drugs
- Agrochemicals
- Pigments
- Explosives
- Protein ligands

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

Additional data
- 13,478 polymorph families
- 174,987 melting points
- 1,075,904 crystal colours
- 951,746 crystal shapes
- 30,275 bioactivity details
- 13,641 natural source data
- > 350,000 oxidation states

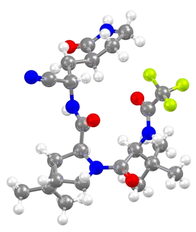
Links and subsets
- DrugBank
- Druglike
- MOFs
- PDB ligands
- PubChem
- ChemSpider
- Pesticide PDB



KEDMEP

C_{46} H_{50} Ag Br_4 Cl Co F_6 Fe N_4 O_{15} P_3 Ru, 2(C H_2 Cl_2)

Most different elements in a structure (12)

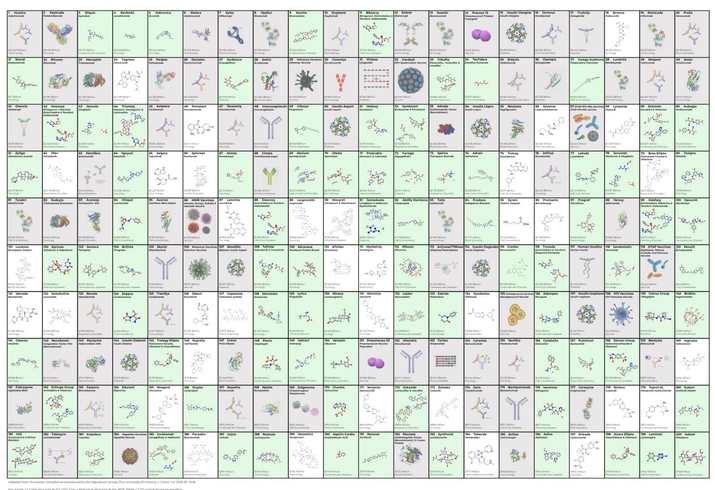


ZIVMEA

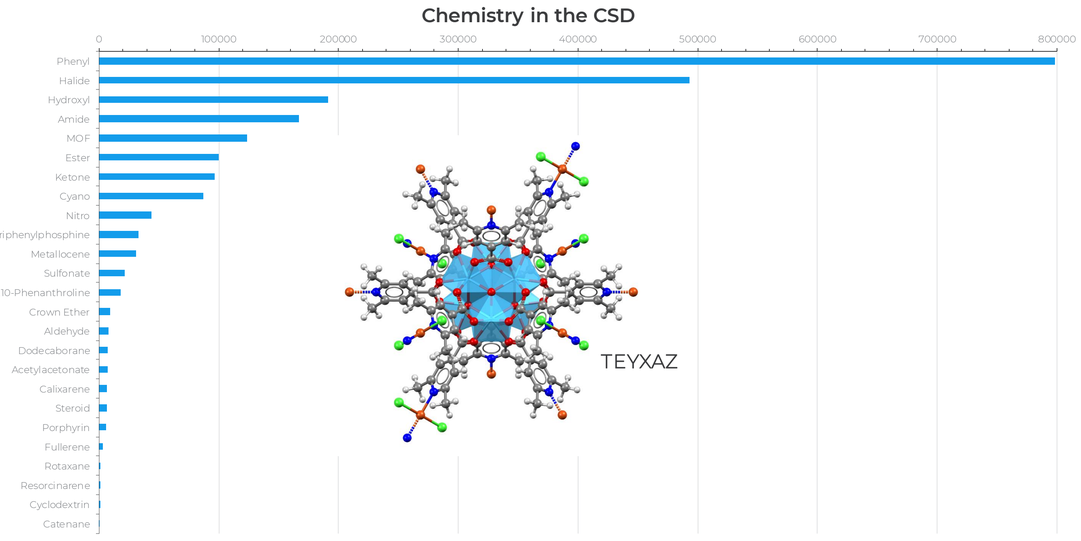
Nirmatrelvir – COVID-19 antiviral

FDA-approved 2023 as Paxlovid (with Ritonavir)

Drugs in the CSD: Top 200 Pharmaceutical Products by Retail Sales 2020

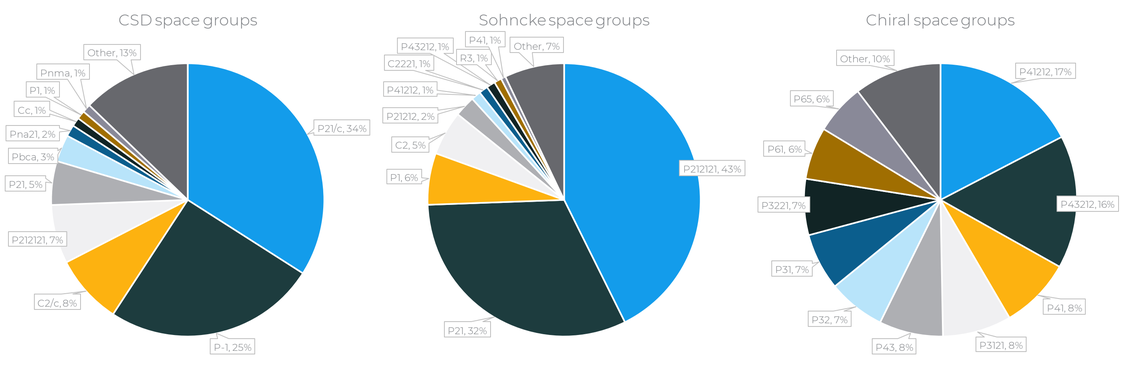


Chemistry in the CSD

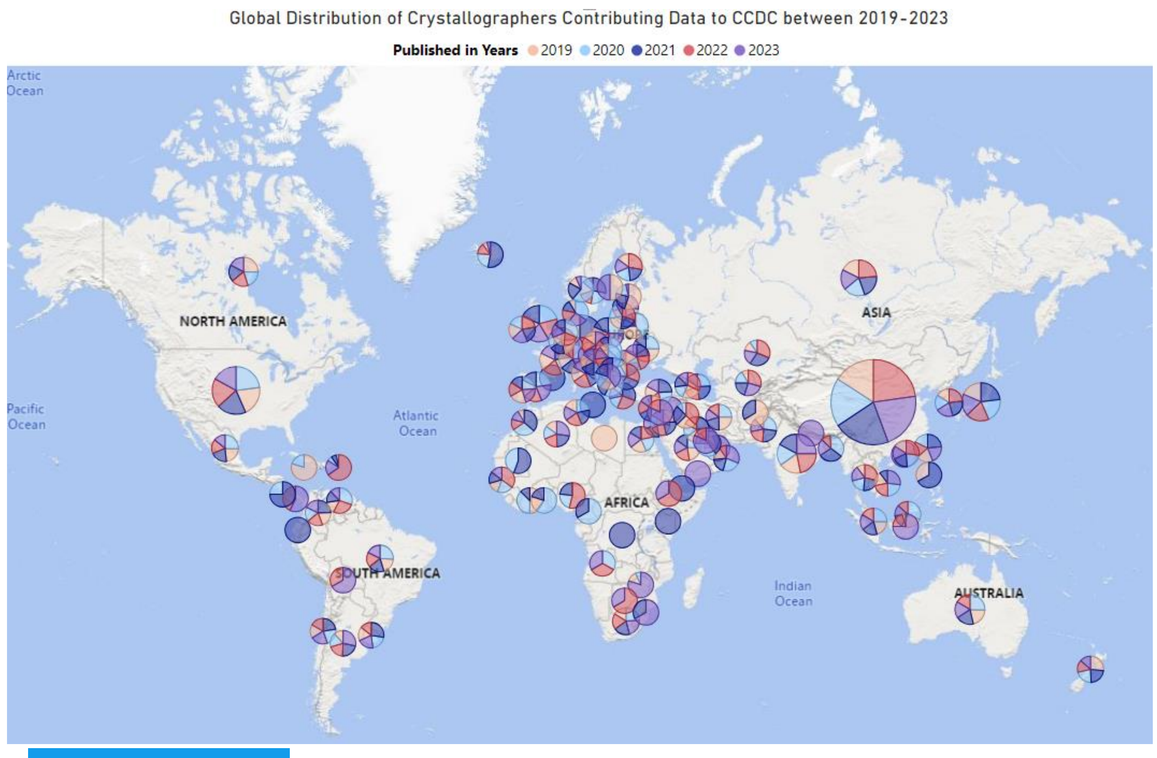


TEYXAZ

Space Groups in the CSD

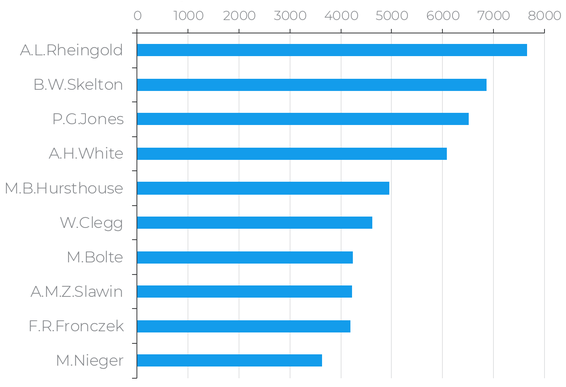


People, Places and Publications



Publication Leaderboard
Inorg.Chem.
Dalton Trans.
J.Am.Chem.Soc.
Organometallics
CSD Communication
Angew.Chem.,Int.Ed.
Chem.-Eur.J.
Acta Crystallogr., Sect.E:Cryst. Commun.
Chem.Commun.
Inorg.Chim.Acta

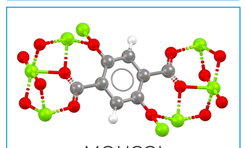
Top 10 Authors in the CSD



Most structures in a single publication

1,853 structures

B.Pato-Doldan, M.H.Rosnes, P.D.C.Dietzel, *ChemSusChem*, **2017**, 10, 1710, DOI: 10.1002/cssc.201601752



MOHGOI

CPO-27-Mg

Using the CSD

CSD-Core applications provide structural chemists with the tools to find structures with WebCSD, search the CSD and local in-house databases with ConQuest, visualise and analyse structures in Mercury, investigate conformations with Mogul, investigate interactions with IsoStar and create tailored scripts to answer targeted research questions using all the CSD's functionality with the CSD Python API.

The CSD-Materials suite provides users with additional tools for interactions with solid form informatics, including exploring intra- and intermolecular interactions, investigating crystal packing with packing features, similarity and motif searches and H-bond propensity analysis and statistics. Hydrate, solvate and aromatic analysers enable solid form risk assessment and users can create CSD-driven analyses and workflows with the CSD Python API. CSD-Particle can be used to understand particle behaviour and anticipate manufacturing bottlenecks to guide formulation decisions.

The CSD-Discovery suite adds computer-aided drug discovery tools to the features available in CSD-Core, with the ability to investigate protein-ligand docking with GOLD, validate receptor-bound ligand geometry and generate conformers derived from CSD data, and explore protein-ligand interactions with Full Interaction Maps and SuperStar. Users can also create CSD-driven Python API analyses and workflows, and mine CSD and PDB data for common interaction patterns with CSD-CrossMiner.

Colin R. Groom, *et al*., The Cambridge Structural Database, *Acta Cryst. B*, 2016, **72**, 171, DOI: 10.1107/S2052520616003954