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

Cambridge Structural Database (WebCSD)

The Cambridge Crystallographic Data Centre
12 Union Road
Cambridge
CB2 1EZ
United Kingdom
Phone: +44 (0)1223 336408
Fax: +44 (0)1223 336033

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, Bruno, Lightfoot, & Ward, 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.

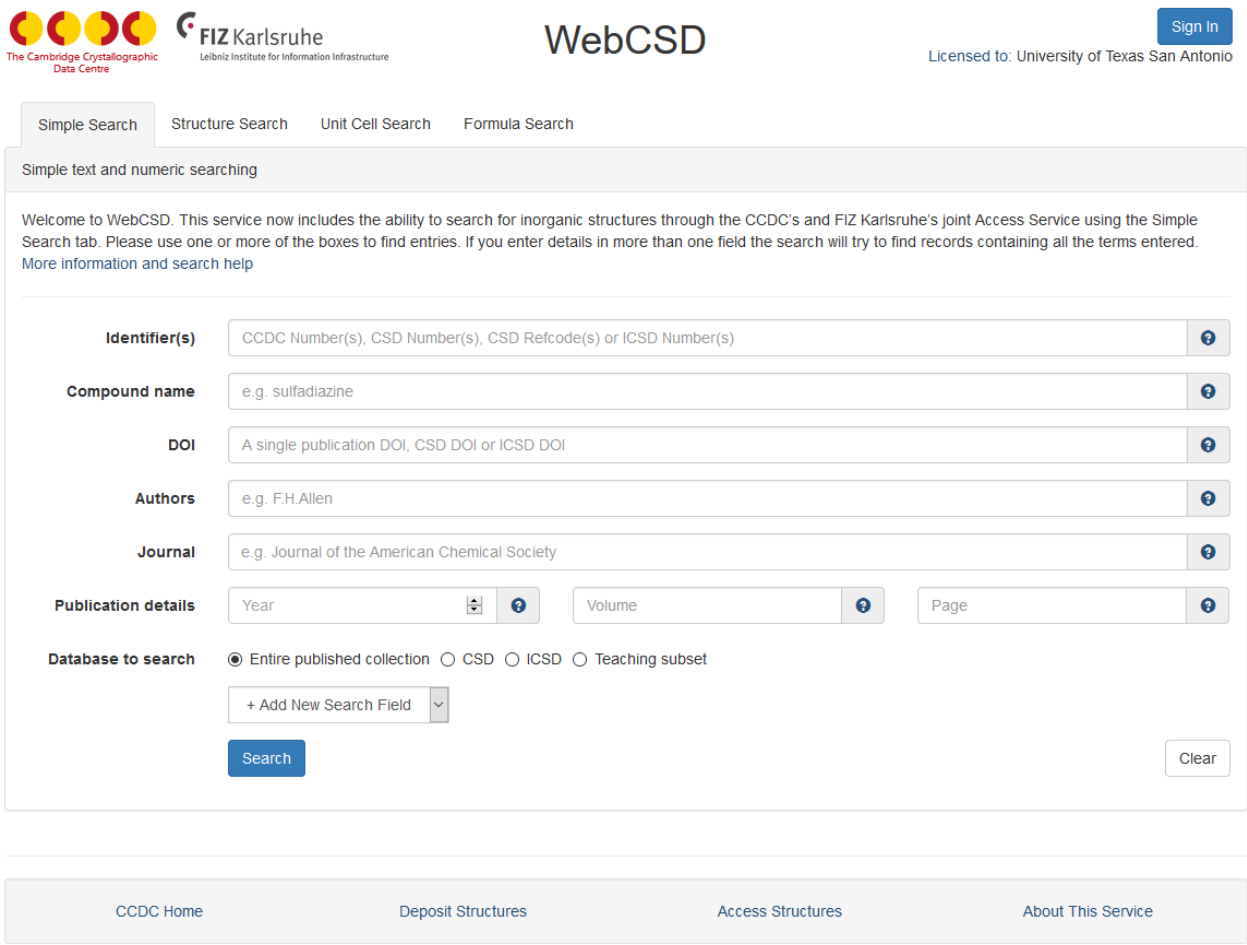
The CSD Software System is intended for in-depth and comprehensive crystallographic searching by advanced users with expert knowledge in crystallography, such as crystallographers, structural chemists, and the drug design community (Thomas et al., 2010). Therefore, this review will focus on WebCSD as the web implementation, aimed at the medicinal and pharmaceutical chemists, 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 (G. Battle & Allen, 2012; G. Battle, Atlen, & Ferrence, 2011; G. M. Battle, Ferrence, & Allen, 2010) and falls outside the scope of this publication.

Searching CSD

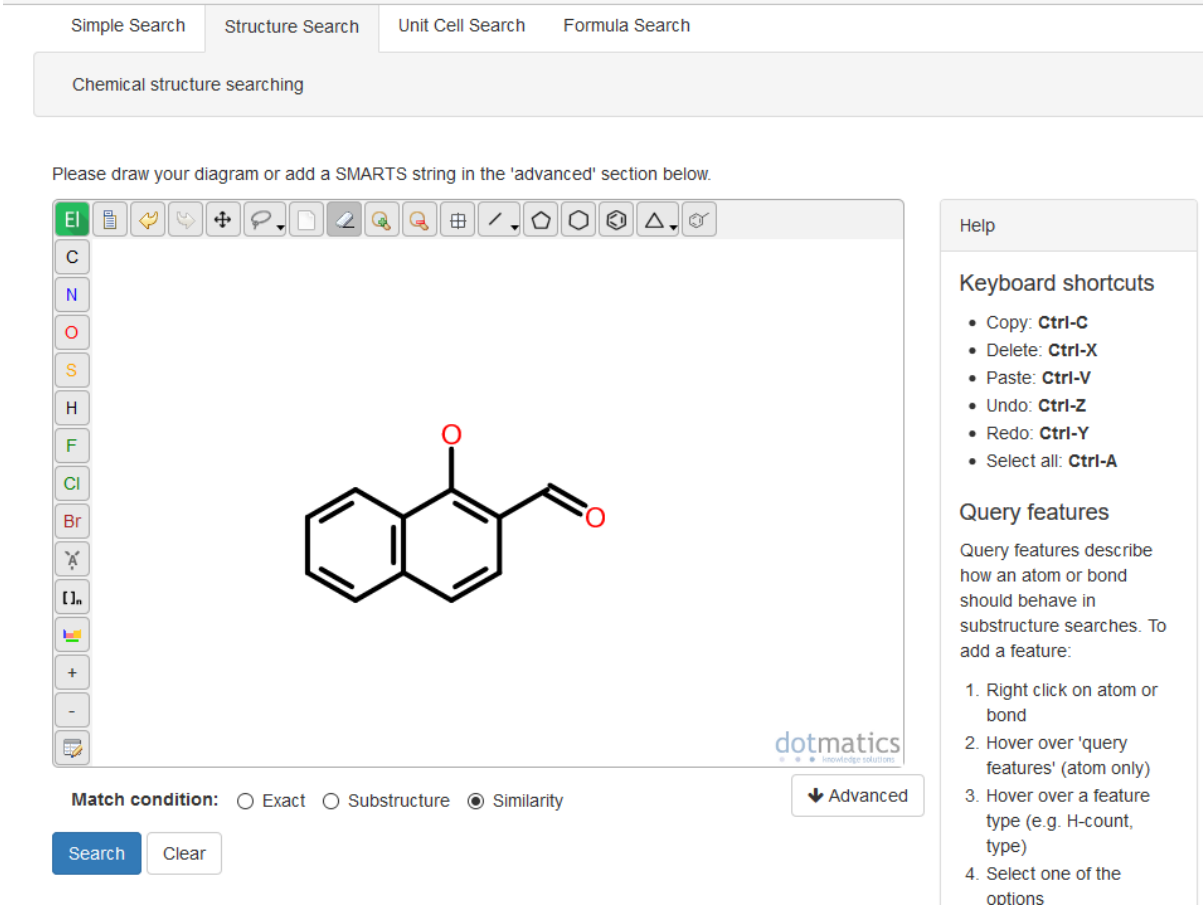
CSD offers several options for searching, arranged in a tabbed frame. Each tab represents a different search type with various options for search and refinement within. The search types are Simple, Structure, Unit Cell, and Formula.

Simple Search –default landing page, search by:

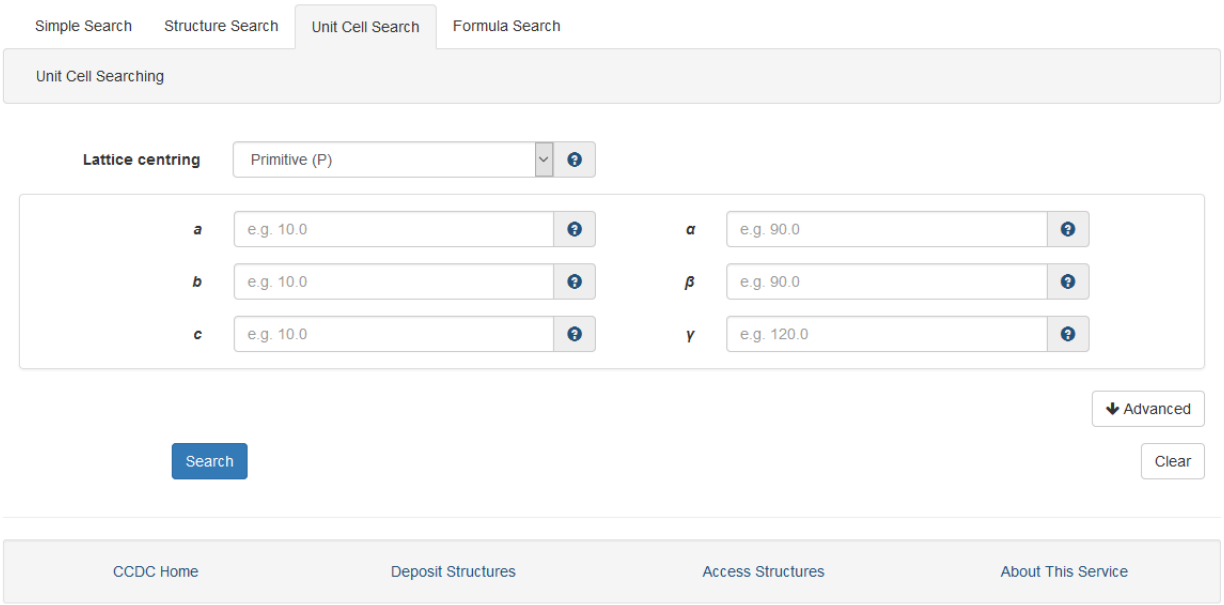
- Identifiers – CCDC/CSD number(s), CSD Refcode(s), or ICS Number(s)
- Compound name – for instance Sodium Chloride
- DOI – publication or CSD DOI
- Authors – Publication author(s)
- Journal – Publication journal title
- Publication Details – Year, volume, page



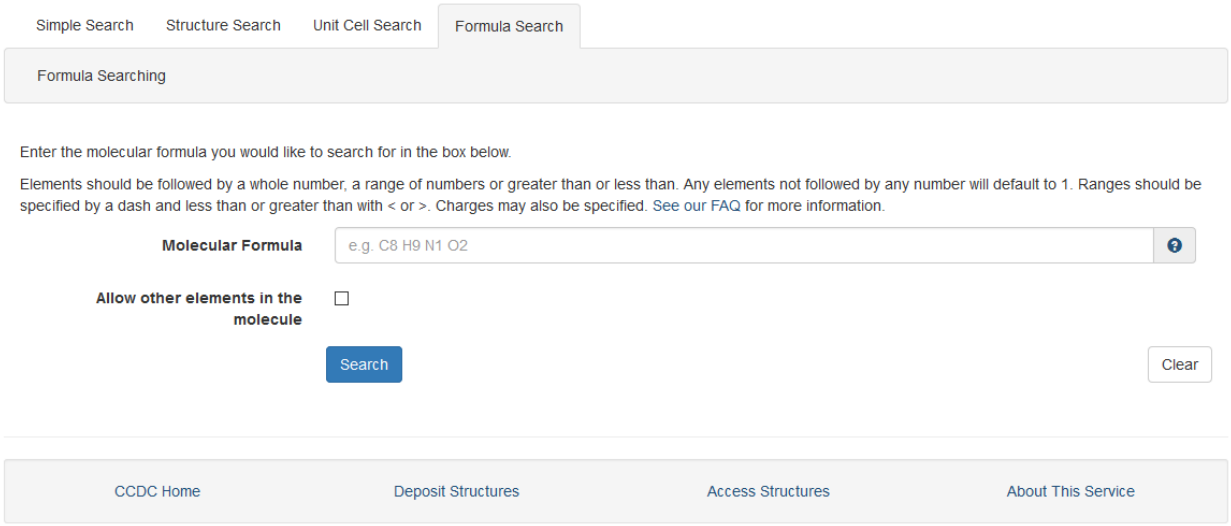
Structure Search – users can draw chemical structures within a Java script applet to search for exact structure, substructure, or similarity. Common rings and elements can be selected from the toolbar, a periodic table, and hand-drawing tools are available for others.



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

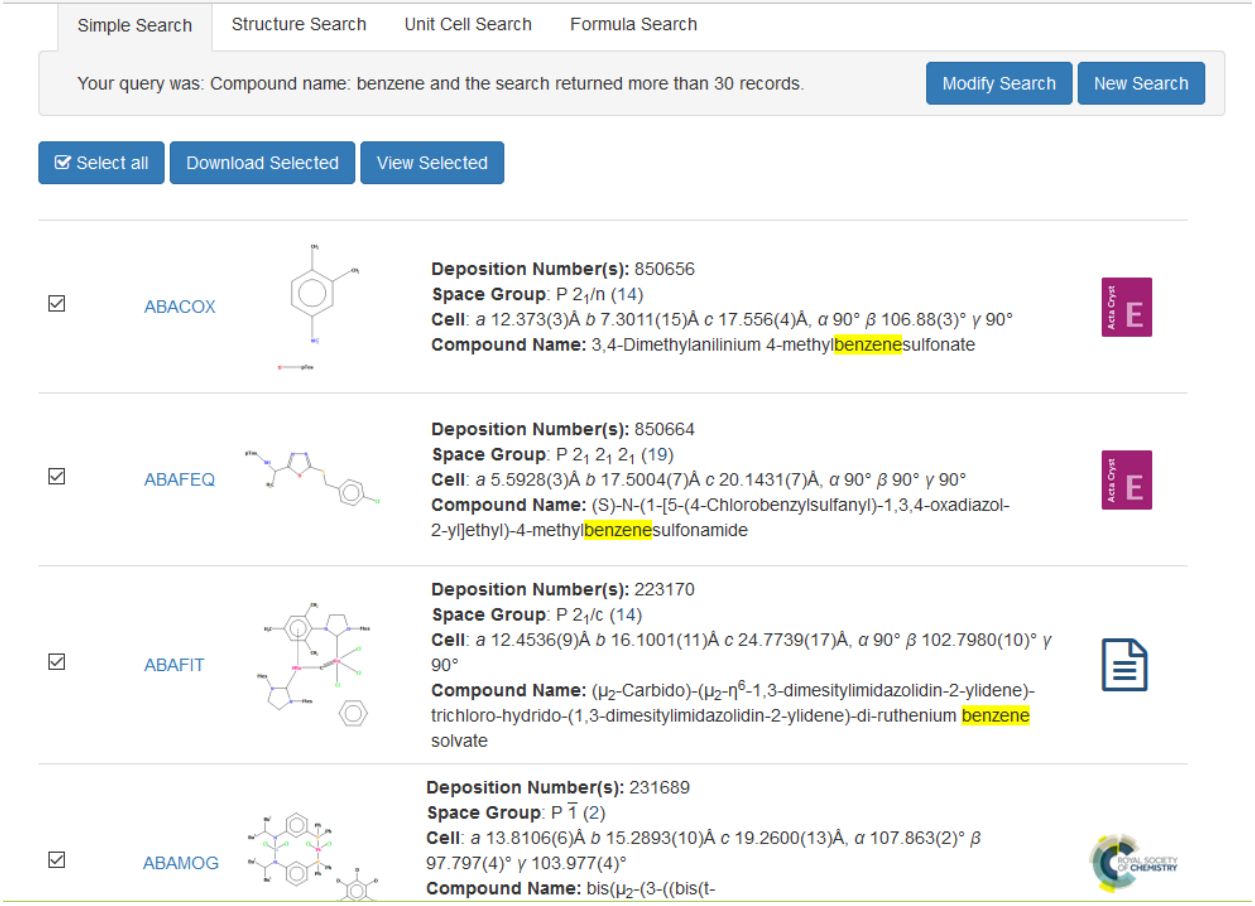


Formula Search – search by molecular formula components (i.e. C5 H6 O2)



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.



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.

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 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.

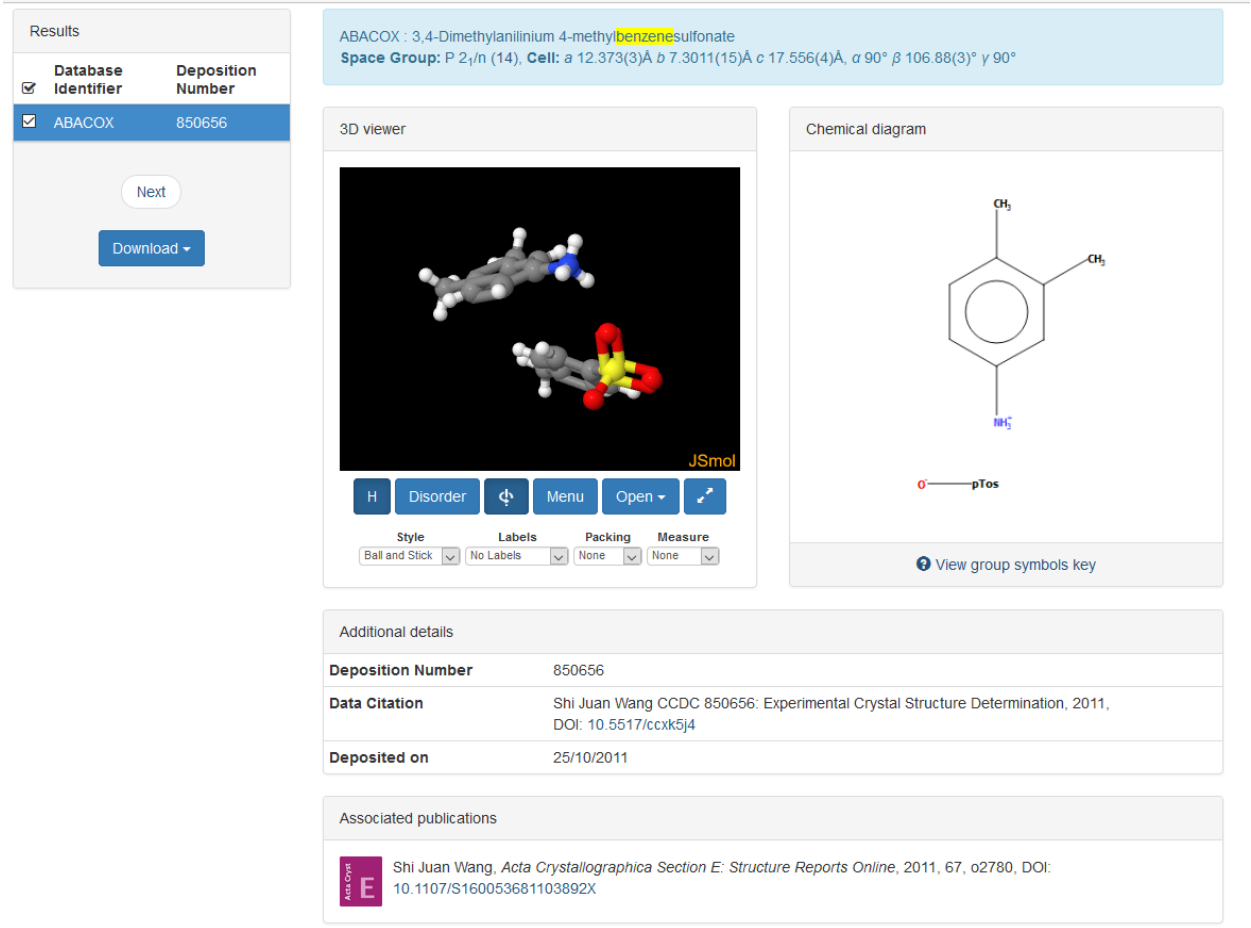
Additional details include the CCDC deposition number, data citation (a reference to the CSD entry, including self-linking DOI), and date on which the structure was deposited.

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, 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.

Crystal details contains more detailed crystallographic information including space group and unit cell information. 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

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 as well as 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.

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

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

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

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

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

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