

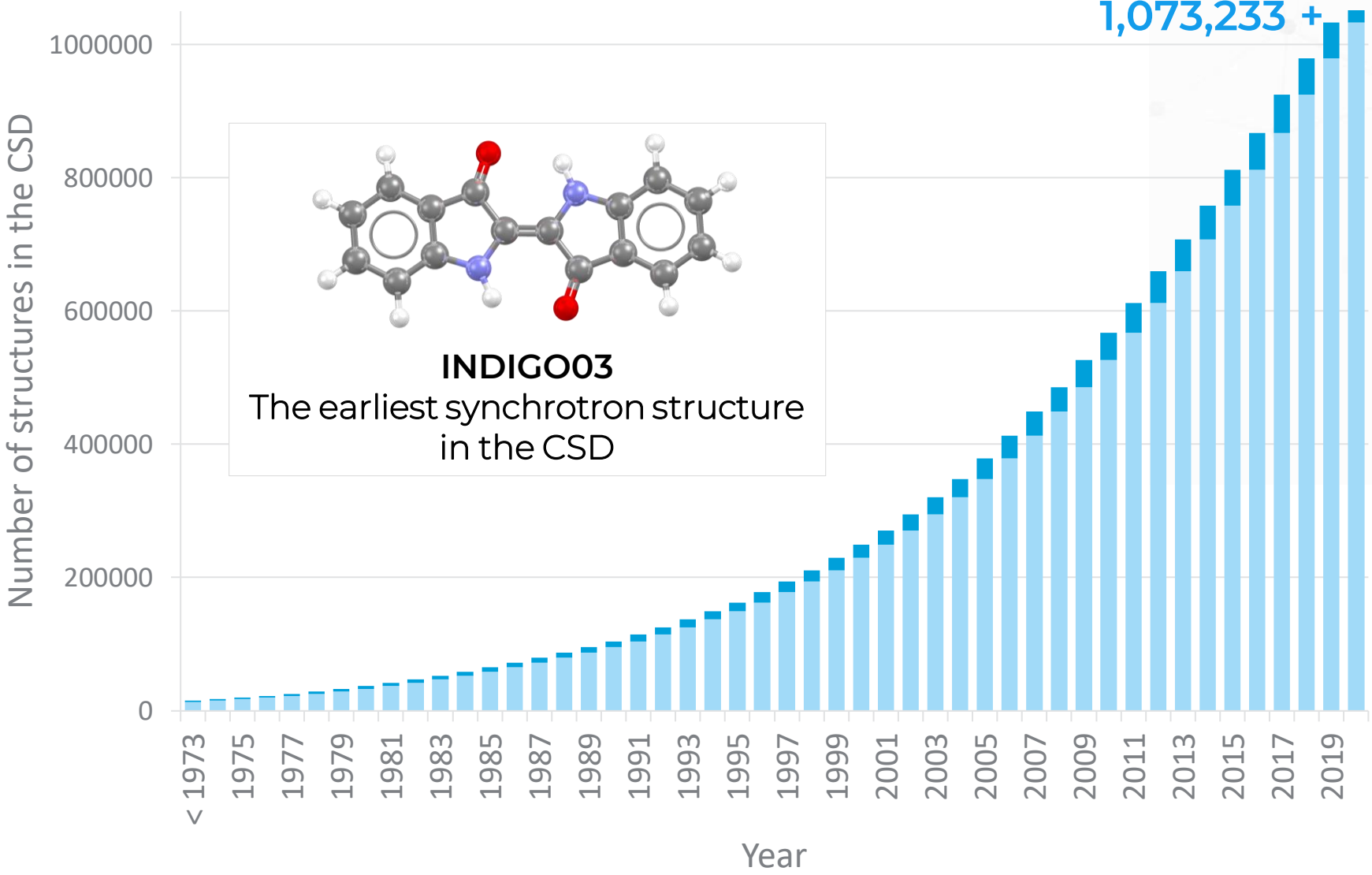
Synchrotron data in the CSD

Natalie Johnson

Data Integrity Research Scientist

22nd August 2020

The Cambridge Structural Database (CSD)



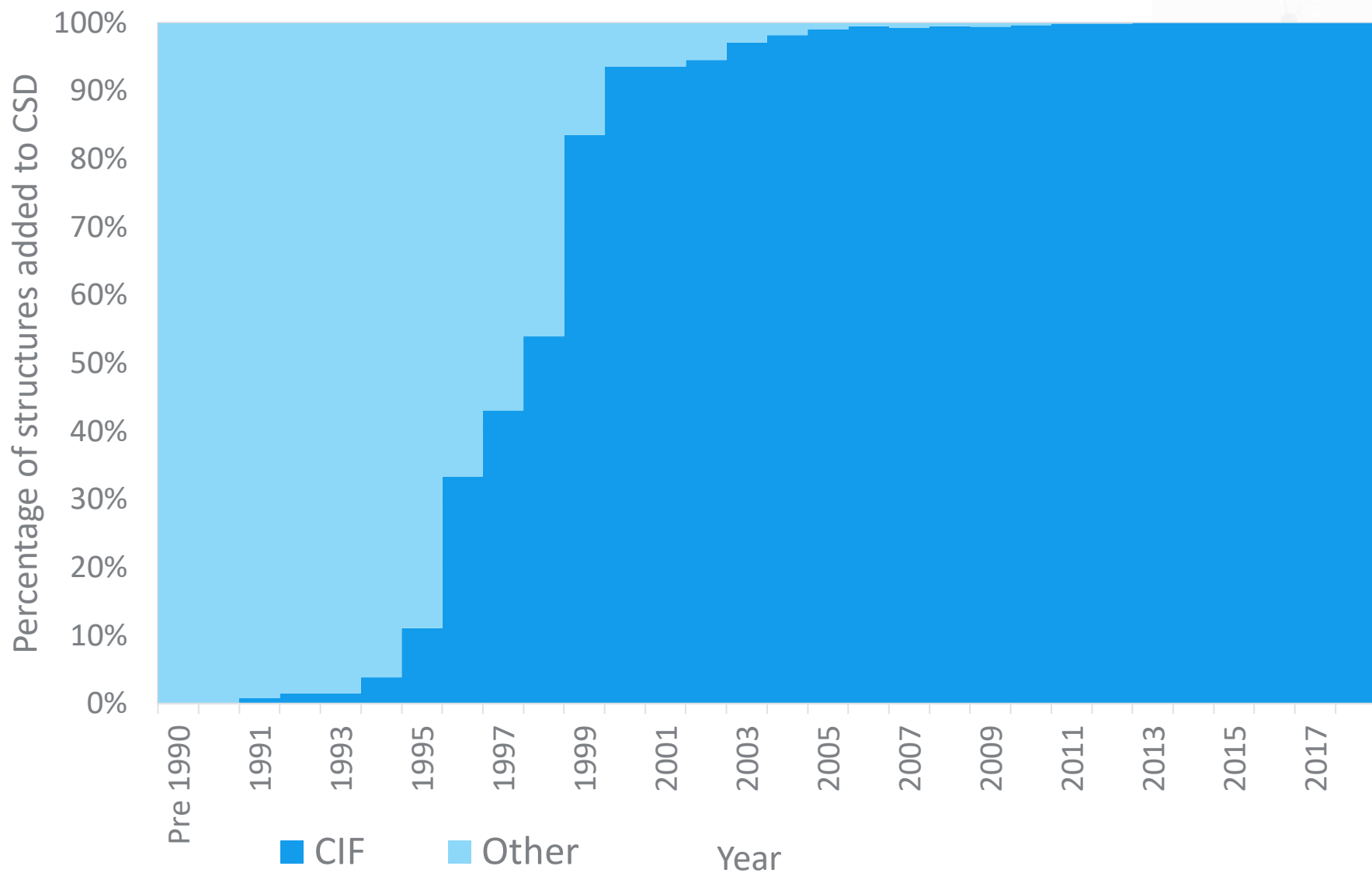
The CSD is a database of small molecule organic and metal-organic crystal structures.

Every entry enriched and annotated by experts.

Every published structure:

- Inc. ASAP & early view
- *CSD Communications*
- Patents
- University repositories

The Cambridge Structural Database (CSD)



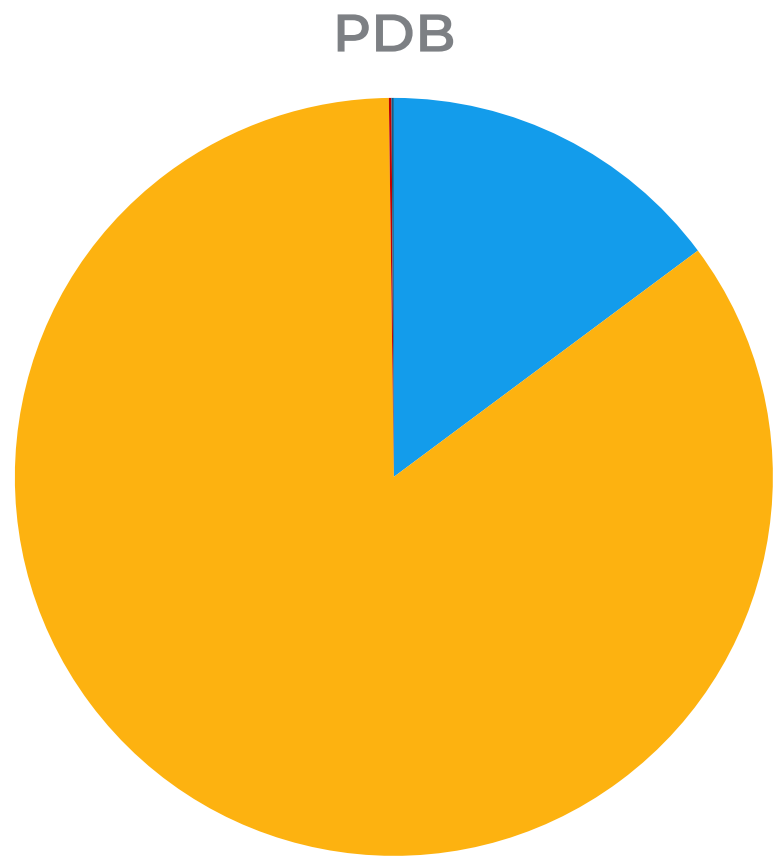
Data can be deposited to CCDC through an online deposition platform or via email.

Pre-CIF, entries were created by typing data from tables in papers.

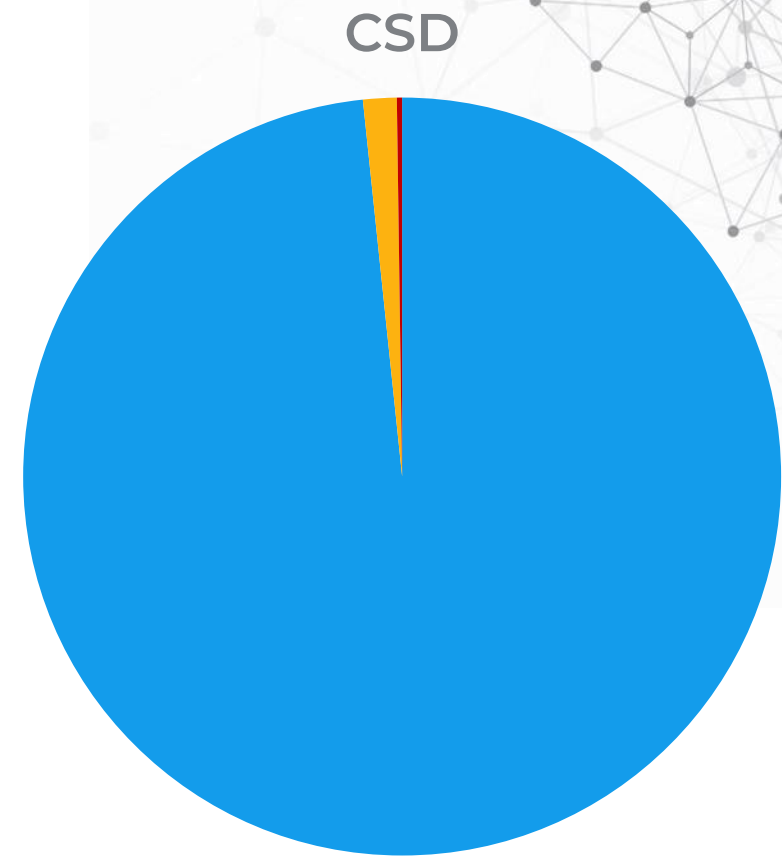
When CIF is not available, CCDC manually creates a CIF from the accessible information.

365 structures (< 0.6%) added to the CSD with a manually created CIF in 2019.

Synchrotron data in the PDB vs CSD



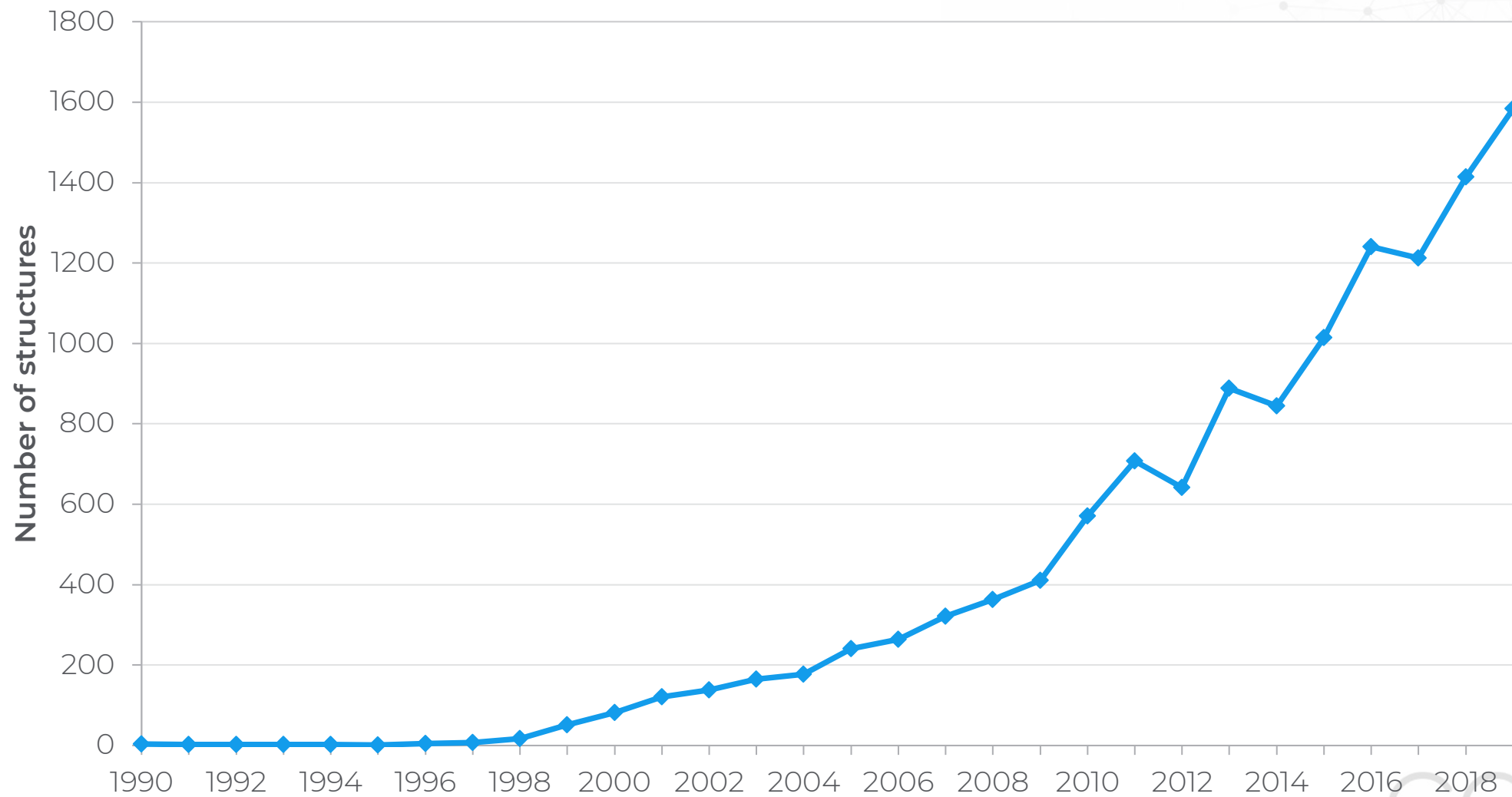
■ Lab X-ray ■ Synchrotron ■ Neutron ■ Electron



■ Lab X-ray ■ Synchrotron ■ Neutron ■ Electron

1 PDB Statistics from Biosync (<http://biosync.sbkb.org/>) and PDB website, accessed 25/06/2020
Diffraction data only.

Synchrotron data in the CSD

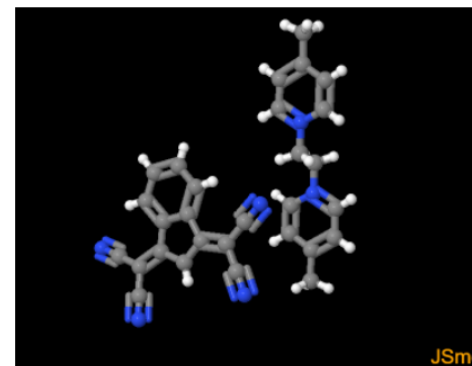


Identification of synchrotron data

- Structures are flagged with 'synchrotron' in the CSD.
- These structures are identified during the data curation process either automatically by the curation software, which checks a select number of CIF fields, or manually labelled by an Editor.

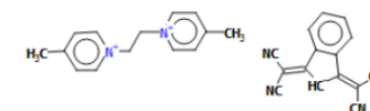
JOPXAQ : 1,1'-(ethane-1,2-diyl)bis(4-methylpyridin-1-ium) bis[1,3-bis(dicyanomethylidene)-2,3-dihydro-1H-inden-2-ide]
Space Group: $P\bar{1}(2)$, **Cell:** a 8.3232(3)Å b 9.7414(4)Å c 10.7328(4)Å, α 87.690(3)° β 83.628(3)° γ 84.506(3)°

3D viewer



Style:
 Labels:
 Packing:
 Measure:

Chemical diagram



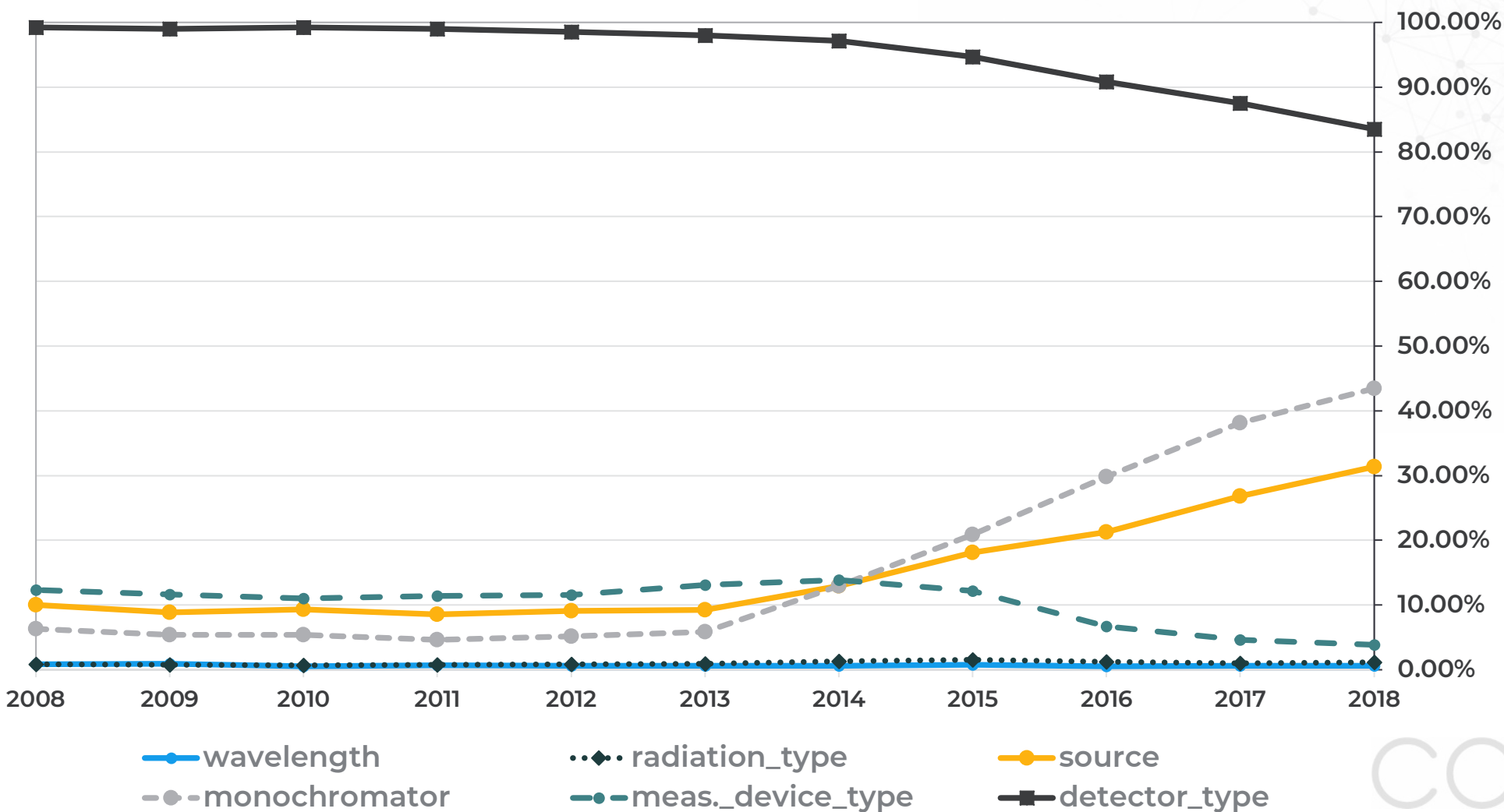
[View group symbols key](#)

Experimental details

R-factor (%)	4.12
Temperature (K)	90
Density (CCDC)	1.3446
Radiation probe	x-ray
Radiation source	synchrotron

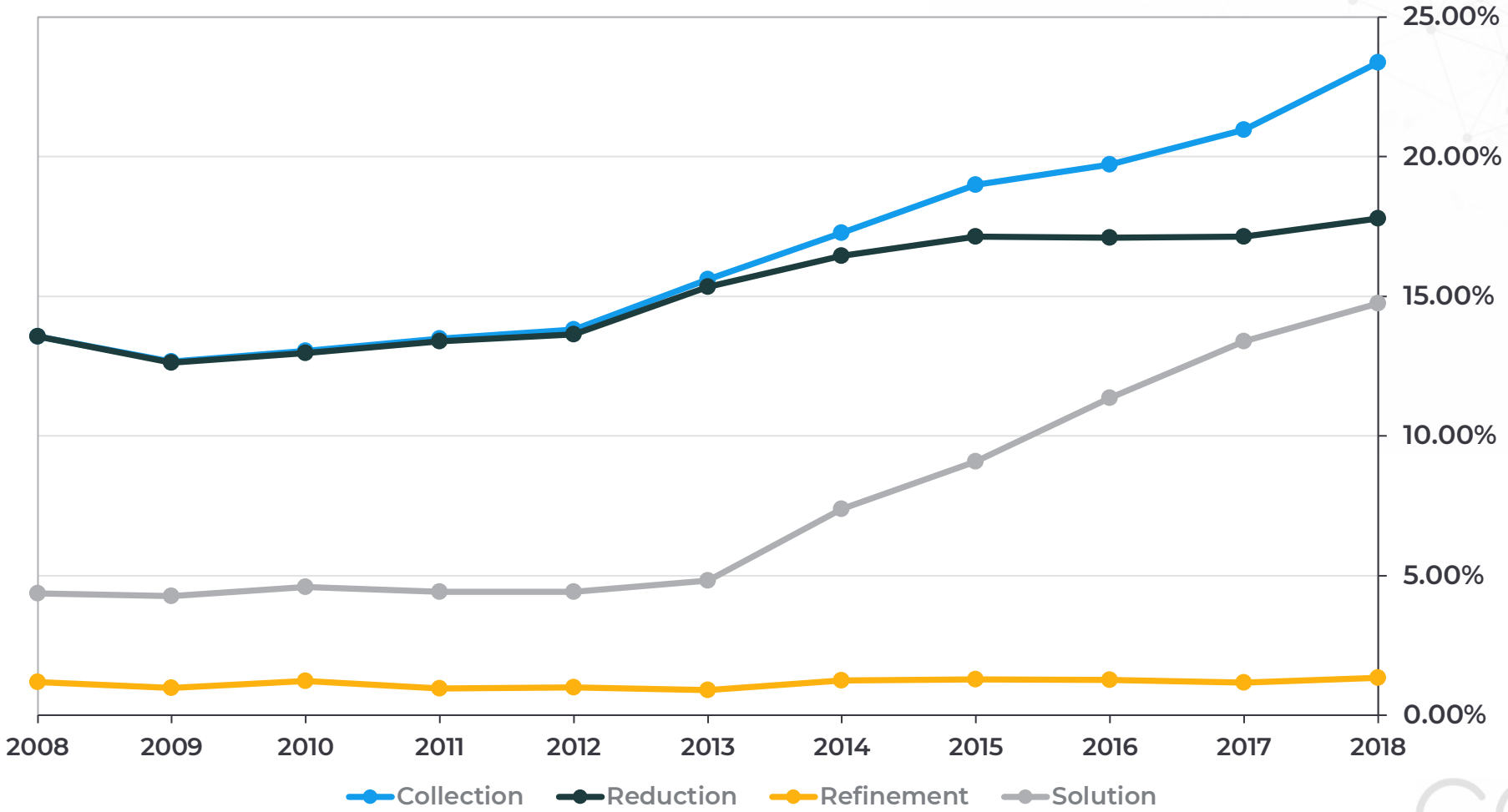
CIF completeness - Experimental

CIFs which did not contain information in fields



CIF completeness - Computational

CIFs which did not contain information in fields

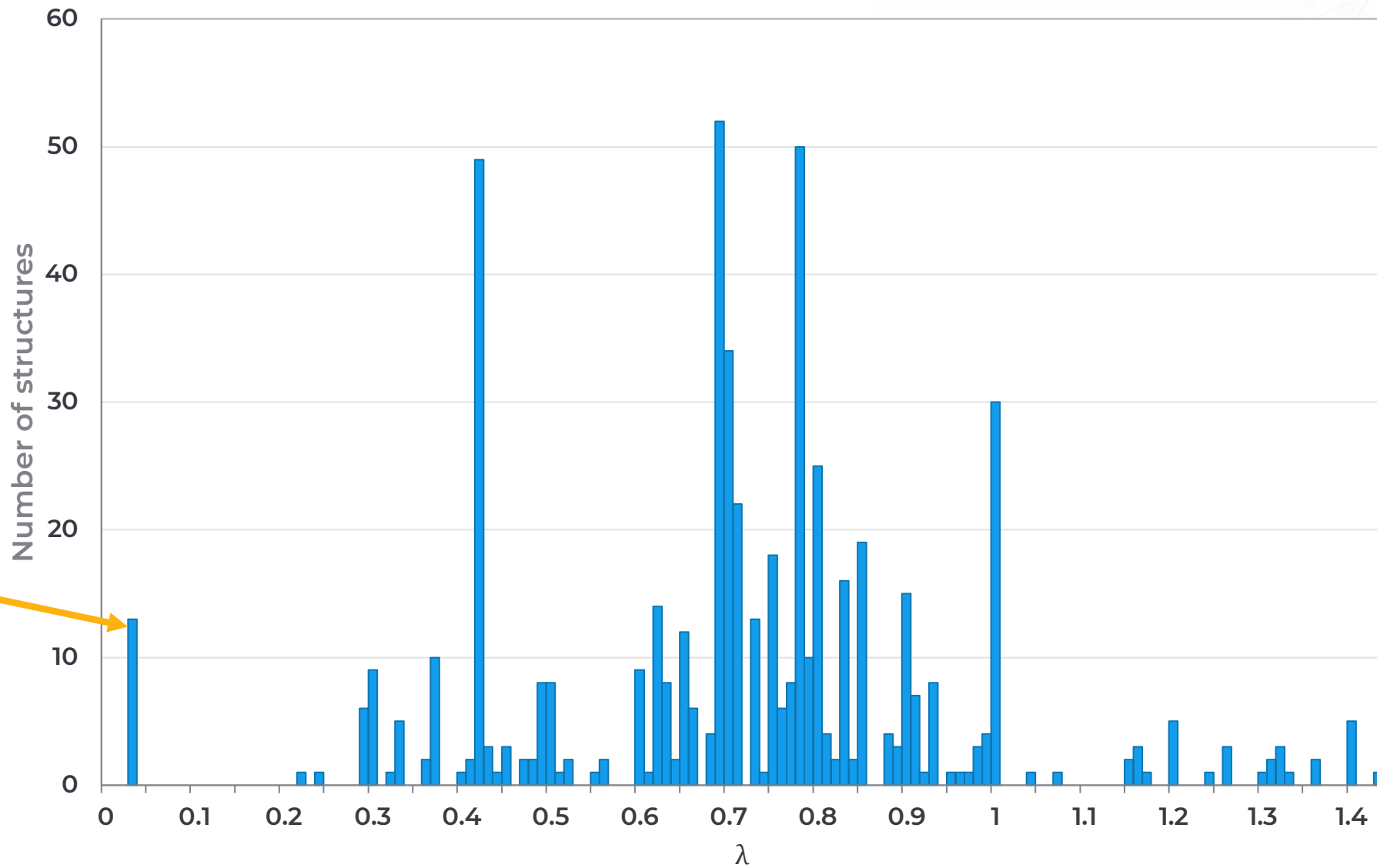


CIF completeness – Outcomes

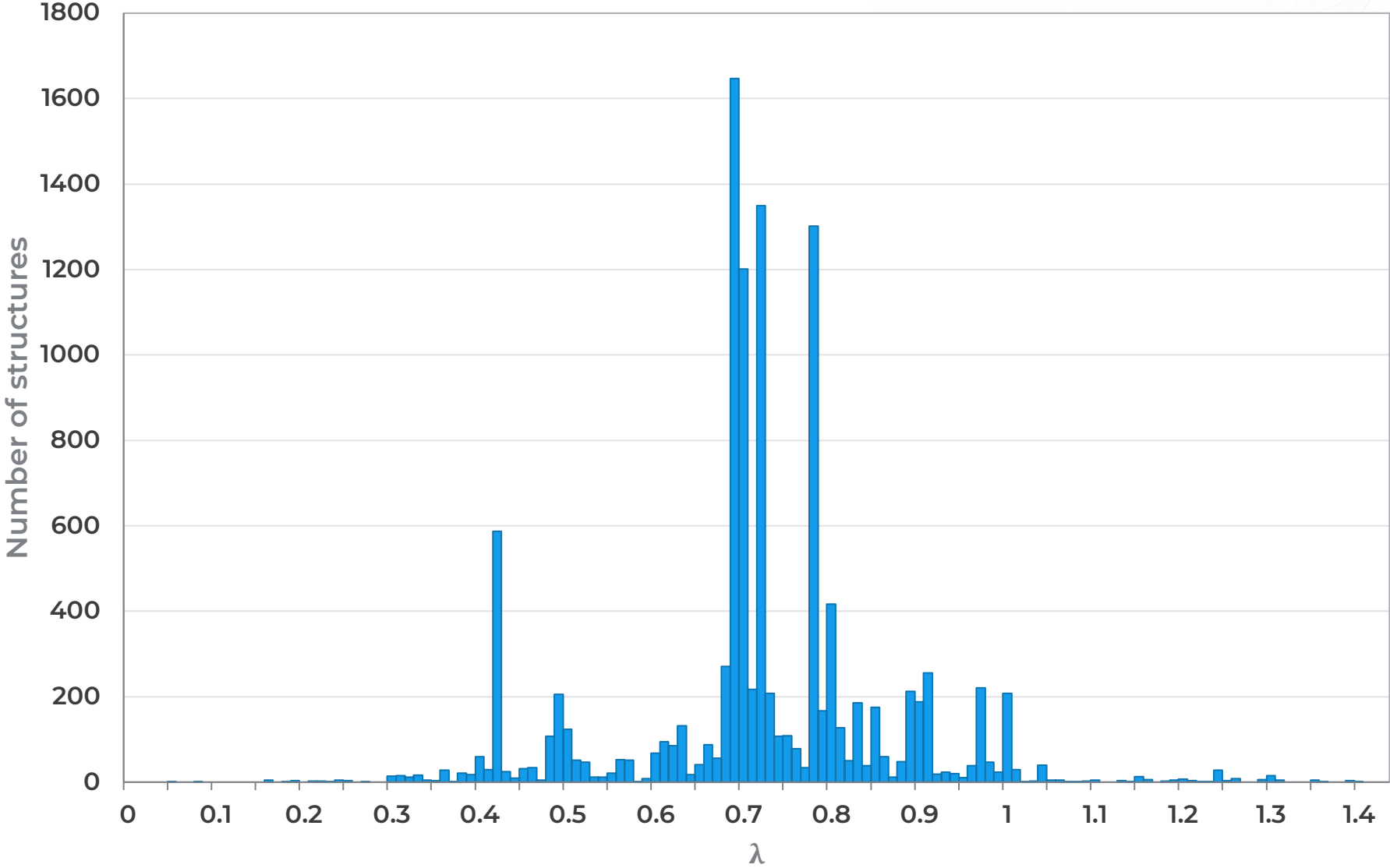
- Some structures were not identifiable from the CIF as it did not contain complete information.
 - 25 structures found by cross-referencing papers associated with the Australian Synchrotron and structures in the CSD.
 - ~ 600 structures report a ‘non-standard’ wavelength, yet contain no synchrotron identifying information.
- Other inconsistencies found in a number of synchrotron CIFs e.g. radiation_type ‘synchrotron’ and _diffrn_source ‘sealed tube’/‘rotating anode’

Structures with a non-standard wavelength

Electron

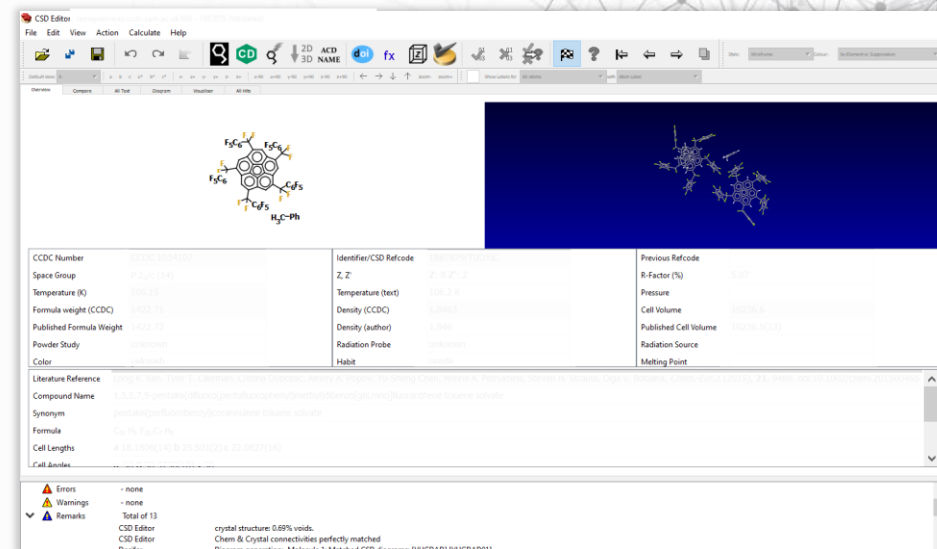


Wavelength of synchrotron structures



FAIR and FACT small molecule data

- Curation to ensure structures are labelled and presented consistently so are **findable** within the CSD.
- Option to download additional information associated with entry in WebCSD/Access Structures.

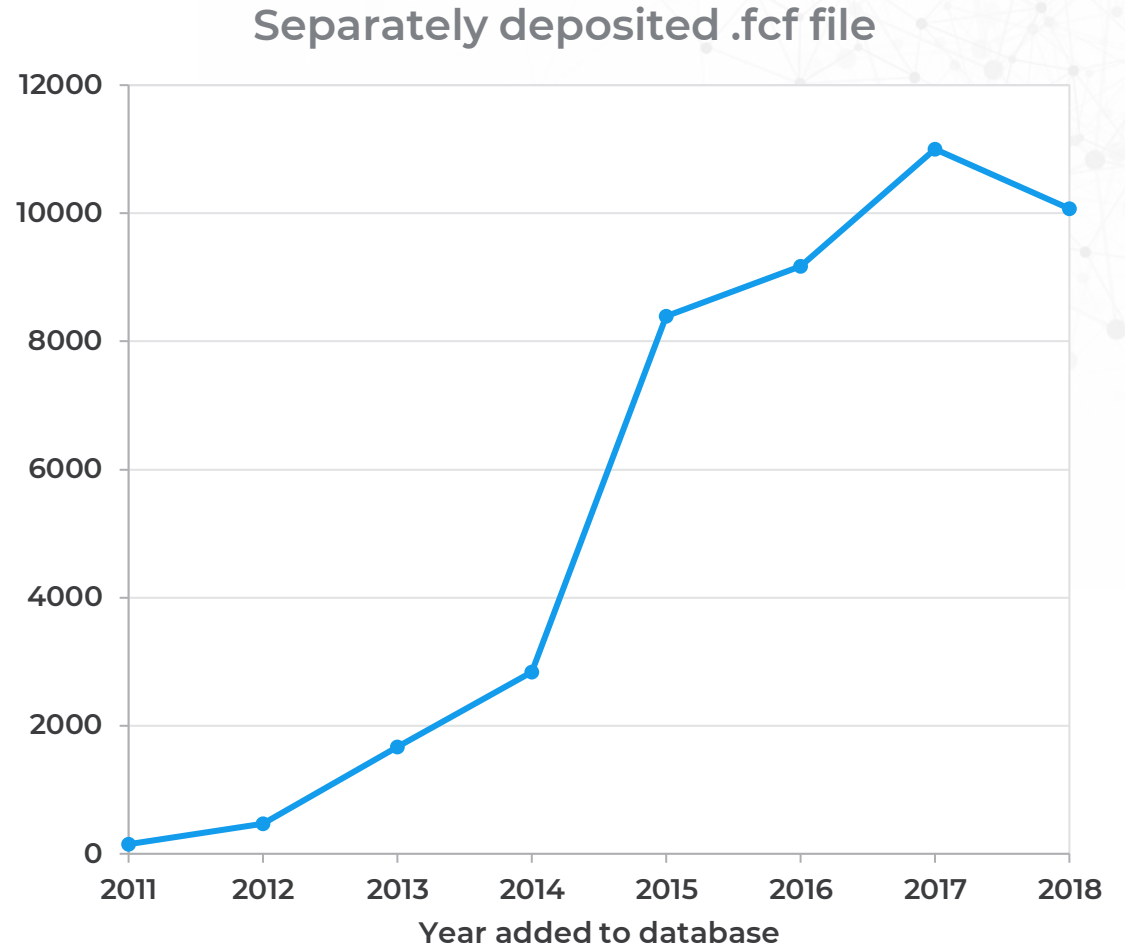


Download deposited CIF

- Deposited CIF(s)
- Deposited CIF(s) without structure factor data
- Deposited file(s) with any available structure factor data and checkCIF reports included
- Include checkCIF reports as a PDF when available

Reflection intensities

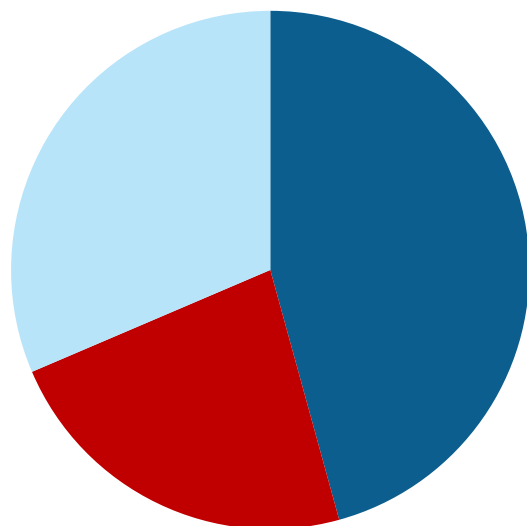
- The deposition of structure factors is not explicitly required by CCDC.
- The number of separately deposited structure factor files is tending to increase.
- Many programs automatically include hkl, res and even structure factor information within the CIF.



Raw Diffraction Data

- CCDC does not currently store raw diffraction data. However, a DOI for the raw data can be linked to the entry.
- 35 structures in CSD have raw data DOI.

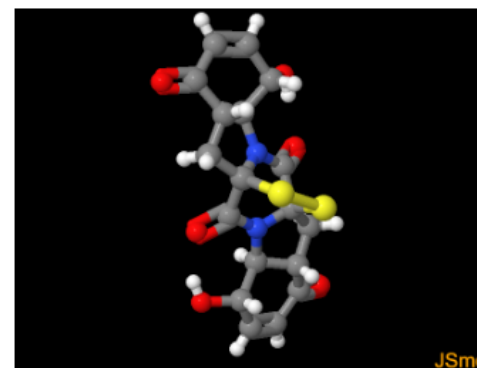
Structures with Raw Data DOIs



■ Electron ■ Neutron ■ X-ray

BISGAO : 4, 11-dihydroxy-4,4a,7,7a,11,11a,14,14a-octahydro-1H,6H,8H,13H-6a,13a-epidithioprazino[1,2-a:4,5-a']diindole-1,6,8,13-tetrone
 Space Group: P 2₁ 2₁ 2₁ (19), Cell: a 10.996(2)Å b 12.452(2)Å c 13.218(3)Å, α 90° β 90° γ 90°

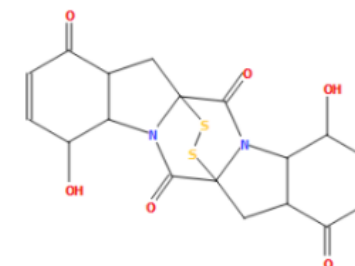
3D viewer



H Disorder Menu Open

Style Labels Packing Measure
 Ball and Stick No Labels None None

Chemical diagram



View group symbols key

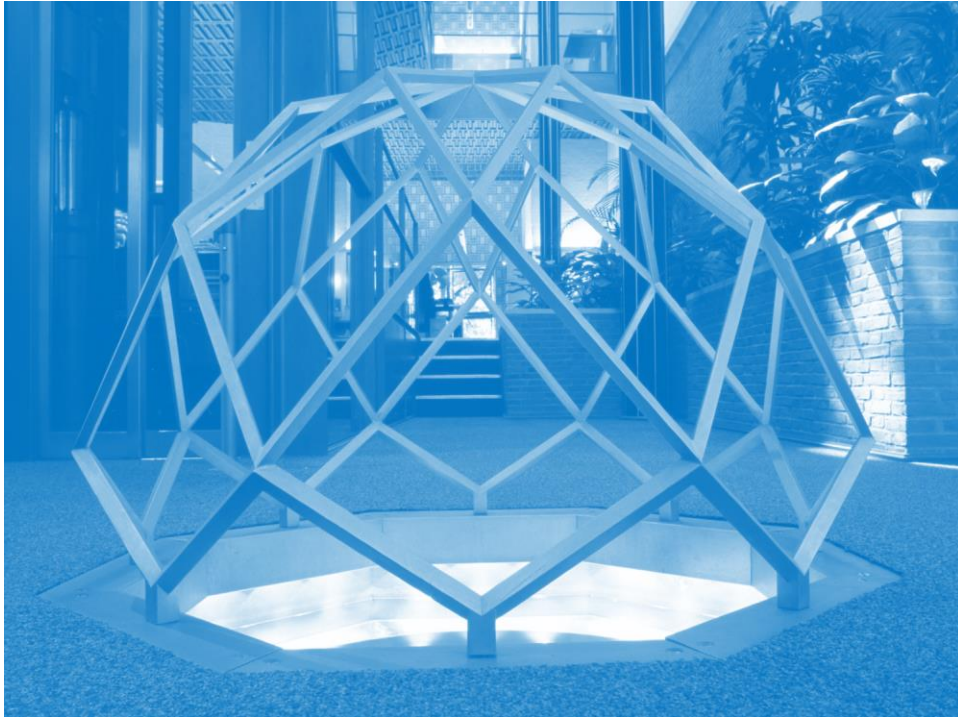
Additional details

Deposition Number	1870981
Data Citation	M.T.B. Clabbers, T. Gruene, E. van Genderen, J.P. Abrahams CCDC 1870981: Experimental Crystal Structure Determination, 2018, DOI: 10.5517/ccdc.csd.cc20sx7z
Synonyms	Epicorazine A
Deposited on	14/11/2018

Raw data DOI(s)

DOI	10.5281/zenodo.1407682
-----	------------------------

Thank you for listening



Acknowledgements

Stephanie Bowers, Jason Price (Australian Synchrotron), **Mike Hoyland, Brian McMahon, Peter Strickland** (IUCr), **Vasily Bunakov, Brian Matthews** (STFC), **Suzanna Ward, Seth Wiggin** (CCDC), **Simon Coles, John Helliwell, Mark Warren**, and **Amy Sarjeant**