The Cambridge Structural Database 2019

What's New?

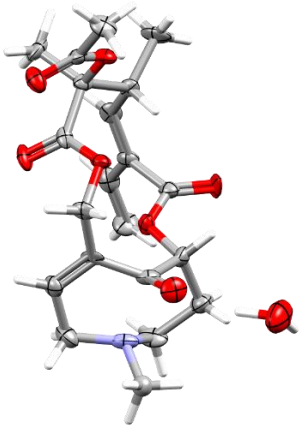
The Cambridge Structural Database (CSD)

The CSD is the only comprehensive and up-to-date database of fully validated organic and metal-organic crystal structures. The 2019 CSD Release contains 957,868 unique structures and 973,630 entries (CSD version 5.40) – an increase of more than 57,000 entries. We are currently on course to reach a million structures by summer 2019! Used by scientists worldwide, the CSD provides the complete crystal structure database for chemists working with organic and metal-organic compounds.

We have now completed work to revolutionise the underlying technologies in the CSD system enabling us to start broadening the data available through the database. This year's release increases the number of entries with anisotropic displacement parameters (ADPs) and occupancy factors, with ADPs now available in >675,000 entries (~75% of CSD structures). This work has also resulted in a streamlined system, meaning that this year's installation has been reduced in size by 4GB.

The 2019 CSD release also contains more targeted enhancements to existing entries. Over 80,000 entries have been been updated to give additional information and ensure more consistent search results. Enhancements this year include:

- Comprehensive review and enhancement of another 14,000 historical entries
- Addition of a further 15,000 metal oxidation states, meaning the CSD now contains oxidation state information for almost 300,000 entries
- Article DOI links added to over 8,000 more entries



The structure of the alkaloid clivorine illustrating the Bürgi-Dunitz angle, with ADPs displayed at the 80% probability level (refcode: CLIVOR11, DOI: 10.5517/ccdc.csd.cc16txtr)

CSD-System – Harness knowledge-based science

For all users of CSD-System, CSD-Materials, CSD-Discovery or CSD-Enterprise

The CSD-System brings you essential structural chemistry capabilities to deliver knowledge from the CSD: powerful 2D/3D search, extensive geometry and interaction analysis, high impact graphics, as well as connectivity via the CSD Python API.

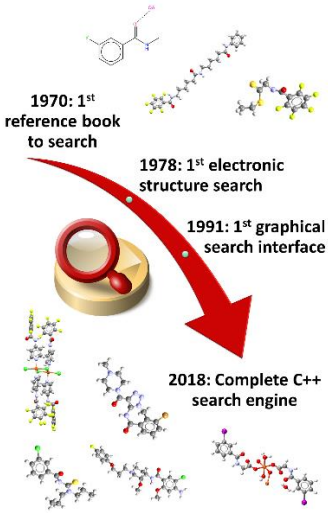
In response to feedback from users and a desire to further develop our systems in the future in new directions, we have now finished rewriting our search engine in C++. The original search engine behind ConQuest was written in the 1970s and was tied to a specific database format (ASER). Since the development of Mercury, we've been writing new, reusable chemistry algorithms in C++ which have been slowly replacing the older parts of the CCDC code base and this has now fully replaced the old code.

The rewriting of ConQuest's fundamental search engine means that:

- Many complex 2D substructure searches are now much faster
- Nearly all types of search are significantly faster than the previous version
- Element and formula searches are more effective
- Many text searches now give more accurate results, especially author searches which now properly handle international name conventions

During 2018, targeted improvements have also been made to WebCSD based on user feedback including:

- Ability to expand the 3D visualiser to full screen
- New formula searching option
- Additional templates for structure search



The development of CSD searching since 1970

CSD-Discovery – Discover new molecules

For all users of CSD-Discovery or CSD-Enterprise

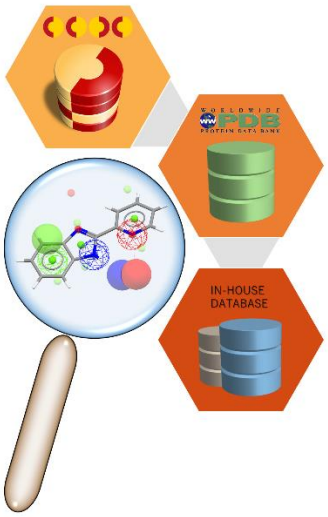
CSD-Discovery provides in one place all the tools you need for discovering new molecules.

Last year, we introduced a new application to CCDC's portfolio: CSD-CrossMiner. During 2018, we've improved and enhanced this software to make it more memory efficient when storing large numbers of hits.

This year we've focussed on improving programmatic access to CSD-CrossMiner through our CSD Python API. The new release allows users to automate pharmacophoric searches in their own workflow using these methods.

We've added a simple but much requested feature in the latest version. One common request for CSD-CrossMiner is to be able to filter down results to remove classes of hits. It's now possible to express 'non-3D' filters on substructures in hit results, so users can restrict hits to remove or only include compounds that contain a particular substructure.

2018 also saw delivery of new CSD Python API methodologies for non-sequence-based cavity searching. The CSD Python API now contains 3 different approaches for cavity & pocket searching in proteins which allow a trade-off between very high speed and very high accuracy.



CSD-CrossMiner: dynamic 3D pharmacophore searching across the CSD, PDB and in-house databases

CSD-Materials – Engineer new materials

For all user of CSD-Materials or CSD-Enterprise

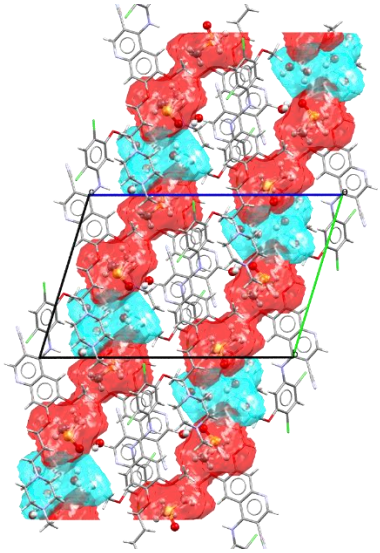
CSD-Materials helps you explore exciting new materials through analysing intra- and intermolecular interactions within the lattice, allowing you to understand your material's behaviour and refine its properties.

This year we have introduced a new component that allows you to effortlessly study the structure of even the most complex solvate systems. The Solvate Analyser allows you to easily identify any number of unique solvents, co-formers or counter-ions and quickly visualise their packing as well as any relevant hydrogen bonds involving the identified components.

We have also introduced three new modules to the CSD Python API in the CSD-Materials area covering H-bond propensity, H-bond coordination and morphology. This enables advanced research projects using these areas of functionality and there is also a selection of example scripts illustrating how users can run complex analyses.

New scripts provided alongside CSD-Materials include:

- Automated polymorph risk assessment *via* hydrogen-bond propensity
- Automated co-crystal screening *via* hydrogen-bond propensity
- Generation of 3D printable morphology object for a crystal structure



The structure of bosutinib DMSO solvate trihydrate with Solvate Analyser display of DMSO space shown in red and water space in light blue (refcode: ABECES, DOI: 10.5517/ccdc.csd.cc1m68kn)

info@ccdc.cam.ac.uk
www.ccdc.cam.ac.uk
UK: +44 1223 336408

Doc no. 108CSDE19

The Cambridge Crystallographic Data Centre