Prepared for the BCA Spring meeting 2021

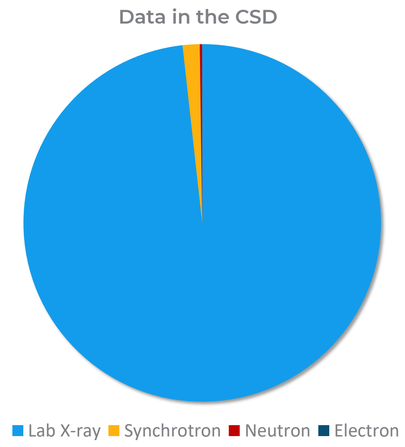
Electron Diffraction Data in the CSD

N. T. Johnson, S. B. Wiggin and S. C. Ward, The Cambridge Crystallographic Data Centre (CCDC)

Email: njohnson@ccdc.cam.ac.uk Website: www.ccdc.cam.ac.uk Twitter: ccdc_cambridge Facebook: ccdc.cambridge YouTube: CCDCCambridge

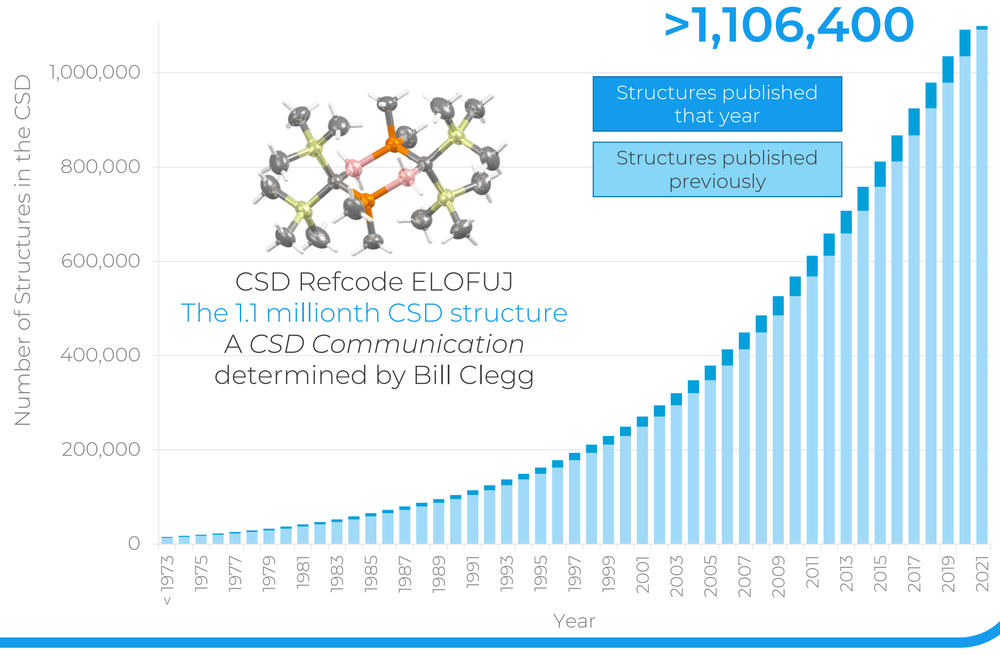
The CSD

The Cambridge Structural Database (CSD) is a curated database of over 1.1 million small molecule organic and metal-organic experimental crystal structures. Data is curated in house, validated and stored in a standardised format so that the structures are searchable, reusable and easily findable.



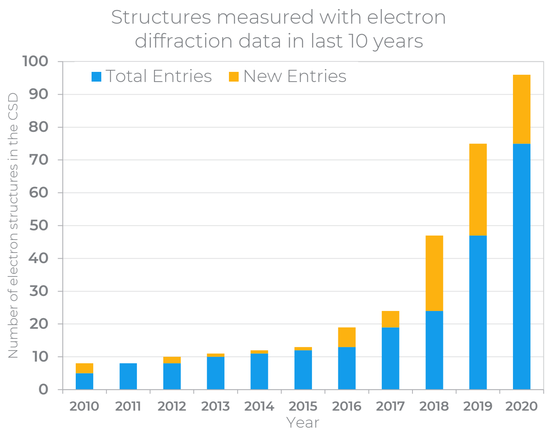
Alongside curating new structures, each year existing entries are enhanced with additional information or to improve the consistency of data and the searchability for users.

Data in the CSD is predominantly from small molecule single crystal experiments. As such, one focus for our CSD Improvement projects is the identification and labelling of non-routine experimental techniques. These investigations have mainly concentrated on synchrotron and neutron data, but we have recently turned our attention to the small but growing number of electron diffraction datasets.

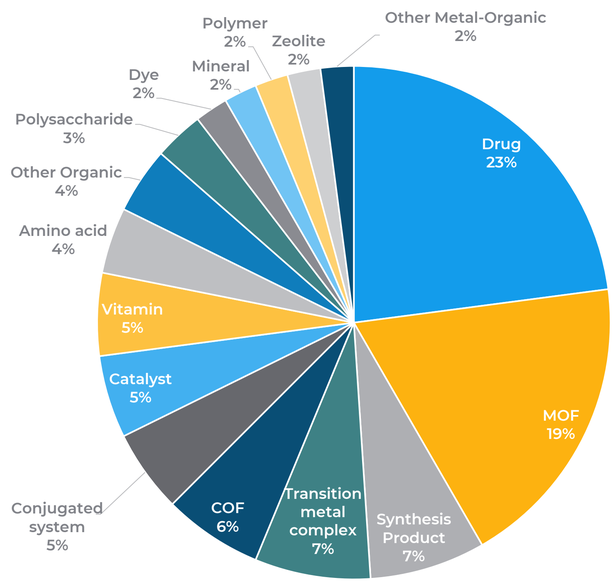


CSD Refcode ELOFUJ
The 1.1 millionth CSD structure
A *CSD Communication*
determined by Bill Clegg

Electron Diffraction Data

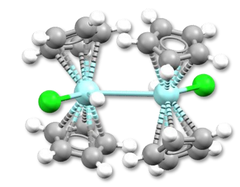


97 unique electron diffraction datasets have been identified in the CSD (v5.42 February update) with a wide variety of different compounds studied. The technique has had renewed interest due to developments in automated data collection and data processing.[1]

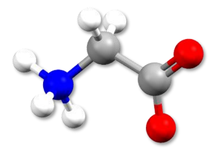


Classification of electron diffraction structures in the CSD

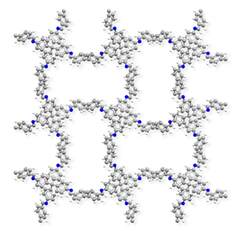
A selection of compounds measured with electron diffraction in the CSD.



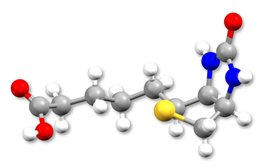
Schwartz reagent
DIZZUK



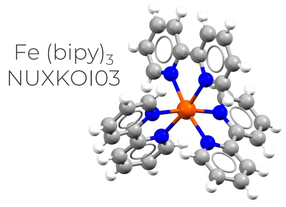
Amino acid (Glycine)
KUFDOH



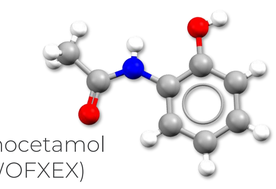
Covalent organic framework (COF)
BOBHOR



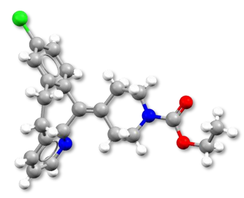
Vitamin B7
BIOTIN14



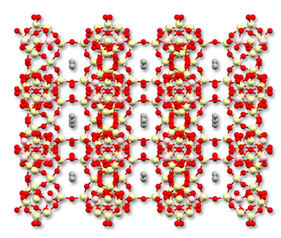
$Fe(bipy)_3$
NUXKOI03



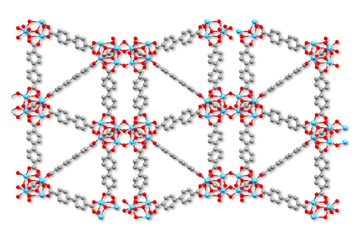
Orthocetamol
(WOFXEX)



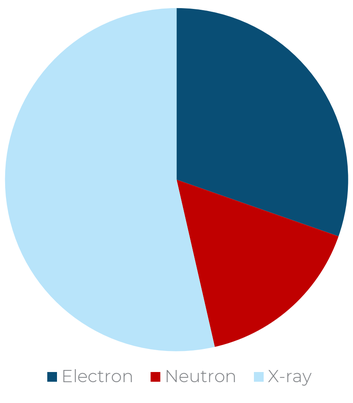
Loratadine (BEQGIN08)
Anti-allergy medication



Zeolite
IKUDOJ



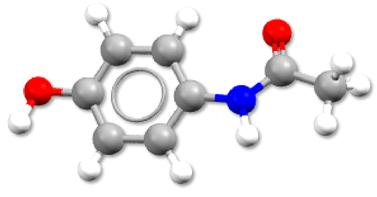
Nanoscale MOF
PAZDOM

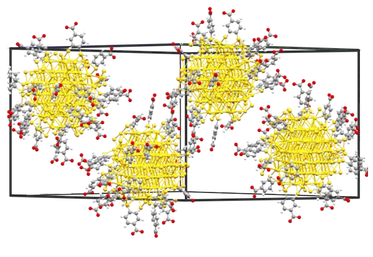


CCDC does not store raw diffraction data. However, a DOI for the raw data, stored in another repository, can be linked to the entry. This 'Raw Data DOI' can be provided during deposition to CCDC or in 'My Structures'.

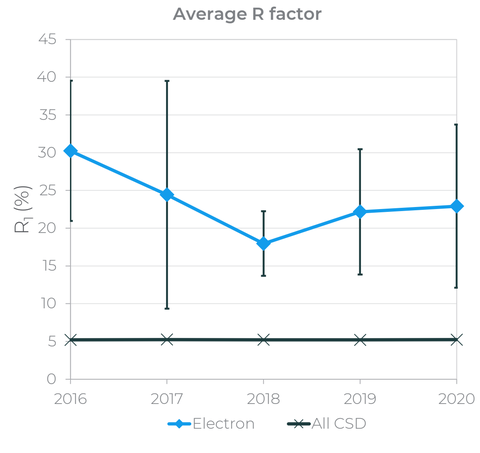
Currently, 56 structures in the CSD (0.005%) have an associated raw data DOI. Interestingly, 14% of electron diffraction structures have a raw data DOI associated with them.

Most Studied Structure
Paracetamol
(4 papers, 6 structures)





Largest Unit Cell Volume
VISXED,
Volume: 91696 Å^3
Gold cluster



The average R factor for electron diffraction data in the CSD is ~ 22% compared to a value of 5% for the whole CSD.

Higher R factors for electron studies may be expected due to dynamical scattering effects, amongst other factors.[2]

1. Gemmi *et al.* (2019). *ACS Central Science*. 5, 1315-1329.

2. Palatinus *et al.* (2013). *Acta Cryst.* A**69**, 171-188.

Identification

There are a number of automatic processes used during CCDC's Data Curation process to identify non-routine structures. These heavily rely on the information in the deposited CIF files. An assessment of electron diffraction data found 61% of data could be identified by the curation process, while other structures were identified after curation from other places in the CIF which are not checked by the curation software. A very small number of structures did not contain identifying information in the CIFs at all.

Identification method	% of CIFs
Automatic from CIF	61 %
Wavelength*	25 %
Information provided elsewhere	13 %
Only identifiable from paper	1 %

* Electron radiation wavelength is typically 0.02-0.03 Å

The CCDC has created an FAQ about electron diffraction data in the CSD, which provides the definition of suggested information to include in a CIF.

CIF field	Information to include
_diffrn_radiation_probe	'electron'
_diffrn_measurement_method	The method used to measure the data e.g. 'Rotation Electron Diffraction'
_diffrn_radiation_type	'electron'
_diffrn_source	The general electron source used in experiment e.g. 'electron microscope'
_diffrn_source_type	Specific information about the electron source used, such as make/model or name of the equipment
_diffrn_measurement_device	The general name of the equipment used to collect the diffraction data
_diffrn_measurement_device_type	Specific make/model of the equipment used e.g. 'FEI Tecnai F30ST'
_diffrn_detector	The general name of the detector used
_diffrn_detector_type	Specific make/model of the detector used
_exptl_special_details	Any additional special details about the experimental setup
_diffrn_special_details	Any additional special details about the diffraction data collection

Future Developments

Although the current number of electron diffraction datasets in the CSD is small, it is important that they are properly curated so they are findable and useful to users. We would like to hear your input on any ways we can support the growth of this technique.

The CCDC is also planning other improvement work, for instance the standardisation of property fields such as melting point data and undertaking an analysis of polymorphic data in the CSD to see how this information could be enhanced to increase consistency.

We would also like to hear your input of what future projects that CCDC should undertake to help users to get the most out of the CSD. Please share your thoughts to **hello@ccdc.cam.ac.uk.**