

20 years of the COD: disseminating crystallographic data

Saulius Gražulis

Grenoble, 2024

Vilnius University Institute of Biotechnology



Id: slides2-policy.tex 2937 2024-12-31 09:52:31Z saulius December 31, 2024



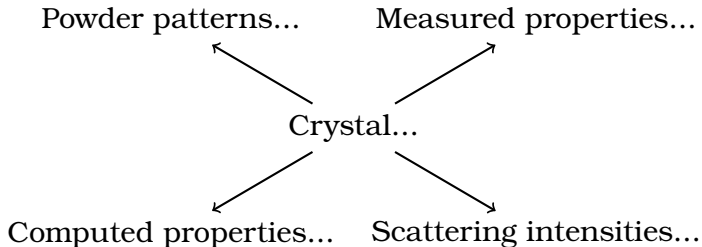
Overview of the talk

- What is COD (with some history);
- COD contents and data curation principles;
- Recent developments with the COD;
- Software for COD data;
- Future prospects;

<https://www.crystallography.net/archives/2024/slides/NOBUGS-Talk/slides2-policy.pdf>

The importance of crystallographic data

All observations *must* be compatible with crystallographic models.



Things to do with Open Data

Crystal structures {
Drug design
Material property prediction
Teaching
Citizen science
Machine learning models
...

PDB is just fine!

The screenshot shows the wwPDB website in a Firefox browser. The header features the 'WORLDWIDE PDB PROTEIN DATA BANK' logo. The main content area is divided into two columns. The left column contains text about the PDB archive and its management by wwPDB, with the phrase 'PDB is freely and publicly available' circled in red. The right column lists three main services: 'Validate Structure or View validation reports', 'Deposit Structure All Deposition Resources', and 'Download Archive Instructions', each with a corresponding icon. The browser's address bar shows the URL 'https://www.wwpdb.org'. The taskbar at the bottom displays various open applications and the system clock.

Since 1971, the Protein Data Bank archive (PDB) has served as the single repository of information about the 3D structures of proteins, nucleic acids, and complex assemblies.

The Worldwide PDB (wwPDB) organization manages the PDB archive and ensures that the PDB is freely and publicly available to the global community.

Learn more about PDB **HISTORY** and **FUTURE**.

- Validate Structure**
or View validation reports
- Deposit Structure**
All Deposition Resources
- Download Archive**
Instructions

Problems with access to data

Proprietary licensing causes a lot of headache in the XXI century...

- CCDC Access Structures Terms and Conditions: “These services must not be used to systematically download or redistribute these structures, data or associated information. Programmatic access to these services is not permitted.”
(<https://www.ccdc.cam.ac.uk/access-structures-terms/>, last accessed 2024-09-20)
- “In the specific case of the article in question, /.../ a small molecule 3-D structure predictor and Web server (COSMOS) /.../ [t]he CCDC vigorously intervened to prevent distribution of such a tool. The statement in the CCDC’s letter that “express permission was immediately granted” is simply false. A dozen librarians and other staff from the University of California (UC) had to intervene under the threat of losing a system-wide license to the CSD.” (Baldi 2011)

The COD project

But what if crystallographers work together to establish a public domain database with all relevant crystallographic data? This would not only overcome the current situation with 'fragmented' databases, it would also prevent for becoming dependent from monopolists.

What would be needed?

1. A small team of engaged scientists with some experience in database and software design to coordinate the project.
2. The authors (i.e. the scientific community = YOU) who provides the project with database entries (note, that if you have'nt sold your experimental results exclusively, you are free to distribute the data to such a database, even if they have already been part of a publication - and a lot of good data have never been published).
3. Free software a) for maintaining the database, b) for data evaluation and calculation of derived data (e.g. calculated powder pattern from crystal structures for search-match purposes), c) for browsing and retrieval.

gemstonede (Dr. Michael BERNDT) Fri Feb 14, 2003 1:26 pm

The Crystallography Open Database (COD)

<https://www.crystallography.net/cod>



Crystallography Open Database

COD Home

[Home](#)
[What's new?](#)

Accessing COD Data

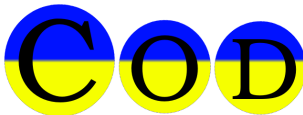
[Browse](#)
[Search](#)
[Search by structural formula](#)

Add Your Data

[Deposit your data](#)
[Manage data positions](#)
[Manage/release prepublications](#)

Documentation

[COD Wiki](#)
[Obtaining COD License](#)
[Privacy and GDPR](#)
[Querying COD](#)
[Citing COD](#)
[COD Mirrors](#)
[Advice to donors](#)
[Useful links](#)



Open-access collection of crystal structures of organic, inorganic, metal-organic compounds and minerals, excluding biopolymers.

Including data and [software](#) from [CrystalEye](#), developed by Nick Day at the [department of Chemistry](#), the University of Cambridge under supervision of [Peter Murray-Rust](#).

All data on this site have been placed in the [public domain](#) by the contributors.

Currently there are **309888** entries in the COD.
Latest deposited structure: [7159763](#) on **2024-01-11** at **01:32:14 UTC**



[CIFs Donators](#)



Advisory Board

Daniel Chateigner, Xiaolong Chen, Marco Ciriotti,
Robert T. Downs, Saulius Gražulis, Werner Kaminsky, Arnel Le Bail, Luca Lutterotti,
Yoshitaka Matsushita, Andrius Merkys, Peter Moeck, Peter Murray-Rust, Miguel Quirós Olozábal,
Hareesh Rajan, Antanas Vaitkus, Alexandre F.T. Yokochi

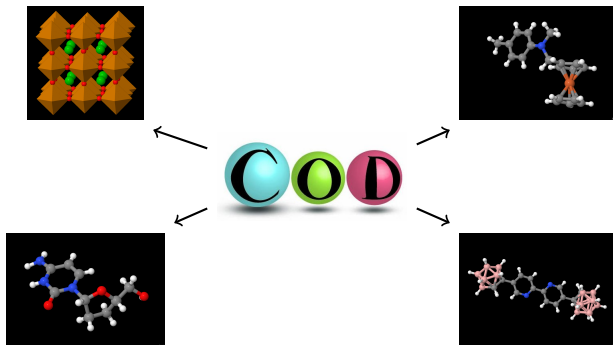
If you find bugs in the COD or have any feedback, please contact us at
cod-bugs@ibt.lt

[Top of the page](#)

All data in the COD and the database itself are dedicated to the public domain and licensed under the [CC0 License](#). Users of the data should acknowledge the original authors of the structural data

COD contents

<https://www.crystallography.net/cod>



509 888 records as of 2024-01-11, available under **CC0 License**

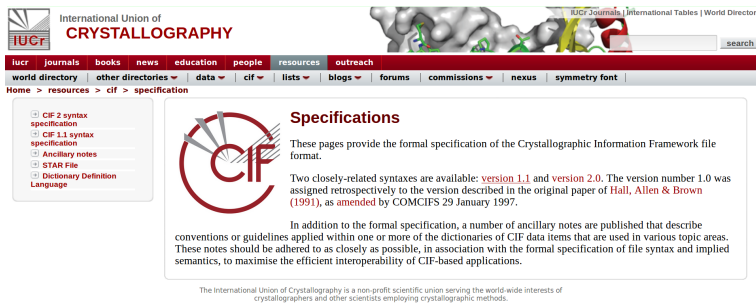
All data are presented in a standardised, machine-readable form (Gražulis et al. 2009; Gražulis et al. 2012).

COD data sources

- Peer-reviewed publications;
- Preprints, dissertations;
- Depositions by crystallographers (pers. comm., pre-publ.);
- Other databases; notably **AMCSD**, maintained by the group of Robert Downs (Downs et al. 2003; Rajan et al. 2006)

<https://rruff.geo.arizona.edu/AMS/amcsd.php>





The screenshot shows the IUCr Crystallography website. The header includes the IUCr logo and the text "International Union of CRYSTALLOGRAPHY". A navigation bar contains links for iucr, journals, books, news, education, people, resources, and outreach. Below this is a secondary navigation bar with links for world directory, other directories, data, cif, lists, blogs, forums, commissions, nexus, and symmetry font. The main content area is titled "Specifications" and features a large red "CIF" logo. The text explains that these pages provide the formal specification of the Crystallographic Information Framework file format. It mentions two versions: version 1.1 and version 2.0, with version 1.0 being assigned retrospectively to the original paper of Hall, Allen & Brown (1991), as amended by COMCIFS 29 January 1997. It also notes that a number of ancillary notes are published to describe conventions or guidelines applied within one or more of the dictionaries of CIF data items. A footer note states that the IUCr is a non-profit scientific union serving the world-wide interests of crystallographers and other scientists employing crystallographic methods.

International Union of
CRYSTALLOGRAPHY

iucr journals books news education people resources outreach

world directory | other directories | data | cif | lists | blogs | forums | commissions | nexus | symmetry font

Home > resources > cif > specification

☐ CIF 2 syntax specification
☐ CIF 1.1 syntax specification
☐ Ancillary notes
☐ STAR File
☐ Dictionary Definition Language

Specifications

These pages provide the formal specification of the Crystallographic Information Framework file format.

Two closely-related syntaxes are available: [version 1.1](#) and [version 2.0](#). The version number 1.0 was assigned retrospectively to the version described in the original paper of [Hall, Allen & Brown \(1991\)](#), as amended by COMCIFS 29 January 1997.

In addition to the formal specification, a number of ancillary notes are published that describe conventions or guidelines applied within one or more of the dictionaries of CIF data items that are used in various topic areas. These notes should be adhered to as closely as possible, in association with the formal specification of file syntax and implied semantics, to maximise the efficient interoperability of CIF-based applications.

The International Union of Crystallography is a non-profit scientific union serving the world-wide interests of crystallographers and other scientists employing crystallographic methods.

(Hall et al. 1991; Bernstein et al. 2016)

The Crystallographic Interchange File/Framework (CIF):

- Provides standard means for data publishing and exchange;
- Is suitable for archiving;
- Is maintained by the IUCr;

Accessing the COD

COD data can be accessed:

- 1 Via the Web page:

<https://www.crystallography.net/cod/7159763.html>

- 2 Via the COD REST API:

<https://www.crystallography.net/cod/7159763.cif>

<https://www.crystallography.net/cod/result?text=perovskite>

- 3 Via the OPTIMADE API (Andersen et al. 2021):

[https://www.crystallography.net/cod/optimade/structures?filter=elements+HAS+\"U\"](https://www.crystallography.net/cod/optimade/structures?filter=elements+HAS+\)

- 4 Via SQL:

```
mysql -u cod_reader -h sql.crystallography.net cod -e \  
'select file from data where formula = \"- H2 O -\"'
```

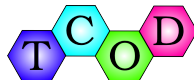
- 5 By downloading to your computer using Subversion, rsync or simple Web download:

<https://wiki.crystallography.net/howtoobtaincod>

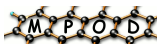
COD “sisters”



<http://www.crystallography.net/cod>
> 500 000 entries



<http://www.crystallography.net/tcod>
> 7400 entries (ready to grow to > 10^7 ?)



<http://mpod.cimav.edu.mx/>
> 300 entries



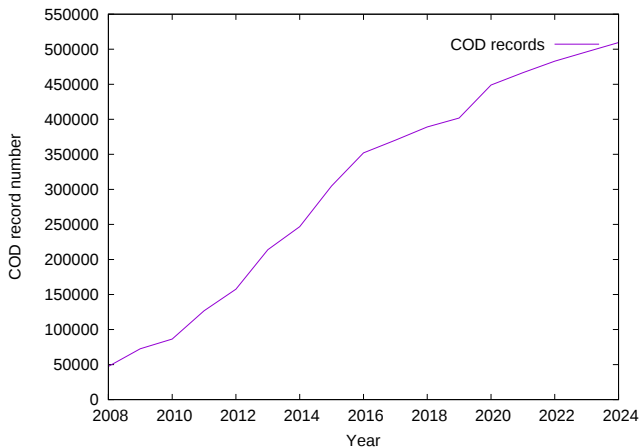
<http://www.crystallography.net/pcod>
> 10^6 entries (ready to grow to > 10^8 ?)



<http://solsa.crystallography.net/rod/>
> 1100 entries

(Gražulis et al. 2009; Gražulis et al. 2012; Pepponi et al. 2012; Fuentes-Cobas et al. 2017; Mendili et al. 2019)

COD growth



Data curation in the COD:

```
svn log -r283960 --diff svn://www.crystallography.net/cod/cif/9
```

```
--- 00/15/9001556.cif (revision 283959)
+++ 00/15/9001556.cif (revision 283960)
@@ -68,8 +68,24 @@
 _atom_site_fract_y
 _atom_site_fract_z
 _atom_site_U_iso_or_equiv
 {+_atom_site_type_symbol+}
 {+_atom_site_attached_hydrogens+}
 Fe 0.25000 0.25000 0.25000 0.00490 {+Fe 0+}
 O-H1 0.50000 0.17800 0.30800 0.00100 {+O 1+}
 O-H2 0.19500 0.19000 0.50000 0.00100 {+O 1+}
 O-H3 0.31800 0.50000 0.32300 0.00100 {+O 1+}
 Wat 0.00000 0.50000 0.50000 0.00640 {+O 2+}
 /.../
```

COD data curation

Inputs from COD users

Thomas Dortmann (2013), PANalytical, “COD-minerals.xlsx”:

*The entries that now have the mineral name are minerals,
the rest are not.*

> 3 500 unique mineral names assigned 104 “atypical” names¹.

Update (2024):

> 4 257 unique mineral names, 566 “atypical” names

¹Not matching the RE `/^[A-Z] [-a-zA-Z ()]+$/`

COD versioning

Essential for reproducibility

All COD changes are tracked in a Subversion repository.

▼ Version history

Revision	Date	Message	Files
277834 (current)	2022-09-14	cif/ Added space group information derived from the space group operation list using the 'cif_filter' program.	2000000.cif
199748	2017-08-14	cif/2/00/00/ (antanas@echidna.ibt.lt) Removing 43 symmetrically equivalent atoms in entry 2000000.	2000000.cif

- The latest revision has a stable URI:
<https://www.crystallography.net/cod/2000000.cif>
- A URI with a specific revision allows to reconstruct the *specific byte stream*:
<https://www.crystallography.net/cod/2000000.cif@199748>

<https://wiki.crystallography.net/howtoobtaincod/>

[COD wiki](#) / Obtaining COD

[Edit](#) [RecentChanges](#) [Preferences](#) [?Discussion](#)

Obtaining COD

COD is an open-access database, and you can freely obtain all data contained in it. You can get the database using one of the following methods outlined below:

COD

Using Subversion

COD, PCOD and TCOD are available as [Subversion](#) repositories, open for anonymous download – you are welcome to update your copies on a regular basis (say, each midnight). This is probably the best and the most complete method to get the databases – after the initial checkout (which, admittedly, can take long), you will have a possibility to update your working copies of the COD,

Partial COD checkout

Saves your disk space and time :)

Recommended: partial checkout of the COD SVN working copy:

```
$ svn co --depth empty svn://crystallography.net/cod
$ cd cod
$ svn up --set-depth infinity cif/
$ svn up --set-depth infinity hkl/
$ svn up --set-depth infinity smi/

$ svn up --set-depth exclude hkl/
```

COD validation and deposition Web site

<https://www.crystallography.net/cod/deposit>

https://www.crystallography.net/cod/initiate_deposition.php

Crystallography Open Databases...

Data block 739121:

- » `_journal_name_full` is undefined
- » neither `_journal_year` nor `_journal_volume` is defined
- » `_journal_page_first` is undefined

Tip: if you need to add bibliography common to all structures in this file, you can add a **data_global** section below, and the data will be distributed into all other sections.

Fetch bibliography by DOI (<http://www.doi.org>):

Save and check Fetch ☐ Pubmed ☒ crossref

Your CIF File contents:

```
data_global
loop
  _publ_author_name
  'Sabiah, Shahulhameed'
  'Lee, Chen-Shiang'
  'Hwang, Wen-Shu'
  'Lin, Ivan J. B.'
  _publ_section_title
;
Facile C-N Bond Cleavage Promoted by Cuprous Oxide: Formation
of C-C-Coupled Bimidazole from Its Methylene-Bridged Congener
;
_journal_issue 2
_journal_name_full Organometallics
_journal_page_first 290
_journal_volume 29
_journal_year 2010
data_714906
_chemical_formula_sum 'C16 H20 Cl4 Cu2 N8'
_chemical_formula_weight 593.28
```

COD validation and deposition Web site

<https://www.crystallography.net/cod/deposit>

https://www.crystallography.net/cod/initiate_deposition.php

The screenshot shows a web browser window titled "Crystallography Open Database: CIF Validator - Mozilla Firefox". The address bar shows the URL <http://www.crystallography.net/store.php?f=0&CODESESSION=ZY0lg8DU9KTyEi-KIIS,grO54O4>. The page content includes the title "Crystallography Open Database Validation and Deposition Interface" and a progress bar with five steps: "Log in", "Upload a file", "Validate data", "Deposit structures", and "Finish". The "Validate data" step is currently active. Below the progress bar, there is a button labeled "Deposit to COD all valid files". A table displays the validation results for a file named "om9010406_si_002.cif", showing its status as "valid" and providing "Edit" and "Deposit to COD" actions. A green message box at the bottom states: "File [om9010406_si_002.cif] is correct".

**Crystallography Open Database
Validation and Deposition Interface**

Log in Upload a file **Validate data** Deposit structures Finish

Deposit to COD all valid files

File	Status	Actions
om9010406_si_002.cif	valid	Edit Deposit to COD

File [om9010406_si_002.cif] is correct

COD data validation policies:

- 1 Syntactic checks:
`$ cifparse 7234818.cif`
- 2 Semantic validation (against dictionaries)
`$ cif_validate -D cif_core.dic 7234818.cif`
- 3 Database-specific checks
`$ cif_cod_check 7234818.cif`

Commands from the `cod-tools` package:

[svn://cod.ibt.lt/cod-tools](https://cod.ibt.lt/cod-tools)

<https://github.com/cod-developers/cod-tools>

COD entry checks – IUCr criteria checks

- Checks on prepublications and Personal communications;
- Checks on published structures;
- *Statistics of structures in the database*

IUCr data validation criteria (Version: 2000.06.09,
<ftp://ftp.iucr.ac.uk/pub/dvntests> or
<ftp://ftp.iucr.org/pub/dvntests>)

COD entry checks – IUCr criteria checks

- Checks on prepublications and Personal communications;
- Checks on published structures;
- *Statistics of structures in the database*

IUCr data validation criteria (Version: 2000.06.09,
<ftp://ftp.iucr.ac.uk/pub/dvntests> or
<ftp://ftp.iucr.org/pub/dvntests>)

`cif_cod_check 3000424.cif`

```
/usr/bin/cif_cod_check: 3000424.cif data_3000424: NOTE, data item  
  '_refine_ls_R_factor_gt' value '0.1120' is > 0.1.  
/usr/bin/cif_cod_check: 3000424.cif data_3000424: NOTE, data item  
  '_refine_ls_wR_factor_ref' value '0.3195' is > 0.25.  
/usr/bin/cif_cod_check: 3000424.cif: NOTE, 2 NOTE(s) encountered.
```

COD deposition types

① Published structure

- Must have full bibliography or at least DOI;
- Assumed to be peer-reviewed, no IUCr checks enforced;

② Pre-publication

- Can be on-hold until publication (default 6 months, extendable);
- IUCr checks performed and enforced;

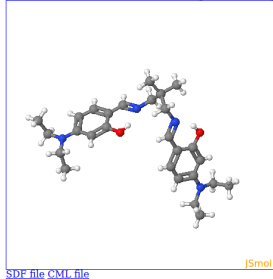
③ Personal communication

- Becomes public immediately;
- IUCr checks performed and enforced;

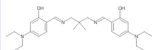
COD chemical repertoire

<http://molecules.crystallography.net/>

[Previous \(2227696\)](#) [Next \(2227698\)](#) [Original COD entry](#)



Reduced structural formula



Reduced canonical SMILES:

CCN(c1ccc(c(c1)O)/C=N/CC(C/C=N=C/c1ccc(cc1O)N(CC)CC)(C)C)CC (**x1**) [PubChem](#)

Unique components

SMILES

CCN(c1ccc(c(c1)O)/C=N/CC(C/C=N=C/c1ccc(cc1O)N(CC)CC)(C)C)CC

InChI

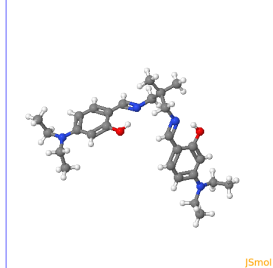
InChI=1S/C27H40N4O2/c1-7-30(8-2)23-13-11-21(25(32)15-23)17-28-19-27(5,6)2/h11-18,32-33H,7-10,19-20H2,1-6H3/b28-17+,29-18+

See also poster by Merkys et al. (<https://bit.ly/3BKZ5vG>)

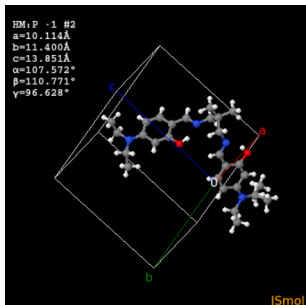
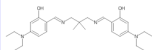
COD chemical repertoire

<http://molecules.crystallography.net/>

[Previous \(2227696\)](#) [Next \(2227698\)](#) [Original COD entry](#)



Reduced structural formula



(Vaitkus et al. 2023)

Reduced canonical SMILES:

CCN(c1ccc(c(c1)O)/C=N/CC(C/N=C/c1ccc(cc1O)N(CC)CC)(C)C)CC (x1) [PubChem](#)

Unique components

SMILES

CCN(c1ccc(c(c1)O)/C=N/CC(C/N=C/c1ccc(cc1O)N(CC)CC)(C)C)CC

InChI

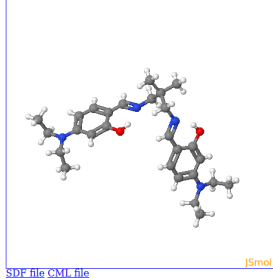
InChI=1S/C27H40N4O2/c1-7-30(8-2)23-13-11-21(25(32)15-23)17-28-19-27(5,6)2/h11-18,32-33H,7-10,19-20H2,1-6H3/b28-17+,29-18+

See also poster by Merkys et al. (<https://bit.ly/3BKZ5vG>)

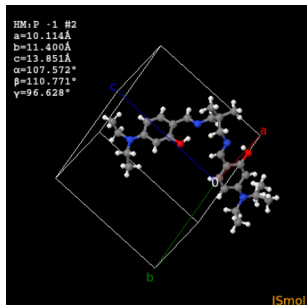
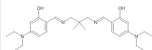
COD chemical repertoire

<http://molecules.crystallography.net/>

[Previous \(2227696\)](#) [Next \(2227698\)](#) [Original COD entry](#)



Reduced structural formula



(Vaitkus et al. 2023)

Reduced canonical SMILES:

CCN(c1ccc(c(c1)O)/C=N/CC(C/C=N/C/c1ccc(cc1O)N(CC)CC)(C)C)CC (x1) [PubChem](#)

Unique components

SMILES

CCN(c1ccc(c(c1)O)/C=N/CC(C/C=N/C/c1ccc(cc1O)N(CC)CC)(C)C)CC

InChI

InChI=1S/C27H40N4O2/c1-7-30(8-2)23-13-11-21(25(32)15-23)17-28-19-27(5,6)2/h11-18,32-33H,7-10,19-20H2,1-6H3/b28-17+,29-18+

See also poster by Merkys et al. (<https://bit.ly/3BKZ5vG>)

COD use cases

COD and PubChem

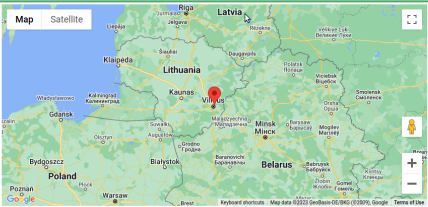
<https://pubchem.ncbi.nlm.nih.gov/source/849>

DATA SOURCES

Crystallography Open Database

The Crystallography Open Database is an open-access collection of crystal structures of organic, inorganic, metal-organics compounds and minerals, excluding biopolymers.

Organization	Vilnius University Institute of Biotechnology
Category	Research and Development
URL	https://www.crystallography.net/cod/
Contact Name	Saulius Gražulis
Address	Saukietikio al. 7, Vilnius, Lithuania, LT-10257
Data Source ID	849
Data in PubChem	203,088 Live Substances
Last Updated	2021/05/17



COD use cases

COD and PubChem

<https://pubchem.ncbi.nlm.nih.gov/substance/164348954>

SUBSTANCE RECORD

6-(2-Bromobenzylamino)purine monohydrate

PubChem SID	164348954
Structure	 2D
Source	Crystallography Open Database
External ID	2210002
Source Category	Research and Development
Version	1 Revision History
Status	Live
Related Compounds	PubChem CID CID 71768516 (6-(2-Bromobenzylamino)purine monohydrate) Component CID CID 962 (Water) CID 61402401 (N-[(2-bromophenyl)methyl]-7H-purin-6-amine) Parent CID CID 61402401 (N-[(2-bromophenyl)methyl]-7H-purin-6-amine)

Cite

Download

CONTENTS

Title and Summary

1 2D Structure

2 3D Conformer

3 Identity

4 Depositor Comments

5 Related Records

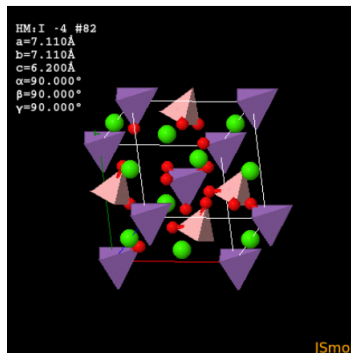
6 Information Sources

Data cross-referencing

External links

Links to external databases are implemented and populated:

- Implemented: AMCSD, Wikidata, Wikipedia, MPOD, ChemSpider;
- Planned: PubChem, **raw diffraction data**;



Coordinates

[9016740.cif](#)

External links

[AMCSD](#); [Wikidata](#); [Wikipedia](#)

Interlinked data in COD



Crystallography Open Database

COD Home

[Home](#)
[What's new?](#)

Accessing COD Data

[Browse](#)
[Search](#)
[Search by structural formula](#)

Add Your Data

[Deposit your data](#)
[Manage depositions](#)
[Manage/release prepublications](#)

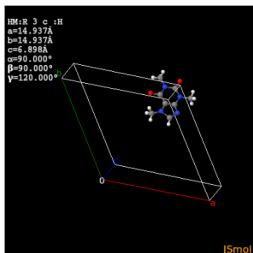
Documentation

[COD Wiki](#)
[Obtaining COD](#)
[Querying COD](#)
[Citing COD](#)
[COD Mirrors](#)
[Advices to donors](#)
[Useful links](#)

Information card for 2100202

[2100201](#) << [2100202](#) >> [2100203](#)

Preview



[Display in Jmol](#)

Coordinates

Original IUCr paper

External links

[2100202.cif](#)

[HTML](#)

[ChemSpider](#); [DrugBank](#); [PubChem](#); [Wikipedia](#)

```
select * from wikipedia_x_cod
```

-----+			
id	ext_id	cod_id	relation_id
-----+			
1	Ibuprofen	2006278	1
2	Caffeine	2100202	1
3	Serotonin	2019147	1
4	Pristinamycin	1000001	1
5	Cucurbituril	1516465	1
6	Rubrene	1516682	1
-----+			

▼ Structure parameters

Group theory in Ada/SPARK

examples/group_theory.ads

```
pragma Spark_Mode (On);
```

```
generic
```

```
  type Element is private;
```

```
  Identity : Element;
```

```
  with function "*" (E, F: Element) return Element is <>;
```

```
function Is_Closed_On_Multiplication (G : Group) return Boolean
```

```
is (for all E of G =>
```

```
    (for all F of G => (Belongs_To (E*F, G))))
```

```
  with Ghost;
```

```
function All_Elements_Have_Inverses (G : Group) return Boolean
```

```
is (for all E of G => Has_Inverse (E, G))
```

```
  with Ghost;
```

```
function Is_Group (G : Group) return Boolean
```

```
is (Has_Identity (G) and then
```

```
    All_Elements_Have_Inverses (G) and then
```

```
    Is_Closed_On_Multiplication (G)
```

```
  )
```

```
  with Ghost;
```

(Petrauskas et al. 2022)

Automatic compilation of proven code

Ada and SPARK

examples/make_group.ads

```
8  type Ring_Element is mod 37;
```

```
29 function Build_Group (E : Ring_Element) return Group
30 with
31   Post => Is_Group (Build_Group' Result);
```

gnatprove -P main.gpr --report=all make_group.adb

```
make_group.ads:23:14: info: postcondition proved
make_group.ads:27:14: info: postcondition proved
make_group.ads:31:14: info: postcondition proved
group_theory.ads:16:15: info: postcondition proved, in instantiation at make_group.ads:16
```

```
saulius@tasmanijos-velnias spacegroups/ $ ./run_make_group 8
(1, 8, 27, 31, 26, 23, 36, 29, 10, 6, 11, 14)

saulius@tasmanijos-velnias spacegroups/ $ ./run_make_group 7
(1, 7, 12, 10, 33, 9, 26, 34, 16)
```

Why Ada/SPARK?

- 1 Durable design – first designed in 1983!
- 2 Modern language – latest standard is Ada 2022;
- 3 Mostly backwards compatible;
- 4 Good F/LOSS compiler available – GNAT;
- 5 Ada is statically very strictly typed;
- 6 Programs are easy to read (Level (Ada) > Level (C));
- 7 Ada & SPARK have a rich type system;
- 8 Language level concurrent programming;
- 9 Produces fast optimised native code, links with any language;
- 10 Not controlled by any private company;



(Amiard et al. 2022)

Why is Ada not popular (yet)?

- ① The language is complex and difficult to implement;
 - ② No good compilers in the 1990's;
 - ③ Procured by the DOD, used for “war fighting software”;
 - ④ Poor academic outreach in the 20th century;
- (National Research Council 1997)

Why is Ada not popular (yet)?

- ❶ ~~The language is complex and difficult to implement;~~
- ❷ The language is rich and convenient to program/design in;
- ❸ ~~No good compilers in the 2020's;~~
- ❹ Very nice compiler and dev. system available: gnat;
- ❺ Procured by the DOD, used for “war fighting software”;
- ❻ Used for mission-critical software (avionics, spacecraft ctrl., railways, plant ctrl...)
- ❼ Poor academic outreach in the 21st century;
- ❽ SPARK allows formal verification of the code (!);

Why is Ada not popular (yet)?

- ① ~~The language is complex and difficult to implement;~~
- ② The language is rich and convenient to program/design in;
- ③ ~~No good compilers in the 2020's;~~
- ④ Very nice compiler and dev. system available: gnat;
- ⑤ Procured by the DOD, used for “war fighting software”;
- ⑥ Used for mission-critical software (avionics, spacecraft ctrl., railways, plant ctrl...)
- ⑦ Poor academic outreach in the 21st century;
- ⑧ SPARK allows formal verification of the code (!);

A code example

From https://github.com/sauliusg/decode-Hall-symbol:examples/decode_hall.adb

```
Build_Group (Symmetry_Operators, N_Symmetry_Operators);

-- Add centering and inversion matrices:

declare
  M : Positive := N_Symmetry_Operators;
  New_Symmetry_Operator : Symmetry_Operator;
begin
  for I in 1..N_Inversions loop
    for C in 1..N_Centering loop
      if I /= 1 or else C /= 1 then
        for S in 1..N_Symmetry_Operators loop
          New_Symmetry_Operator :=
            Symmetry_Operators (S) * Centering (C) * Inversions (I);
          if not Has_Symmetry_Operator (Symmetry_Operators, M,
                                         New_Symmetry_Operator) then

            M := M + 1;
            Symmetry_Operators (M) := New_Symmetry_Operator;
          end if;
        end loop;
      end if;
    end loop;
  end loop;
  N_Symmetry_Operators := M;
end;
```

Compile and use it!

From <https://github.com/sauliusg/decode-Hall-symbol>:

```
$ decode_hall '-P 2ab'
```

```
saulius@starta slides/ $ decode_hall '-P 2ab'
```

```
X,Y,Z
```

```
-X+1/2,-Y+1/2,Z
```

```
-X,-Y,-Z
```

```
X+1/2,Y+1/2,-Z
```

Where to go further?

- Collect more structures;
- Find all papers with crystal structures;
- Write reliable, readable software;
- Apply machine learning;
- Expand the community – **your contributions are invaluable!**

Acknowledgements

VU LSC IBT (KICIS)

Andrius Merkys
Antanas Vaitkus
Algirdas Grybauskas
Yaroslav Rozdobudko

QM community

Audrius Alkauskas
Vytautas Žalandauskas
Lukas Razinkovas
Nicola Marzari
Giovanni Pizzi
Lubomir Smrcok
Linas Vilčiauskas
Rickard Armiento

VU MIF II (FMG)

Linas Laibinis
Karolis Petrauskas

COD Advisory board

Daniel Chateigner
Robert T. Downs
Werner Kaminsky
Armel Le Bail
Luca Lutterotti
Peter Moeck
Peter Murray-Rust
Miguel Quirós

Cheminf community

Evan Bolton
Paul Thiessen
Thomas Sander

Enormous thanks for our commercial users and supporters: PANalytical, Rigaku, Bruker

Funding:

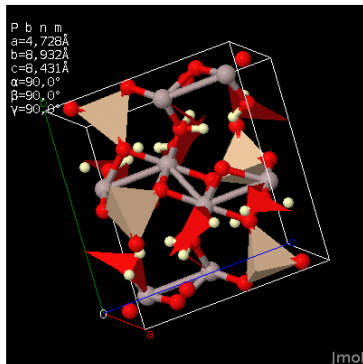
Lithuanian-French Program “Gilibert” (RCoL grant S-LZ-23-3); CECAM; RCoL grants S-MIP-20-21, S-MIP-23-87, VU Intramural funding.

Data Centre: **Research Council of Lithuania Project No.: S-A-UEI-23-11**

Thank you!



<http://en.wikipedia.org/wiki/Topaz>



Coordinates

[2207377.cif](#)

Original IUCr paper

[HTML](#)

<http://www.crystallography.net/2207377.html>

<https://www.crystallography.net/archives/2024/slides/NOBUGS-Talk/slides2-policy.pdf>

References I



Amiard, Raphaël et al. (Apr. 2022). *Learning Ada*. Ed. by Richard Kenner. WEB: <https://learn.adacore.com/>. AdaCore. URL: https://learn.adacore.com/pdf_books/learning-ada.pdf.



Andersen, Casper W. et al. (Aug. 2021). “OPTIMADE, an API for exchanging materials data”. In: *Scientific Data* 8.1, pp. 1–10. DOI: 10.1038/s41597-021-00974-z.



Baldi, Pierre (2011). “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”. In: *Journal of chemical information and modeling* 51, p. 3029. DOI: 10.1021/ci200460z. URL: <http://pubs.acs.org/doi/abs/10.1021/ci200460z>.



Bernstein, Herbert J. et al. (Feb. 2016). “Specification of the Crystallographic Information File format, version 2.0”. In: *Journal of Applied Crystallography* 49.1, pp. 277–284. ISSN: 1600-5767. DOI: 10.1107/s1600576715021871. URL: <http://dx.doi.org/10.1107/S1600576715021871>.



Downs, Robert T. et al. (2003). “The American Mineralogist crystal structure database”. In: *American Mineralogist* 88, pp. 247–250. URL: http://geo.arizona.edu/xtal/group/pdf/am88_247.pdf.



Fuentes-Cobas, Luis E. et al. (Aug. 2017). “The representation of coupling interactions in the Material Properties Open Database (MPOD)”. In: *Advances in Applied Ceramics* 116.8, pp. 428–433. DOI: 10.1080/17436753.2017.1343782.

References II



Gražulis, Saulius et al. (2009). “Crystallography Open Database – an open-access collection of crystal structures”. In: *Journal of Applied Crystallography* 42, pp. 726–729. DOI: 10.1107/S0021889809016690. URL: <http://dx.doi.org/10.1107/S0021889809016690>.



Gražulis, Saulius et al. (2012). “Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration”. In: *Nucleic Acids Research* 40, pp. D420–D427. DOI: 10.1093/nar/gkr900. URL: <http://nar.oxfordjournals.org/content/40/D1/D420.abstract>.



Hall, S. R. et al. (1991). “The crystallographic information file (CIF): a new standard archive file for crystallography”. In: *Acta Crystallographica Section A* 47, pp. 655–685. DOI: 10.1107/S010876739101067X. URL: <http://dx.doi.org/10.1107/S010876739101067X>.



Mendili, Yassine El et al. (May 2019). “Raman Open Database: first interconnected Raman–X-ray diffraction open-access resource for material identification”. In: *Journal of Applied Crystallography* 52.3, pp. 618–625. DOI: 10.1107/s1600576719004229.



National Research Council (Jan. 1997). *Ada and beyond*. National Academies Press. DOI: 10.17226/5463.

References III



Pepponi, Giancarlo et al. (2012). “MPOD: A Material Property Open Database linked to structural information”. In: *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 284.0. E-MRS 2011 Spring Meeting, Symposium M: X-ray techniques for materials research-from laboratory sources to free electron lasers, pp. 10–14. ISSN: 0168-583X. DOI: 10.1016/j.nimb.2011.08.070. URL: <http://www.sciencedirect.com/science/article/pii/S0168583X11008639>.



Petrauskas, Karolis et al. (May 2022). “Proving the correctness of the algorithm for building a crystallographic space group”. In: *Journal of Applied Crystallography* 55.3, pp. 515–525. DOI: 10.1107/s1600576722003107.



Rajan, H. et al. (2006). “Building the American Mineralogist Crystal Structure Database: A recipe for construction of a small Internet database”. In: *Geoinformatics: Data to Knowledge*. Ed. by A.K. Sinha. Vol. 397. Geological Society of America Special Papers. Boulder, CO, United States: Geological Society of America, pp. 73–80. DOI: 10.1130/2006.2397(06).



Vaitkus, Antanas et al. (Dec. 2023). “A workflow for deriving chemical entities from crystallographic data and its application to the Crystallography Open Database”. In: *Journal of Cheminformatics* 15.1. ISSN: 1758-2946. DOI: 10.1186/s13321-023-00780-2.