Open access crystallographic databases & COD in its 11th year

Peter Moeck¹, Daniel Chateigner², Robert T. Downs³, Saulius Gražulis⁴, Armel Le Bail⁵, Luca Lutterotti⁶, Miguel Quirós⁷, & all other members of our International Advisory Board

¹ Dep. Physics, Portland State University, Portland, OR 97201 ² CRISMAT-ENSICAEN, Université de Caen-Basse Normandie,F-14032 Caen, France, ³ Dep. Geosciences, University of Arizona, Tucson, AZ 85721-0077, ⁴ Institute of Biotechnology, Vilnius University, Graiciuno 8, LT- 02241, Vilnus, LT-02241, Lithuania, grazulis@ibt.lt, ⁵ Lab. des Oxydes et Fluorures, Université du Maine, 72085 Le Mans Cedex 9, France, ⁶ Dep. Materials Engineering, University of Trento, 38050 Trento, Italy, ⁷ Dep. de Química Inorgánica, Universidad de Granada, 18071 Granada, Spain

Funding: Research Council of Lithuania, NorthWest Academic Computing Consortium, National Science Foundation, PANalytical, Crystal Impact & our various home institutions
1. Motivation: the time was right to do this ten years ago, so we did, now we have something quite significant to show for.

2. Open access crystallographic databases

“There is nothing so powerful as an idea whose time has come.”
Victor Hugo, 1802 - 1885

3. Crystallography Open Database in its 11th year

4. Efforts by Portland State’s Nano-Crystallography Group / future Bicrystallography Open Database
“Changing the culture of science from one where publications were viewed as the primary product of the scientific enterprise to one that also equally values data.” William Michener (of the DataONE project), in: Nature 461 (2009) 160-163
Crystallographic database

A crystallographic database is a database specifically designed to store information about crystal structures. A crystal structure is a repeating pattern of atoms in three dimensions, having, in all three dimensions of space, a regularly repeating arrangement of atoms, ions, or molecules that are held together by either covalent, ionic, hydrogen, or van der Waals forces. Crystallographic databases are created using data from X-ray crystallography, neutron diffraction, and other crystallographic techniques.

Crystallographic databases are used to store information about the structure of materials, such as minerals, metals, and organic compounds. They are also used to store information about biological macromolecules, such as proteins and nucleic acids. Crystallographic databases are often used by researchers to analyze the structure of materials and to develop new materials with desired properties.

Crystallographic databases can be accessed through the Internet or through local servers. They are often used in the development of new materials and in the study of the structure of biological molecules. Crystallographic databases are an important tool for researchers in many fields, including materials science, chemistry, and biology.
Advisory Board
Daniel Chateigner, Xiaolong Chen, Marco Ciriotti, Robert T. Downs, Saulius Gražulis, Armel Le Bail, Luca Lutterotti, Yoshitaka Matsushita, Peter Moeck, Miguel Quirós Olozábal, Hareesh Rajan, Alexandre F.T. Yokochi

http://www.crystallography.net
mirrors worldwide http://cod.ibt.lt
http://cod.ensicaen.fr
http://qiserver.ugr.es/cod
http://nanocrystallography.org
web portal: http://nanocrystallography.net

more than 235,000 entries
Neutron and x-ray structure refinements between 15 and 1083 K of piezoelectric gallium arsenate, Ga As O₄: temperature and pressure behavior compared with other $\alpha$-quartz materials

COD Advisory Board thanks The Research Council of Lithuania for their financial publication "Crystallography Open Database (COD): an open-access collection of crystal structures for world-wide collaboration", Nucleic Acids Research. (2012) PDF version

We thank Crystal Impact GbR for their financial support of the publication "Crystallography Open Database - an open-access collection of crystal structures", J. Appl. Crystallogr. (2009) PDF version

Currently there are 235786 entries in the COD.
Latest deposited structure: 7140974 on 2013-05-17 at 12:48:22 UTC
### Result: There are 100 entries in the selection

You can download the COD numbers of the selection as a text file.
You can download all files as a single ZIP archive.

#### Searching COD ID like 201710

<table>
<thead>
<tr>
<th>CIF file</th>
<th>Formula: C5 H9 N O6 S</th>
</tr>
</thead>
<tbody>
<tr>
<td>HKL data</td>
<td>Comments: Minkov, Vasily S.; Boldyreva, Elena V.</td>
</tr>
<tr>
<td>Original</td>
<td>&lt;small&gt;DL&lt;/small&gt;-Cysteinium semioxalate Acta</td>
</tr>
<tr>
<td>IUCr paper</td>
<td>Crystallographica Section C 65(5) (2009) o245-o247</td>
</tr>
</tbody>
</table>

**Space group:** P 1  
**Cell volume:** 443.62  
**Cell parameters:** 5.6664; 9.0149; 9.7749; 109.349; 102.282; 100.119;

---

### Information card for 2017100

#### Preview

**Acta Crystallographica Section C**  
**Crystal Structure Communications**  
**Volume 65, Part 5 (May 2009)**  
**organic compounds**

**DL-Cysteinium semioxalate**  
**V. S. Minkov and E. V. Boldyreva**

**Abstract:** Two chiral counterparts (L- and D-cysteinium) are present in the structure of the title compound, C₃H₈NO₂S⁺·C₂H₂O₄⁻. The carboxy group of the cysteinium cation relative to the amino group. The crystal structure is built from cysteinium cations are connected to each other not direct semioxalate anions linked to each other via O-H...O hydrogen bonds.

**Formula:** C₃H₈NO₂S⁺·C₂H₂O₄⁻

---

### CIF file

**Download 2017100.cif**

---

### HKL data file

**Download 2017100.hkl**

---

### Original IUCr paper

**Chemical name:** DL-cysteinium semioxalate

---

### Structure parameters

#### Common name

DL-cysteinium semioxalate

---
Crystallography Open Database

CIF Information Card

Information card for 7150000

Coordinates 7150000 cif

Structure parameters

Formula
Calculated formula
Title of publication

Related to carba-pyranoses: synthesis of acetylated derivatives of 4-amino-2,4-dideoxy-3-O-(betahydroxypropyl) D-glucose and carba-pyranose from a D-glucose template.
The "FPSM method" uses a Rietveld like fitting procedure to test all possible crystal structures from a Database, rank them and find the more probable in your diffraction pattern. In the end a Rietveld phase quantification is done with the phases identified. Be aware that if a phase is not present in the database (COD is used here), it cannot be found nor quantify; so this method is limited to only the phases for which a crystal structure has been determine and has been uploaded to the COD database. This page has been constructed to permit other people to use the method and test it on their data. To use it you need to upload a datafile in proper format. Use .prn, a double column format with no title line, first column contains the 2theta coordinates (d space for TOF), the second column the intensity (corrected for incident intensity for TOF). You need also to specify some additional instrument characteristics (wavelength, geometry etc.). If the selected instrumental broadening function much the one of your instrument, then a reasonable analysis of crystallite sizes and microstrain is reported.

Found phases and quantification:

<table>
<thead>
<tr>
<th>Phase ID</th>
<th>name</th>
<th>vol. (%)</th>
<th>wt. (%)</th>
<th>crystallites (Å)</th>
<th>microstrain</th>
</tr>
</thead>
<tbody>
<tr>
<td>9004178</td>
<td>Zincite</td>
<td>20.5148</td>
<td>29.1683</td>
<td>1893.8</td>
<td>4.76139e-08</td>
</tr>
<tr>
<td>1000043</td>
<td>Fluorite</td>
<td>42.3438</td>
<td>33.7285</td>
<td>2154.45</td>
<td>0.00036731</td>
</tr>
<tr>
<td>9007498</td>
<td>Corundum</td>
<td>37.1414</td>
<td>37.1032</td>
<td>1941.94</td>
<td>0.000229095</td>
</tr>
</tbody>
</table>

Final Rietveld analysis, Rw: 0.154707, GoF: 1.90021

http://cod.iutcaen.unicaena.fr/
About this Validation Interface

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

Steps

The process of files deposition, after you have uploaded your data is pretty simple. First step, after files have been uploaded, is validation. Our scripts perform some checks to see if all necessary data are present in the submitted file. Results are displayed to you next to your files.

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.
Theoretical Crystallography Open Database

Open-access collection of theoretically calculated or refined structural information of organic, inorganic, metal-organic compounds and biopolymers

All data on this site have been placed in the public domain by the contributors.

Currently there are 96 entries in the TCOD.

Latest deposited structure: 20000096 on 2013-05-09 at 11:48:23 +0000 (Thu. 09 May 2013)

Advisory Board

Daniel Chatcigner, Xiaolong Chen, Marco Cirera, Robert T. Downs, Sathius Grazulis, Armel Le Bail, Luc Long, Yoshitaka Matsushita, Peter Mocck, Miguel Quiros, Hareesh Rajan, Alexandre F.T. Yokochi

Information card for 3000009

CIF's Donators

Information card for 3000009
Looking for caffeine?

Get it here: www.crystallography.net/cif/2100202

And it's FREE!
http://nanocrystallography.net

International Year of Crystallography: 150 years of modern crystallography http://www.iucr.org/iycr

Welcome to the Open Access Crystallography Resource Portal

Crystallography Open Database: Containing more than 225,000 small molecules and small to medium unit cell crystal structures (including minerals). Main sites www.crystallography.net (in France) and cod.ist.lt (in Lithuania). Less frequently updated mirrors cod.ensicaen.fr (in France), gisember.ugr.es/cod (in Spain), nanocrystallography.org and nanocrystallography.research.pdx.edu/search/cod/miro (in North America).

Predicted Crystallography Open Database: Containing over 1,000,000 inorganic compounds (silicates, phosphates, sulfates of Al, Ti, V, Ga, Nb, Zr, zeolites, fluorides, etc.). Derived product, P202 (Predicted Powder Diffraction Database) contains all powder patterns calculated from the PCOD. Main site www.crystallography.net/pcod/index.html (in France). Less frequently updated mirrors adpd.unige.lemans.fr/cod/pcod (in France) and nanocrystallography.net/pcod/miro/html (in North America).

Portland State University’s Nano-Crystallography Group: nanocrystallography.research.pdx.edu hosts small educational open access databases of common crystal structures (e.g. EDU-COD and CMD). The Wiki Crystallography Database, to which we invite you to contribute, offers over 8,000 entries on minerals.

All good open access crystallography resources, e.g. space group drawings, ...
searchable collection of CIFs for all kinds of simulations and visualizations of grain boundaries to be derived from user inputs and freely modifiable at the atomic level

primitive cubic lattice, 5 atoms per lattice point of dichromatic pattern P4/mm’ m’, zero rigid body shift and expansion for simplicity


$\text{SrTiO}_3 \Sigma = 13a, (510)/[001], 22.6^\circ$ tilt boundary in [001] projection, sectioned at $\frac{1}{2}$ [510], large disks Sr columns, medium disks pure O columns, small disks mixed O and Ti columns
primitive cubic lattice, 5 atoms per lattice point of dichromatic pattern P4/mm’ m’, zero rigid body shift and expansion for simplicity

\textit{Aberration-corrected translation-symmetry averaged STEM Z-contrast, H. Yang et al., Phil. Mag. 2012, 1-11, iFirst Article.}

\[
\text{SrTiO}_3 \Sigma = 13a, (510)/[001], 22.6^\circ \text{ tilt boundary in [001] projection, sectioned at } \frac{1}{4} [510], \text{ large disks Sr columns, medium disks pure O columns, small disks mixed O and Ti columns}
\]
Summary

Open access crystallographic databases, combined some 350,000 – 400,000 entries

COD in it’s 11th year, more than 235,000 entries


TCOD in it’s first few months, 96 entries

Open access crystallography resource portal, nanocrystallography.net

Plans for Bicrystallography Open Database

We are also for collaborations, but open access crystallographic data and web sites are not going to go away.
Lifecycle of data incorporating primary data, derived data, and workflow artifacts with corresponding provenance metadata for quality assessment.

www.dataONE.org
The Coalition for Networked Information (CNI) is an organization dedicated to supporting the transformative promise of digital information technology for the advancement of scholarly communication and the enrichment of intellectual productivity. Some 200 institutions representing higher education, publishing, information technology, scholarly and professional organizations, foundations, and libraries and library organizations make up CNI's members; CNI is entirely funded through membership dues. Semi-annual membership meetings bring together representatives of CNI's constituencies to discuss ongoing and new projects, and to plan for future initiatives.

CNI is based in Washington, DC and led by Executive Director Clifford A. Lynch and Associate Executive Director Joan K. Lippincott.

Our current Program Plan is available online. The Program Plan is a snapshot of our plans and priorities for the year as of early November. It includes some background information about CNI and its collaborative activities, and discusses program initiatives planned for the year (July 1-June 30).

For announcements about the CNI community, subscribe to CNI-ANNOUNCE or point your news reader to CNI News.

More information is available about the history of CNI. See also Key Benefits of CNI Membership and the Membership FAQ.

Last updated: Thursday, December 13th, 2012
Browse By Topic

Home / Browse by Topic

The Topics pages organize some material, produced by CNI staff and/or hosted on our website, around a few themes that have been important to CNI’s work in recent years. Each topic page includes a brief description of that subject, as well as links to some representative materials in that area. Please note that topic page lists are not comprehensive, and, in most cases, they are primarily representative of CNI’s most recent activity.

Topics

- Assessment (37)
- Cyberinfrastructure (66)
- Digital Curation (88)
- Digital Humanities (52)
- Digital Libraries (147)
- Digital Preservation (113)
- E-Books (26)
- E-Journals (27)
- E-Science (76)
- Economic Models (49)
- Electronic Theses & Dissertations (ETDs) (11)
- Identity Management (20)
- Learning Spaces (includes Information/Learning Commons) (38)
- Metadata (77)
- Mobile Technologies (11)
- Net Generation (includes Millennials) (19)
- Personal Archives (7)
- Publishing (57)
- Repositories (110)
- Scholarly Communication (94)
- Social Media (24)
- Special Collections (21)
- Standards (42)
- Teaching & Learning (104)
Scholarly Note-Taking On The Web

http://www.cni.org/

Michael Buckland (UC Berkeley) & Ryan Shaw (UNC Chapel Hill) explore how current Web technology can aid scholarly editing projects by making research notes available through Web publication, as one example. Their presentation was made at CNI’s recent spring 2013 member meeting in San Antonio, TX.
Data-Driven High-Throughput Prediction of the 3-D Structure of Small Molecules: Review and Progress. A Response to the Letter by the Cambridge Crystallographic Data Centre

Pierre Baldi

Department of Computer Science, University of California, Irvine, Irvine, California 92697-3435, United States

Published: November 22, 2011

ultimately for advancing our understanding of the data they are meant to interpret. Accurate prediction of 3-D structures is central to chemistry and drug discovery; thus, any restrictions in this area impact not only the scientists that are involved but ultimately all the tax payers.

As history shows, those who stand in the way of democracy and scientific progress end up losing over the long run. The reactionary attitude of the CCDC staff has started to backfire by energizing academic laboratories around the world to find alternative solutions around the CCDC. There are already several efforts (e.g., Crystallography Open Database<sup>3</sup> and CrystalEye<sup>4</sup>) to produce large, freely available databases of crystallographic structures using the same main source as the CSD—publicly available data. Furthermore, quantum mechanical methods have now reached the level of

---

Predicted Crystallography Open Database

Upload data
or
Search the database

Updated November 2009 : 1,062,771 entries in the PCOD
The largest collection of CIFs in the world
(see the "What is New" page)

Content:
PCOD contains (see the complete list) inorganic compounds (silicates, phosphates, sulfates of Al, Ti, V, Ga, Nb, Zr, zeolites, fluorides, etc) predicted - or enumerated - mainly by ZEFSA II
(898,707 SiO_2 entries) or by GRINSP (163,520 entries), or by other programs.

Derived product:
The P2D2 (Predicted Powder Diffraction Database) contains all powder patterns calculated from the PCOD, assembled in a system allowing for search-match (by EVA from Bruker).

http://www.mwdeem.rice.edu/zefsall/

http://sdpd.univ-lemans.fr/grinsp/index.html

http://www.crystallography.net/pcod/P2D2/EVA/index.html
phase diagrams + crystal structures + physical properties

...together in the world largest database for inorganic compounds

Dear Dr. Villars:

It is surely fine that you are starting a very large scale project to extend the database to cover all non-organic solid state materials. Also, it is fine that you want to call it the PAULING'S FILE.

This letter can serve as my permission.

Sincerely,

Linus Pauling

27 August 1993

Inorganic Material Database (AtomWork)

Outline

The Inorganic Material Database aims to cover all basic crystal structure, x-ray diffraction, property and phase diagram data of inorganic and metallic materials from main literature sources.

You have three choices to search data:
1. "Search material" - Search materials by specifying chemical system, chemical formula, substance name, structure type (prototype), Pearson symbol or space group number.
3. "Search phase diagrams" - Search phase diagrams by specifying chemical system.

As of July 1, 2010, the list of registered data had reached 32,000 crystal structures, 56,000 material properties and 15,000 phase diagrams.

Disclaimer

- National Institute for Materials Science (NIMS) holds the copyright of this database system.
- No reproduction, republication or distribution to third parties of any content is permitted without written permission of NIMS.
- NIMS takes no responsibility for any damage incurred by the user as a result of using this database system.

About AtomWork data

AtomWork data part is a collaboration among: National Institute for Materials Science (NIMS), Tsukuba, Japan, Materials Phases Data System (MPDS), Yizhau, Switzerland, and Japan Science Technology Corporation (JST), Tokyo, Japan.

AtomWork data part is copyrighted by National Institute for Materials Science and Materials Phases Data System.

AtomWork data, release 2010/08, is published by National Institute for Materials Science, Tsukuba, Japan. Copyright©2010. All Rights Reserved.
Copyright? What copyright?

- Copyright covers works of authorship (novels, verse, sci. papers, computer programs)
- Copyright covers only the expression of the ideas
- Copyright does not cover:
  - Ideas
  - (scientific) facts
  - Simple forms (i.e. ones that do not contain individual’s “trace of the hand”)

Idea-expression divide, quoted from Wikipedia

... courts have recognized that there are particular ideas that can be expressed intelligibly only in one or a limited number of ways. The French name for this doctrine is Scènes à faire. Therefore even the expression in these circumstances is unprotected, or extremely limited to verbatim copying only. This is true in the United Kingdom and most Commonwealth countries. In the United States this is known as the merger doctrine, because the expression is considered to be inextricably merged with the idea. United States courts are divided on whether merger constitutes a defense to infringement or prevents copyrightability in the first place, but it is often pleaded as an affirmative defense to copyright infringement.

COD copyright policy

- Include data:
  - _atom_site_fract_x 0.333
- Exclude potentially copyrighted text:
  - _publication_text
  - Introduction

We have solved ...
What Does Copyright Protect?

What does copyright protect?
Copyright, a form of intellectual property law, protects original works of authorship including literary, dramatic, musical, and artistic works, such as poetry, novels, movies, songs, computer software, and architecture. Copyright does not protect facts, ideas, systems, or methods of operation, although it may protect the way these things are expressed. See Circular 1, Copyright Basics, section “What Works Are Protected.”

What Is Not Protected by Copyright?

Several categories of material are generally not eligible for federal copyright protection. These include among others:

- works that have not been fixed in a tangible form of expression (for example, choreographic works that have not been notated or recorded, or improvisational speeches or performances that have not been written or recorded)
- titles, names, short phrases, and slogans; familiar symbols or designs; mere variations of typographic ornamentation, lettering, or coloring; mere listings of ingredients or contents
- ideas, procedures, methods, systems, processes, concepts, principles, discoveries, or devices, as distinguished from a description, explanation, or illustration
Uncreative collections of facts are outside of Congressional authority under the **Copyright Clause** (Article I, § 8, cl. 8) of the **United States Constitution**, therefore no database right exists in the **United States**. Originality is the **sine qua non** of copyright in the United States (see **Feist Publications v. Rural Telephone Service**). This has not stopped database owners lobbying for the introduction of such a right, but so far bills to introduce it in the U.S. have been prevented by the successful lobbying of research libraries, consumer groups and firms who benefit from the free use of factual information.

* Sine qua non, Latin, refers to an indispensable and essential action, condition, or ingredient. It was originally a Latin legal term for "[a condition] without which it could not be," or "but for..." or "without which [there is] nothing".

**Feist Publications, Inc., v. Rural Telephone Service Co.**, 499 U.S. 340 (1991), commonly called **Feist v. Rural**, is an important United States Supreme Court case establishing that information alone without a minimum of original creativity cannot be protected by copyright. In the case appealed, Feist had copied information from Rural's telephone listings to include in its own, after Rural had refused to license the information. Rural sued for copyright infringement. The Court ruled that information contained in Rural's phone directory was not copyrightable and that therefore no infringement existed.
A *database right* is considered to be a *property right*, comparable to but distinct from *copyright*, that exists to recognise the investment that is made in compiling a database, even when this does not involve the 'creative' aspect that is reflected by copyright.

*Sui generis* database right for member states of the European Union

Copyright protection is not available for databases which aim to be "complete", that is where the entries are selected by objective criteria: these are covered by *sui generis* *database rights*. While copyright protects the creativity of an author, database rights specifically protect the "qualitatively and/or quantitatively substantial investment in either the obtaining, verification or presentation of the contents". Database rights are held in the first instance by the person or *corporation* which made the substantial investment, so long as the person is a national or domiciliary of a Member State or the *corporation* is formed according to the laws of a Member State and has its registered office or principal place of business within the European Union.

Article 11(3) provides for the negotiation of treaties to ensure reciprocal treatment outside the EU: as of 2006, no such treaty exists.

*Sui generis* Latin phrase, meaning "of its own kind/genus" and hence "unique in its characteristics".
Re: Crystallographic data copyrights
From: Eben Moglen <moglen@columbia.edu>
Date: Sun, 04 May 2003 09:02:21 -0400 I will have to be brief. If you need to follow up, let me know.

I assume US law governs throughout; an inaccurate but necessary assumption here. If you extract only the actual coordinate data you have no copyright liability. One cannot copyright facts, only the expression incident to factual reporting. This principle was recognized by the US Supreme Court in 1915 with respect to news reports sent by telegraph. The idea/expression distinction has been held by the Supreme Court to prevent assertion of copyright over telephone white pages, where there is no originality in the concept of alphabetic organization of data. More complex forms of association or organization of data might give rise to claims. You should move quickly. Proposals for database protection in the US and Europe will close up vast areas of human knowledge within the next decade. Make this data free soon, or you risk losing the chance. How to license your data so that everyone is compelled to make free their improvements or accessions to it is another subject. Best regards.

--
Eben Moglen voice: 212-854-8382 212-854-8382 FREE
Professor of Law fax: 212-854-7946 moglen@columbia.edu
Columbia Law School, 435 West 116th Street, NYC 10027
General Counsel, Free Software Foundation
http://moglen.law.columbia.edu/
Somebody who reads Wikipedia is “rather in the position of a visitor to a public restroom,” says Mr. McHenry, Britannica's former editor. “It may be obviously dirty, so that he knows to exercise great care, or it may seem fairly clean, so that he may be lulled into a false sense of security. What he certainly does not know is who has used the facilities before him.” One wonders whether people like Mr. McHenry would prefer there to be no public lavatories at all.