



Cambridge Structural Database (WebCSD)

Matt Hayward

STEM Librarian

University of Texas at San Antonio

matt.hayward@utsa.edu

Cambridge Structural Database (CSD) is a chemistry resource compiled and distributed by the Cambridge Crystallographic Data Centre (CCDC). CSD contains a highly-detailed and complete record of all published organic and metal-organic small-molecule crystal structures. CSD is considered the authoritative source for finding and sharing structural chemistry data ([Groom et al. 2016](#)). WebCSD, the online implementation of CSD, is freely available on the internet, although a subscription and individual account is required for advanced searching. The CSD Software System, which includes ConQuest, IsoStar, Mercury, PreQuest, Mogul, and Python API, is available for annual subscription.

CCDC is a non-profit charitable organization started in 1965 by the Organic Chemistry Department at the University of Cambridge ([Groom & Allen 2014](#)). CSD is a vast repository for experimentally-determined small-molecule crystallography data and structures. The database is continually updated with new structures visible in the database within moments of user deposition. CSD contains a complete record of all published organic and metal-organic small-molecule crystal structures. As of this writing, the database contains over 970,000 entries.

The CSD Software System is intended for in-depth and comprehensive crystallographic searching by advanced users with expert knowledge in crystallography, such as crystallographers, structural chemists, and the drug design community ([Thomas et al. 2010](#)). Therefore, this review will focus on WebCSD as the web implementation, aimed at the medicinal and pharmaceutical chemists, but it is more likely to be used by librarians and students. WebCSD is also an excellent tool for chemical education. However, that content has been reviewed in numerous chemical education journals ([Battle and Allen, 2012](#); [Battle et al. 2011](#); [Battle et al. 2010](#)) and falls outside the scope of this publication.

Searching CSD

CSD offers several options for searching, which are arranged in a tabbed frame. Each tab represents a different search type, with various options for search and refinement within the design. The search types are Simple, Structure, Unit Cell, and Formula.

Simple Search – default landing page, search by:

- Identifiers – CCDC/CSD number(s), CSD Refcode(s), or ICS Number(s)
- Compound name – for instance Sodium Chloride
- DOI – publication or CSD DOI
- Authors – Publication author(s)
- Journal – Publication journal title
- Publication Details – Year, volume, page

The screenshot shows the 'Simple Search' tab of the WebCSD interface. At the top, there are logos for The Cambridge Crystallographic Data Centre, FIZ Karlsruhe, and WebCSD, along with a 'Sign In' button and a license statement 'Licensed to: University of Texas San Antonio'. Below the tabs, a message states: 'Welcome to WebCSD. This service now includes the ability to search for inorganic structures through the CCDC's and FIZ Karlsruhe's joint Access Service using the Simple Search tab. Please use one or more of the boxes to find entries. If you enter details in more than one field the search will try to find records containing all the terms entered. More information and search help'. The search fields include: 'Identifier(s)' (placeholder: CCDC Number(s), CSD Number(s), CSD Refcode(s) or ICS Number(s)), 'Compound name' (placeholder: e.g. sulfadiazine), 'DOI' (placeholder: A single publication DOI, CSD DOI or ICS DOI), 'Authors' (placeholder: e.g. F.H. Allen), 'Journal' (placeholder: e.g. Journal of the American Chemical Society), and 'Publication details' (Year, Volume, Page). There are also radio buttons for 'Database to search': 'Entire published collection' (selected), 'CSD', 'ICSD', and 'Teaching subset'. A '+ Add New Search Field' button and a 'Search' button are at the bottom. A 'Clear' button is on the right. At the very bottom, there are links: 'CCDC Home', 'Deposit Structures', 'Access Structures', and 'About This Service'.

Figure 1. Simple search

Structure Search – users can draw chemical structures within a Java applet to search for an exact structure, substructure, or similarity. Common rings and elements can be selected from the toolbar. A periodic table and the hand-drawing tools are available for others.

The screenshot shows the 'Structure Search' tab of the WebCSD interface. At the top, there are tabs for 'Simple Search', 'Structure Search' (selected), 'Unit Cell Search', and 'Formula Search'. Below the tabs, the text 'Chemical structure searching' is displayed. A message says: 'Please draw your diagram or add a SMARTS string in the "advanced" section below.' The main area contains a chemical structure drawing toolbar on the left with elements C, N, O, S, H, F, Cl, Br, I, and a '+', a central drawing canvas showing a naphthalene ring with a carbonyl group, and a 'dotmatics' logo at the bottom right. Below the canvas, there are radio buttons for 'Match condition': 'Exact', 'Substructure', and 'Similarity' (selected). There is an 'Advanced' button with a dropdown arrow. At the bottom left are 'Search' and 'Clear' buttons. On the right side, there is a 'Help' section with 'Keyboard shortcuts' (Copy: Ctrl-C, Delete: Ctrl-X, Paste: Ctrl-V, Undo: Ctrl-Z, Redo: Ctrl-Y, Select all: Ctrl-A) and 'Query features' (Query features describe how an atom or bond should behave in substructure searches. To add a feature: 1. Right click on atom or bond, 2. Hover over 'query features' (atom only), 3. Hover over a feature type (e.g. H-count, type), 4. Select one of the options).

Figure 2. Structure search

Unit Cell Search – search by lattice centring (e.g., primitive, rhombohedral, a-, b-, c-, face- or body-centered) as well as cell lengths and angles.

Figure 3. Unit Cell search

Formula Search – search by molecular formula components (e.g., C5 H6 O2)

Figure 4. Formula search

Search Results

Following a query, the user is presented with a list of reference codes for structure or substructure matches (depending on query type and user selection) and the individual record for the first item in the list.

Simple Search Structure Search Unit Cell Search Formula Search			
Your query was: Compound name: benzene and the search returned more than 30 records.			
Select all Download Selected View Selected			
<input checked="" type="checkbox"/>	ABACOX		Deposition Number(s): 850656 Space Group: P 2 ₁ /n (14) Cell: a 12.373(3)Å b 7.3011(10)Å c 17.506(4)Å α 90° β 106.68(3)° γ 90° Compound Name: 3,4-Dimethylaminium 4-methylbenzenesulfonate
<input checked="" type="checkbox"/>	ABAFEQ		Deposition Number(s): 850654 Space Group: P 2 ₁ /n (14) Cell: a 5.5928(3)Å b 17.5004(7)Å c 20.1431(7)Å α 90° β 90° γ 90° Compound Name: (5)-(4-chlorobenzylsulfanyl)-1,3,4-oxadiazol-2-yl(ethyl)-4-methylbenzenesulfonamide
<input checked="" type="checkbox"/>	ABAFIT		Deposition Number(s): 223170 Space Group: P 2 ₁ /c (14) Cell: a 12.4536(9)Å b 16.1001(11)Å c 24.7739(17)Å α 90° β 102.7980(10)° γ 90° Compound Name: (μ2-Carbid)-[μ2-η5-1,3-dimesitylimidazolidin-2-ylidene)-trichloro-hydrido-(1,3-dimesitylimidazolidin-2-ylidene)-di-ruthenium benzene solvate
<input checked="" type="checkbox"/>	ABAMOG		Deposition Number(s): 231689 Space Group: P 1 (2) Cell: a 13.8106(6)Å b 15.2893(10)Å c 19.2600(13)Å α 107.863(2)° β 97.797(4)° γ 103.977(4)° Compound Name: bis(μ2-3-((bis(4-methylphenyl)imino)phenyl)phosphino)ethane)nickel(II) dinitrate hexahydrate

Figure 5. Search Results

Individual record - Individual records retrieved consist of several sections: reference code, compound name, 3-dimensional structure, chemical diagram, additional details, data citation, associated publication(s), and other chemical, crystal, and experimental details.

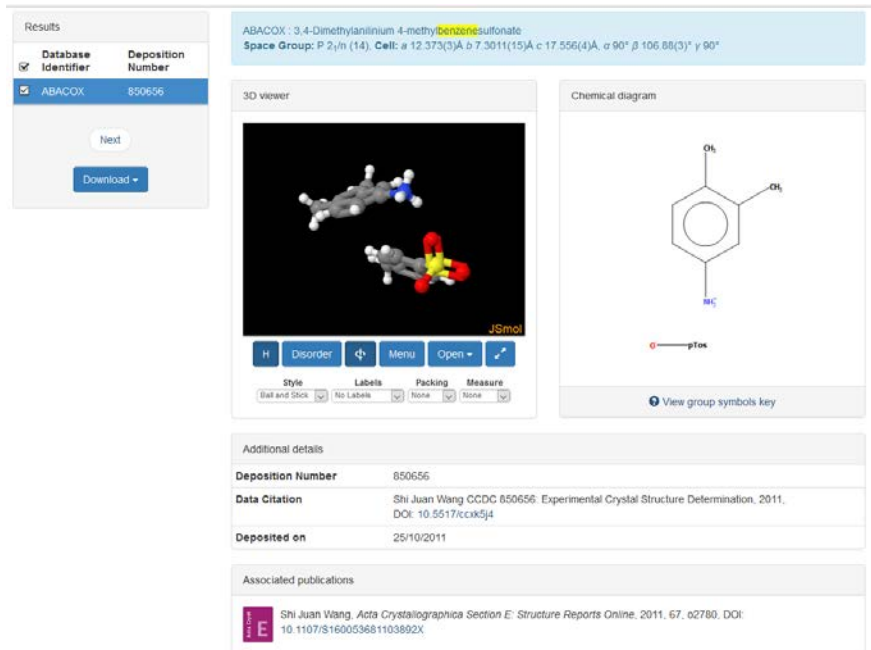


Figure 6. Individual record

Reference code (refcode) is a six-character unique identifier that is assigned to all entries in the database. Two additional characters may be assigned to indicate a record for an item that has already been deposited, but under different experimental conditions or by a different research group.

Compound name provides the name for the chemical compound, as well as space-group and unit-cell information.

3D structures are shown within a Java applet and can be viewed in full-screen mode and manipulated in several ways. These manipulations include, but are not limited to, changing the representation style, labels, colors and highlighting (to emphasize specific elements and bond types), zooming in on atoms and bonds, viewing in from any angle and in full screen, and showing measures like bond length, angle, and torsion.

Chemical diagram simply presents a standard skeletal structural formula for the compound.

Additional details include the CCDC deposition number, data citation (a reference to the CSD entry, including self-linking DOI), and the date on which the structure was deposited.

Associated publication(s) contains citation(s) for the journal publication associated with the structural determination. Hyperlinked DOIs are included where available. Note that these articles are not part of CSD and require additional subscriptions. These articles come predominantly from traditional chemistry resources that most institutions already subscribe to, such as Wiley, Taylor & Francis, and ScienceDirect.

Crystal details contains more detailed crystallographic information including space-group and unit-cell information (lengths of a, b, c and angles of α , β , γ). Further details, such as crystallization, cell volume, crystal habit, and polymorph information, are also provided, if available.

Experimental details provides experimental conditions, including R-factor, temperature, density, radiation probe, experiment type, and sensitivity, as provided by author.

Technical Information & Limitations

A CSD subscription, annually renewed, includes an unlimited-use license, with user authentication based on IP address, as well as local installation and updates for the CSD System Software. This authentication allows for secure searching locally using the on-site server, as well as integration with in-house databases and proxy connection for off-campus users. For records where the originally published articles are provided, DOIs can be linked to library holdings. These records are linked under the “Associated publications” section and have been shown to be a source of confusion for some first-time users. Some users infer from the verbiage that these records are other related publications outside of the original structural determination.

While WebCSD can be accessed from any device, the structure search and 3D structural applets can infrequently be finicky. Most problems that users encounter with WebCSD can be resolved through updating Java, switching browsers (Firefox seems to work best), or clearing cache & cookies. There are occasional off-campus access interruptions, especially during heavy-use periods, for instance, when class assignments that require WebCSD are due. CCDC recommends having users create their own (free) CCDC accounts using the License Site Number and License Confirmation Code, which can be requested from the subject librarian. While the CCDC support team is very fast to respond when issues arise, their response time is sometimes hindered for US-based institutions by the time-zone difference.

There are other resources available that offer comparable features, such as structure and compound searching, 3D visualization, and structural properties, but CCDC provides the most comprehensive coverage for organic crystal structures, while also incorporating innovative searching techniques and thorough experimental and physical details. It is important to note that while CSD does contain all published crystal structures, as of 2015, it was estimated that only about 15% of determined structures were published ([Groom et al. 2016](#)). Additionally CSD does not include the following: inorganic structures, proteins, high-molecular-weight compounds, polypeptides and polysaccharides consisting of greater than 24 units, or oligonucleotides.

Patrons searching for those structures should consider: Inorganic Crystal Structure Database (ICSD), NRCC Metals Crystallographic Data File (CRYSTMET), Protein Databank (PDB), or ICDD NIST Crystal Data File. While many databases offer some features of CSD, such as the chemical structure search, no other available resources offer the full search capabilities or comprehensive records afforded by CSD.

For more information contact:

The Cambridge Crystallographic Data Centre

<https://www.ccdc.cam.ac.uk/solutions/csd-system/components/csd/>

12 Union Road, Cambridge, CB2 1EZ, United Kingdom.
Phone: +44 (0)1223 336408, Fax: +44 (0)1223 336033.

References

Battle, G., & Allen, F. 2012. Learning about Intermolecular Interactions from the Cambridge Structural Database. *Journal of Chemical Education* 89(1): 38. DOI: [10.1021/ed200139t](https://doi.org/10.1021/ed200139t).

Battle, G., Atlen, F., & Ferrence, G. 2011. Teaching Three-Dimensional Structural Chemistry Using Crystal Structure Databases. 4. Examples of Discovery-Based Learning Using the Complete Cambridge Structural Database. *Journal of Chemical Education* 88(7): 891. DOI: [10.1021/ed1011025](https://doi.org/10.1021/ed1011025).

Battle, G. M., Ferrence, G. M., & Allen, F. H. 2010. Applications of the Cambridge Structural Database in chemical education. *Journal of Applied Crystallography* 43(5-2): 1208-1223. DOI: [10.1107/S0021889810024155](https://doi.org/10.1107/S0021889810024155).

Groom, C. R., & Allen, F. H. 2014. The Cambridge Structural Database in Retrospect and Prospect. *Angewandte Chemie International Edition* 53(3): 662-671. DOI: [10.1002/anie.201306438](https://doi.org/10.1002/anie.201306438).

Groom, C. R., Bruno, I. J., Lightfoot, M. P., & Ward, S. C. 2016. The Cambridge Structural Database. *Acta Crystallographica Section B* 72(2): 171-179. DOI: [10.1107/S2052520616003954](https://doi.org/10.1107/S2052520616003954).

Thomas, I. R., Bruno, I. J., Cole, J. C., Macrae, C. F., Pidcock, E., & Wood, P. A. 2010. WebCSD : the online portal to the Cambridge Structural Database. *Journal of Applied Crystallography* 43(2): 362-366. DOI: [10.1107/S0021889810000452](https://doi.org/10.1107/S0021889810000452).