Open-access Crystallographic Databases (for the teaching of geometric structural crystallography and much more)

Peter Moeck, Physics, Portland State University, & all other Members of International Advisory Board of the Crystallography Open Database (COD): Daniel Chateigner, Xiaolong Chen, Marco Ciriotti, Robert T. Downs, Saulius Gražulis, Armel Le Bail, Luca Lutterotti, Yoshitaka Matsushita, Miguel Quirós Olozábal, Hareesh Rajan, Alexandre F.T. Yokochi

Financial support from the Northwest Academic Computing Consortium and Portland State University (to PM) is gratefully acknowledged. The Vilnius development group of the Crystallography Open Database (under the direction of SG) is supported by the Research Council of Lithuania, contract No. MIP-124/2010.
Outline

- What kind of crystallographic data are there in open access?
- What might be done with them for teaching purposes?
- What else might be done with some of these data?
- Anyone interested in uploads of data to the COD in the Crystallographic Information File (CIF) format?
- Anyone interested in downloads of CIFs from the COD or American Mineralogist Crystal Structure Database for powder XRD, EBSD, or other purposes?
Crystallographic database

From Wikipedia, the free encyclopedia
(Redirected from Crystallographic database)

A crystallographic database is a database specifically designed to store information about crystals and crystal structures. Crystals are solids having, in all three dimensions of space, a regularly repeating arrangement of atoms, ions, or molecules. They are characterized by symmetry, morphology, and directionally dependent physical properties. A crystal structure describes the arrangement of atoms, ions, or molecules in a crystal.

Crystal structures of crystalline material are typically determined from X-ray or neutron single-crystal diffraction data and stored in crystal structure databases. They are routinely identified by comparing reflection intensities and lattice spacings from X-ray powder diffraction data with entries in powder-diffraction fingerprint databases.

Crystal structures of nanometer sized crystalline samples can be determined via structure factor amplitude information from single-crystal electron diffraction data or structure factor amplitude and phase angle information from Fourier transforms of HRTEM images of crystals. They are stored in crystal structure databases specializing in nanocrystals and can be identified by comparing zone axis subsets in lattice-fringe fingerprint plots with entries in a lattice-fringe fingerprinting database.

Crystallographic databases can be categorized as crystallographic information from supersets or sub-sets of inorganics, metals/biaryls, organics, and biological macromolecules. They differ in access and usage rights and offer varying degrees of search and analysis capacity. Many provide structure visualization capabilities. They can be browser based or installed locally. Newer versions are built on the relational database model and support the Crystallographic Information File (CIF) as a universal data exchange format.

Contents (show)[edit]

1 Overview
2 Trends
3 Search
4 Crystal phase identification
   4.1 Powder diffraction fingerprinting (TCP)
   4.2 Lattice-fringe fingerprinting (G2)
   4.3 Morphological fingerprinting (3D)
   4.4 Lattice matching (CE)
5 Visualization
   5.1 Crystal structures
   5.2 Morphology and physical properties
6 References
7 See also
8 External links
   8.1 Crystal structures

116,000 entries

External links

- American Mineralogical Crystal Structure Database (AMCD) (contains crystal structures of minerals, access free, size medium)
- Cambridge Structural Database (CSD) (contains crystal structures of organics and inorganics, access medium, size small)
- Cambridge Crystallographic Data Centre (CCDC) (contains crystal structures of organics, inorganics, heavy elements, and small molecules, access free, size small)
- Crystallography Open Database (COD) (contains crystal structures of organics, inorganics, heavy elements, and small molecules, access free, size small)
- Database of Inorganic Structures (contains crystal structures of inorganics, access free, size small)
- Crystallography User Portal (CUP) (contains crystal structures of organics, inorganics, heavy elements, and small molecules, access free, size small)
- Inorganic Crystal Structure Database (ICSD) (contains crystal structures of inorganics, access free, size medium)
- Materials Structure Database (CRYSTEM) (contains crystal structures of materials, access free, size small-medium)
- Molecular Crystal Structures (contains crystal structures of materials, access free, size small-medium)
- NIST Structure Database (contains crystal structures of materials, access free, size small-medium)
- NIST Diffraction Structures Database (contains crystal structures of materials, access free, size small-medium)
- PDB-1 crystal structures database (contains crystal structures of proteins, access free, size small-medium)
- Protein Data Bank (PDB) (contains crystal structures of organic and inorganic macromolecules, access free, size medium)
- Cambridge Crystallographic Data Centre (CCDC) (contains crystal structures of organics, inorganics, heavy elements, and small molecules, access medium, size small-medium)
- International Centre for Diffraction Data (ICDD) (contains crystal structures of materials, access free, size medium)
- Materials Structure Database (CRYSTEM) (contains crystal structures of materials, access free, size small-medium)
- Powder Diffraction File (PDF) (contains crystal structures of minerals and inorganics, access free, size small-medium)
- Powder Diffraction File (PDF) (contains crystal structures of minerals and inorganics, access free, size small-medium)
- Rietveld Method (contains crystal structures of materials, access free, size small-medium)
- Special Features of Crystal Structures (contains crystal structures of materials, access free, size small-medium)
- X-ray Crystallography Database (XCD) (contains crystal structures of organics, inorganics, heavy elements, and small molecules, access medium, size small-medium)
- X-ray Crystallography Database (XCD) (contains crystal structures of organics, inorganics, heavy elements, and small molecules, access medium, size small-medium)
- X-ray Crystallography Database (XCD) (contains crystal structures of organics, inorganics, heavy elements, and small molecules, access medium, size small-medium)
- X-ray Crystallography Database (XCD) (contains crystal structures of organics, inorganics, heavy elements, and small molecules, access medium, size small-medium)
- X-ray Crystallography Database (XCD) (contains crystal structures of organics, inorganics, heavy elements, and small molecules, access medium, size small-medium)
CIF

The International Union of Crystallography is the sponsor of the Crystallographic Information Framework, a standard for information interchange in crystallography.

The acronym CIF is used both for the Crystallographic Information File, the data exchange standard file format of Hall, Allen & Brown (1991) (see Documentation), and for the Crystallographic Information Framework, a broader system of exchange protocols based on data dictionaries and relational rules expressible in different machine-readable manifestations, including but not restricted to Crystallographic Information File and XML.

CIF was developed by the IUCr Working Party on Crystallographic Information in an effort sponsored by the IUCr Commission on Crystallographic Data and the IUCr Commission on Journals, and was adopted in 1990 as a standard file structure for the archiving and distribution of crystallographic information. It is now well established and is in regular use for reporting crystal structure determinations to Acta Crystallographica and other journals. It is often cited as a model example of integrating data and textual information for data-centric scientific communication. In 2006 the importance of CIF and the value of its accompanying web-based service for the validation of structural data, checkCIF, were recognized by the Award for Publishing Innovation of the Association of Learned and Professional Society Publishers (ALPSP). In their report, the judges were impressed with the way in which CIF and checkCIF are easily accessible and have served to make critical crystallographic data more consistently reliable and accessible at all stages of the information chain, from authors, reviewers and editors through to readers and researchers. In doing so, the system takes away the decisions from ensuring that the results of scientific research are trustworthy without detracting from the value of human judgement in the research and publication process.

This part of the IUCr web site provides comprehensive documentation and software resources for users and developers of CIF software.
Neutron and x-ray structure refinements between 15 and 1083 K of piezoelectric gallium arsenate, GaAsO₄: temperature and pressure behavior compared with other $\alpha$-quartz materials.

<table>
<thead>
<tr>
<th>Atom Type Symbol</th>
<th>Oxidation Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ga³⁺</td>
<td>3.000</td>
</tr>
<tr>
<td>As⁵⁺</td>
<td>5.000</td>
</tr>
<tr>
<td>O²⁻</td>
<td>-2.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Atom Site Label</th>
<th>Atom Site Type Symbol</th>
<th>Symmetry Equiv Pos As Xyz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ga₁</td>
<td>Ga³⁺</td>
<td>x,y,z</td>
</tr>
<tr>
<td>As₁</td>
<td>As⁵⁺</td>
<td>-y,x-y,1/3+z</td>
</tr>
<tr>
<td>O₁</td>
<td>O²⁻</td>
<td>y-x,-x,2/3+z</td>
</tr>
<tr>
<td>O₂</td>
<td>O²⁻</td>
<td>y,x,-z</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-x,y,x,1/3-z</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-y,y,y,2/3-z</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Symmetry Space Group Name H-M</th>
<th>P 31 2 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symmetry Int Tables Number</td>
<td>152</td>
</tr>
<tr>
<td>Symmetry Cell Setting</td>
<td>Trigonal</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cell Length A</th>
<th>4.9940(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Length B</td>
<td>4.9940(1)</td>
</tr>
<tr>
<td>Cell Length C</td>
<td>11.3871(4)</td>
</tr>
<tr>
<td>Cell Angle Alpha</td>
<td>90</td>
</tr>
<tr>
<td>Cell Angle Beta</td>
<td>90</td>
</tr>
<tr>
<td>Cell Angle Gamma</td>
<td>120</td>
</tr>
<tr>
<td>Cell Volume</td>
<td>245.9</td>
</tr>
<tr>
<td>Cell Formula Units Z</td>
<td>3</td>
</tr>
<tr>
<td>Symmetry Space Group Name H-M</td>
<td>P 31 2 1</td>
</tr>
<tr>
<td>Symmetry Int Tables Number</td>
<td>152</td>
</tr>
<tr>
<td>Symmetry Cell Setting</td>
<td>Trigonal</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Atom Site Aniso Label</th>
<th>Atom Site Aniso U₁₁</th>
<th>Atom Site Aniso U₁₂</th>
<th>Atom Site Aniso U₁₃</th>
<th>Atom Site Aniso U₂₂</th>
<th>Atom Site Aniso U₂₃</th>
<th>Atom Site Aniso U₃₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ga₁</td>
<td>0.00202(1)</td>
<td>0.0</td>
<td>0.00077(16)</td>
<td>-0.00093(11)</td>
<td>0.00163(15)</td>
<td>0.00201(1)</td>
</tr>
<tr>
<td>As₁</td>
<td>0.000120(12)</td>
<td>0.0</td>
<td>0.000177(16)</td>
<td>-0.00009(11)</td>
<td>0.00163(15)</td>
<td>0.00201(1)</td>
</tr>
<tr>
<td>O₁</td>
<td>0.000552(14)</td>
<td>0.000311(11)</td>
<td>-0.00111(9)</td>
<td>0.000395(13)</td>
<td>-0.000138(10)</td>
<td>0.000432(12)</td>
</tr>
<tr>
<td>O₂</td>
<td>0.00074(15)</td>
<td>0.000327(11)</td>
<td>-0.00089(10)</td>
<td>0.000372(14)</td>
<td>-0.00109(11)</td>
<td>0.000446(13)</td>
</tr>
<tr>
<td>Refine ls R Factor All</td>
<td>0.025</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
130 crystal structures and individual molecules for Materials & Technology, CIF downloadable

about 250 more structures over alphabetical list
Graphite is a polymorph of the element carbon.

Chemical Formula: C

Lawson's explanation: In the mid-16th century, a violent storm knocked over several trees in Borrowdale, England, uncovering a large deposit of a black substance that was first thought to be lead. More than 200 years later, an English scientist discovered that the substance was not actually lead, but a type of carbon instead. The substance was named graphite after the Greek word meaning "to write," since that is how people used the substance.
500 CIFs of important organics can be displayed and downloaded

http://crystdb.nims.go.jp/index_en.htm
National Institute for Materials Science, Tsukuba, Japan

former? Linus Pauling File, 82,000 crystal structures, 55,000 material properties, 15,000 phase diagrams, not clear how many CIFs downloadable or inter-actively displayable in 3D
probably all useful crystallographic data for teaching already in open access

http://cod.ibt.lt

mirrors worldwide
www.crystallography.net
COD.ENSICAEN.FR
www.nanocrystallography.org
jCOD.NANOCRystallography.NET
COD.NANOCRystallography.NET/COD
WEB PORTAL
http://www.nanocrystallography.net
http://nanocrystallography.research.pdx.edu/CIF-searchable

Interactive Crystallography Databases

COD Subset
COD Subset
Educational subset of COD
Nano-Crystallography Database
Wiki Crystallography Database
LFTP Matching (coming soon)

Tools
Login
Main Sponsor
Links

more than 170,000 hits since January 1st, 2008
## COD Subset Search

### Search by these properties:
- **Text (1 or 2 words)**: 
- **Include these elements**: 
- **Without these elements**: 
- **Strict number of elements**: 

### Interactive Table:

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>Si</td>
</tr>
<tr>
<td>Na</td>
<td>Mg</td>
</tr>
<tr>
<td>K</td>
<td>Ca</td>
</tr>
<tr>
<td>Al</td>
<td>Fe</td>
</tr>
<tr>
<td>Si</td>
<td>Mg</td>
</tr>
<tr>
<td>Ca</td>
<td>Fe</td>
</tr>
</tbody>
</table>

- **Cell volume (min, max)**: 
- **a (min, max)**: 

---

*NanoMegas*

*Advanced Tools for electron diffraction*

*Page maintained by [webadmin] Portland State University 2010*

*Visits since January 19, 2008: 17*
Formula: Fe₂₆₉O₇₄

Space group: F d 3 m
Cell parameters: a = 8.395 Å, b = 8.395 Å, c = 8.395 Å

User comments: None
all 116,000+ entries of COD can be displayed at http://jcod.nanocrystallography.net
Also this interface for COD and capability to display structures interactively in 3D over the portal nanocrystallography.net
www.crystallography.net

Crystallography Open Database

Deposit your data (NEW!)
or
Upload data
or
Search the database

View the Petition for Open Data in Crystallography
Call to Volunteers
See also the COD, Predicted Crystallography Open Database
More on the COD project: what's new

Recent open access paper regarding COD development was published in Journal of Applied Crystallography. PDF and HTML versions are available.
COD Advisory Board thanks Crystal Impact GmbH for their financial support of this paper publication.
Crystallography Open Database
Validation and Deposition Interface

About this Validation Interface
This interface allows you to upload, validate and edit CIF files before submitting them for deposition.

Steps
The process of file deposition, after you have uploaded your data is pretty simple:
First step, after files have been uploaded, it validation. Our script performs some checks to see if all necessary data are present in the submitted file. Results are displayed to you next to your file:
If a file is correct, you can deposit it to COD. After the deposition, COD numbers for the newly deposited structures will be displayed.
If a file is not correct you can edit it file in your browser window and validate it once more.

File formats
Currently we accept two types of files:
- Plain CIF files,
- ZIP archives, which does contain CIF files.
checkCIF/PLATON full publication check

Welcome to the checkCIF/PLATON service, which is operated by the IUCr. You may use this form to upload files for a full publication check.
Please upload your CIF using the form below.

File name

This full publication check (required for the journals Acta Crystallographica Section C and Acta Crystallographica Section A) includes checks on:

- CIF syntax and constitution
- Cell and geometry details
- Space-group symmetry
- Anisotropic displacement parameters
- Publication items

Structure-factor checking is currently being tested on articles submitted to Acta Crystallographica Sections C and Acta Crystallographica Section A. These tests may be carried out with a local version of PLATON in the test service here.

If you intend to submit to another section of Acta Crystallographica, Journal of Applied Crystallography or Journal of Synchrotron Radiation, you should make sure that at least basic structural checks are run on the final version of your CIF prior to submission.

If you wish to carry out a full structural check (but without publication checks) please click here.

IMPORTANT NOTICE: Acta Crystallographica Section B: Structure Reports Online is now an open-access journal. For more information, please visit http://journals.iucr.org/services

Useful links:
- How to use checkCIF/PLATON reports
- Details of checkCIF/PLATON tests
- Submit to IUCr Journals
- Core CIF dictionary
Now COD is available for PDXL: Integrated Powder X-ray Diffraction Software!

If you own PDXL, COD can be used along with phase identification in PDXL after installing the index file of COD which Rigaku provides. Please install it according to the following procedure.

NOTE: For any publications, including academic papers, presentations at conferences, etc. in which you report the results of phase identification, Rietveld analysis, etc. obtained using COD in PDXL, please make reference to:

“Crystallography Open Database — an open-access collection of crystal structures”, 

Installation:

1. Download the installer PDXLCODSetup.exe (7MB)

2. Execute it

3. Click OK for the following message

4. Wizard dialog box will appear. Click Next.
**Match!**

**Phase Identification from Powder Diffraction**

Match! is an easy-to-use software for phase identification from powder diffraction data, which has become a daily task in materials scientists work.

Match! compares the powder diffraction pattern of your sample to a database containing reference patterns in order to identify the phases which are present. Single as well as multiple phases can be identified based on both peak data and raw (profile) data.

As reference database, you can apply the included **free-of-charge CSD database** and/or **ICSD/Retriever** (if you have a valid license). You can also **PDF product** and/or create a user database based on your own diffraction patterns. The user database patterns can be edited manually, imported from peak files, calculated from crystal structure data (e.g. CIF files), or imported from your colleague's user database. A list of Match!'s most prominent features can be found here.
<table>
<thead>
<tr>
<th>Name</th>
<th>Last modified</th>
<th>Size</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent Directory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AmMin_April2010_4HS.zip</td>
<td>27-Apr-2010 08:31</td>
<td>39M</td>
<td></td>
</tr>
<tr>
<td>AmMin_Jan2010_4HS.zip</td>
<td>04-Feb-2010 09:44</td>
<td>38M</td>
<td></td>
</tr>
<tr>
<td>AmMin_Mar2010_4HS.zip</td>
<td>26-Feb-2010 14:17</td>
<td>38M</td>
<td></td>
</tr>
<tr>
<td>COD_April2010_4HS.zip</td>
<td>27-Apr-2010 08:57</td>
<td>659M</td>
<td></td>
</tr>
<tr>
<td>COD_Jan2010_4HS.zip</td>
<td>04-Feb-2010 16:16</td>
<td>618M</td>
<td></td>
</tr>
<tr>
<td>COD_Mar2010_4HS.zip</td>
<td>26-Feb-2010 15:14</td>
<td>618M</td>
<td></td>
</tr>
<tr>
<td>PANalytical.txt</td>
<td>03-Jun-2010 14:02</td>
<td>1.3K</td>
<td></td>
</tr>
<tr>
<td>README.txt</td>
<td>03-Jun-2010 14:02</td>
<td>1.3K</td>
<td></td>
</tr>
<tr>
<td>md5sum.dat</td>
<td>03-Jun-2010 14:05</td>
<td>490</td>
<td></td>
</tr>
<tr>
<td>shalsum.dat</td>
<td>03-Jun-2010 14:05</td>
<td>562</td>
<td></td>
</tr>
<tr>
<td>wget-log</td>
<td>28-May-2010 08:58</td>
<td>62K</td>
<td></td>
</tr>
<tr>
<td>wget-log_1</td>
<td>28-May-2010 09:31</td>
<td>1.0M</td>
<td></td>
</tr>
</tbody>
</table>
Hands-on crystallography has its merits, but ...

some instructors & students either don’t get it or too much of it.