Searching the CSD using ConQuest

CCDC Virtual Workshop 2020 – Session 1

Ilaria Gimondi, Suzanna Ward, Yinka Olatunji-Ojo, Natalie Johnson, Jeff Lengyel

November 2020
Learning outcomes for today

• How to search over a million published crystal structures using our desktop software
• How to refine your results to structures in targeted subsets allowing you to explore drugs, pesticides and MOFs in more detail
CSD Refcodes

What is XOPCAJ?
- A CSD Refcode
- A database reference code
- Containing 6-8 characters
- Used to identify entries in the CSD

Refcode families
- The same substances are assigned the same 6 letter code plus an additional 2 numbers
  - Polymorphs
  - New determinations or re-refinements of the same substance
  - Determinations at different temperatures/pressures
- Stereoisomers or different solvates, co-crystals, etc are assigned different refcode families
The CSD software
CSD-System – What does it enable?

- It makes it easy to answer very specific structure questions on your desktop using highly flexible 3D searching of the CSD.
- Quickly identify the most relevant structures across the CSD based on a wide range of properties including chemical constraints such as cyclicity.
- Use interaction and structure-property knowledge gained to drive design decisions.
CSD-System: Search

On your desktop: ConQuest

On a browser: WebCSD
What is ConQuest?

• Enables search and retrieval of information from the CSD
• Provides full range of text / numeric database search options
• More complex search functionality includes:
  • Chemical structure searching
  • 3D Geometrical searching
  • Intermolecular non-bonded contact searching
ConQuest

- Large range of filtering options based on experimental or chemical considerations (e.g. R-factor, temperature, only organics)
- Combine searches using various Boolean options and manage hitlists post-search
- Export results to Mercury for visualisation and advanced numerical analysis & plotting
ConQuest features

- Search for information relating to the structure determination
- Structure sketcher, enabling set-up of basic substructure searches and complex 3D queries
- Bibliographic search
- Generic text search
- Analysis of results
- Elemental make-up
Show One: ConQuest – Opening and search options
ConQuest – Searching and filtering

[Image of ConQuest software interface showing search setup and filters]

- Search Name: search1
- Available Databases: CSD version 5.41 (November 2019) + 3 updates
- Show Updates separately
- Filters section includes options for 3D coordinates, R factor, order/disorder, errors, polymeric, and ions.

You can search complete database(s) or a subset (e.g., hits found in a previous search).

Single query being used. Search will find structures where this query is true:
ConQuest – Author and journal search

The image shows a software interface for author and journal search, with options to build queries, combine search criteria, and refine the search with specific parameters like author names, journal names, and other details. The interface includes sections for drawing, peptide search, author/journal, name/class, elements, formula, space group, unit cell, Z/density, experimental, all text, and refcode (entry ID). The search setup allows for specifying databases, search names, available databases, search parameters, and advanced options such as 3D coordinates determined, R factor, and other criteria to refine the search.
Left clicking each tab will display different information.

Query highlighted

A hit list will be returned within the View Results tab of ConQuest. Left clicking on each REFCODE will toggle on and off that hit for future manipulation.
A hit list will be returned within the “results” tab of ConQuest. Left clicking on each REFCODE will toggle on and off that hit for future manipulation.

Left clicking each tab will display different information

Right clicking on 3D visualiser enables you to change the display style
You can export results in a range of different formats to save data for future manipulation.
ConQuest – Compound name search

Name/Class (1) - New

**Compound Name**
- Ignore non-alphabetic characters, e.g. “butadiene” will match “buta-1,3-diene”
- Find exact word, e.g. “hydrazine” will not match “acetylhydrazine”

**Chemical Class**
- ---- not defined ----
- Carbohydrates
- Nucleosides & nucleotides
- Amino-acids, peptides & complexes
- Porphyrins, corrins & complexes
- Steroids
- Terpenes
- Alkaloids
- Organic polymers

Find entries classified as: ---- not defined ---- and: ---- not defined ----
ConQuest – Space Group and cell search
ConQuest – Elements and formula search
ConQuest – Z/Density and experimental info search
ConQuest – Draw/Structure search

In the demo we will search for:

\[
\text{O} - C - C - 7A - O
\]

where 7A represents any halogen.
What else can you explore in ConQuest?

3D Search
Dihedral angle between planes: 20° - 80°

Export & Analyse Data

Intermolecular Interactions

Search In-House Databases

... and more!
Data analysis in Mercury
Want to explore more?

Educational Resources

The wealth of information contained within the Cambridge Structural Database (CSD) extends far beyond a collection of crystal structures. Knowledge derived from these materials informs much of chemistry, biochemistry, and biology. Chemical and structural concepts are often difficult to grasp without real-world, interactive examples for students to explore.

The CCDC and our colleagues continually produce educational materials for use in classroom and computer lab settings, or as independent study modules. Many of these materials make use of the Teaching Subset - a freely available set of over 750 structures that can be investigated with the free version of our Mercury visualisation and analysis program. Of course, our database of over one million entries are available for free through our Access Structures portal.

If you are an educator looking for supplementary teaching materials, find out more about the Teaching Database here. If you have developed your own modules using the CSD and would like to share them with the broader community, please contact us at education@ccdc.cam.ac.uk.

Self-guided workshops about ConQuest in https://www.ccdc.cam.ac.uk/Community/educationalresources/workshop-materials/csd-system-workshops/

YouTube and LabTube channels links from https://www.ccdc.cam.ac.uk/Community/educationalresources/ccdc-videos/