

# The impact of the Crystallography Open Database on Open Science

Saulius Gražulis

Calgary, 2026

Vilnius University, Life Sciences Center, Institute of Biotechnology



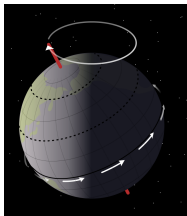
Id: slides.tex 3459 2026-08-11 14:18:36Z saulius August 11, 2026



# Data importance

## Hipparchus (c. 190 – c. 120 BCE)

- measured the longitude of Spica and Regulus and other bright stars
- compared his measurements with data from his predecessors, Timocharis and Aristillus, who lived  $\approx$ **100** years before him,
- discovered what is now called *the precession of the equinoxes*



By NASA, Public Domain

([Wikipedia](#), see also articles on [Timocharis](#) and [Aristyllus](#))

# Computing in crystallography

## History, background

Crystallography – one of the earliest applications of computers in science :) (Shaffer et al. 1946);

650

SHAFFER, SCHOMAKER, AND PAULING

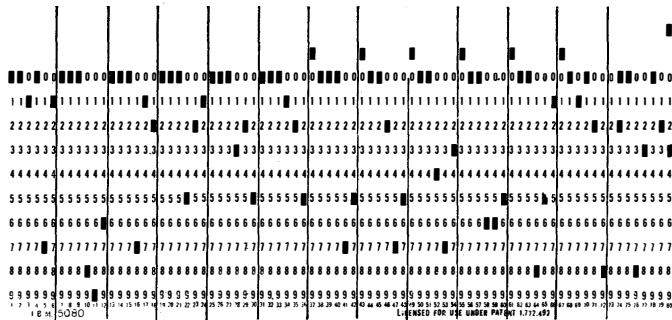


FIG. 1. A typical punched card. The vertical lines indicate the twelve function fields and the classification field for the card design described in the text.

# Distribution of data



US National Weather Records Center: 400 million punched cards (1960)

<https://www.noaa.gov/heritage/resource-collections/managing-legacy-climate-data-in-20th-century>



FAIR (Wilkinson et al. 2016), FAIR/O



**PROTEIN  
DATA BANK**



# PDB is excellent!

wwPDB: Worldwide Protein Data Bank - Mozilla Firefox

File Edit View History Bookmarks Tools Help

wwPDB: Worldwide Protein Data Bank

https://www.wwpdb.org

249%

Search

DuckDuckGo Google Moodle3 My Moodle Wiktionary SG wwPDB RCSB PDB PDBe PDBj Eramet Portal JabRef Discourse Swedbank

**WORLDWIDE PDB**  
PROTEIN DATA BANK

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**.

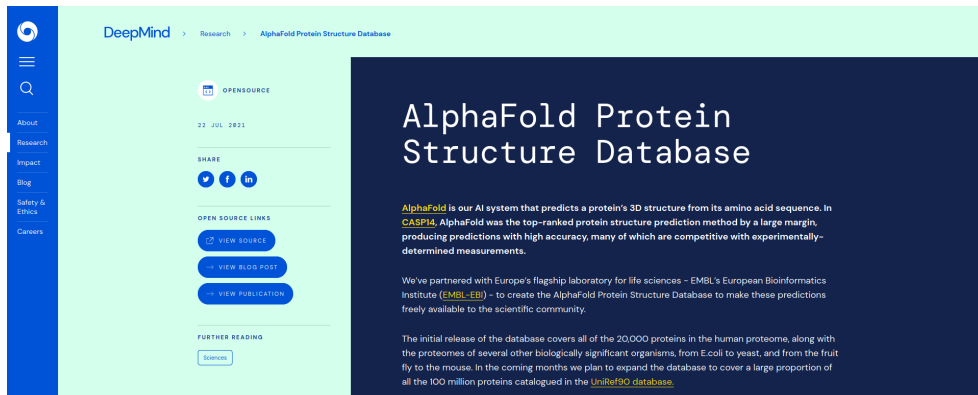
**Validate Structure**  
or View validation reports

**Deposit Structure**  
All Deposition Resources

**Download Archive**  
Instructions

# Consequences: AlphaFold

<https://deepmind.com/research/open-source/alphafold-protein-structure-database><sup>1</sup>



“Our models are trained on structures extracted from the PDB” (Senior et al. 2020).

<sup>1</sup>(accessed 2021-11-23)

# The COD Project

## Michaelio Berndto manifesto

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 COD Project

## Michaelio Berndto manifesto

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

Establishment of COD (2003),  
Announcement at the IUCr  
(2005):



**COD (CRYSTALLOGRAPHY OPEN DATABASE) and  
PCOD (PREDICTED)**

**COD Advisory Board :**

Daniel Chateigner (France), Xiaolong Chen (China), Marco E. Cifelli (Italy), Lachlan M.D. Cranswick (Canada),  
Robert T. Doornik (USA), Arnel Lu Badi (France), Luca Lutterotti (Italy), Alexandre P.T. Yokochi (USA)  
New Members : Yoshitaka Matsushita (Japan), Miguel Quirós Olazábal (Spain)



(Chateigner et al., 2005)



# The COD Project

## Michaelio Berndto manifesto

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

Establishment of COD (2003),  
Announcement at the IUCr  
(2005):



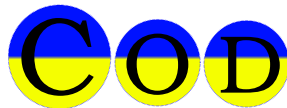
COD (CRYSTALLOGRAPHY OPEN DATABASE) and  
PCOD (PREDICTED)

COD Advisory Board :  
Daniel Chateigner (France), Xiaolong Chen (China), Marco E. Cifelli (Italy), Lachlan M.D. Cranswick (Canada),  
Robert T. Doornik (USA), Armel Le Bail (France), Luca Lutterotti (Italy), Alexandre P.T. Yokochi (USA)  
New Members : Yoshitaka Matsushita (Japan), Miguel Quirós Clotubal (Spain)



(Chateigner et al., 2005)

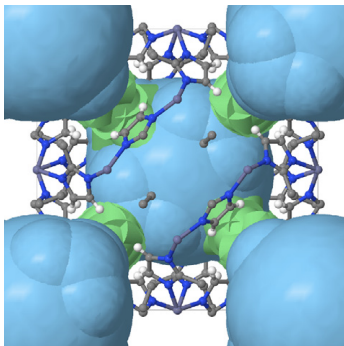
23 years later...



there are **534673** entries in the  
COD <https://crystallography.net>

# One of the first uses for the COD

„MOFomics: Computational pore characterization of metal–organic frameworks“ (First et al. 2013):

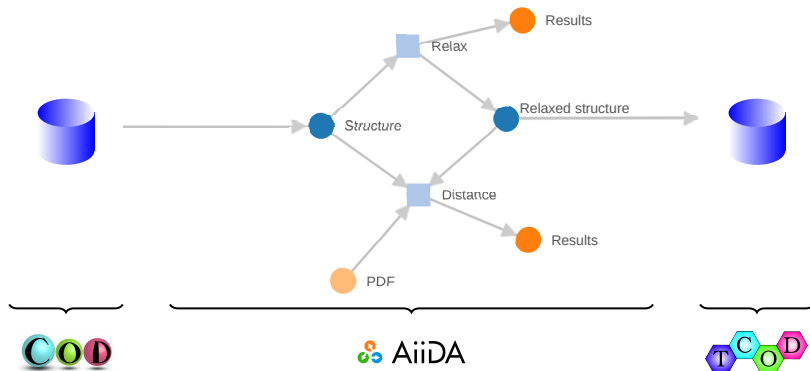


**Table 2**

Characterization results for selected structures. Volume and area are accessible to hydrogen (diameter 2.18 Å).

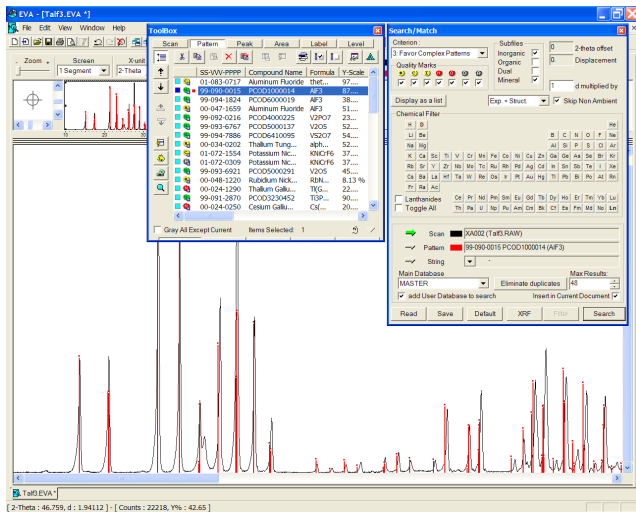
|                        | LCD (Å) | PLD (Å) | Volume (cm <sup>3</sup> /g) | Area (m <sup>2</sup> /g) |
|------------------------|---------|---------|-----------------------------|--------------------------|
| GWMOF-3 (COD 4300942)  | 5.4     | 3.7     | 0.226                       | 625                      |
| MIL-47 (CSD IDIWOH)    | 6.8     | 6.8     | 0.405                       | 1781                     |
| MIL-53 (COD 4103388)   | 7.1     | 7.1     | 0.500                       | 2203                     |
| MOF-5 (CSD SAHYIK)     | 15.0    | 7.8     | 1.186                       | 2297                     |
| MOF-501 (COD 4300890)  | 11.0    | 5.1     | 0.747                       | 2132                     |
| NOTT-401 (COD 7106668) | 7.6     | 4.1     | 0.279                       | 1080                     |
| ZIF-6 (CSD EQOCOC01)   | 9.5     | 6.7     | 0.749                       | 1076                     |
| ZIF-8 (CSD VELVOY)     | 11.4    | 3.4     | 0.485                       | 1531                     |
| Hyp. MOF 5082          | 9.8     | 4.0     | 0.850                       | 2320                     |
| Hyp. MOF 18075         | 7.2     | 4.7     | 1.060                       | 1435                     |
| Hyp. MOF 32532         | 5.9     | 3.2     | 0.345                       | 1416                     |

# The use of COD



• Computational materials science  
(Merkys et al. 2017; Pizzi 2018)

# The use of COD



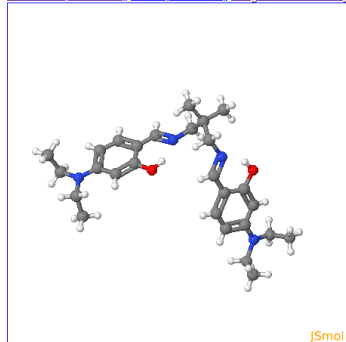
Use of PCOD and COD for materials identification

Courtesy Armel Le Bail  
(Le Bail 2008)

# The use of COD

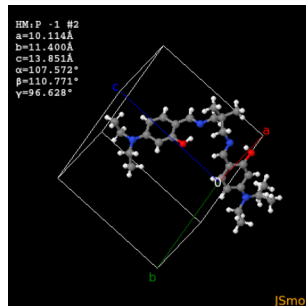
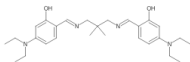
<https://molecules.crystallography.net/>

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



[SDF file](#) [CML file](#)

## 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](#)

# Searching for infinite nets

Building quotient graphs for covalent and coordination polymers.

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

Polymer Viewer Home | Search | Coordination Sequences

[Original COD entry](#) [Download CIF](#) [Download DOI](#)

### Structure 2310128

[2310113](#) << [2310128](#) >> [2310145](#)

**Citation:** Graf, D.L.; Bradley, W.F. (1962). The crystal structure of huntite, Mg<sub>3</sub> Ca (C O<sub>3</sub>)<sub>4</sub>. Acta Crystallographica, 15, pp. 238-242.

**Formula:** C<sub>4</sub> Ca Mg<sub>3</sub> O<sub>12</sub>    **Chemical name:** Trimagnesium Calcium Carbonate

**Cell parameters:** a = 6.075 Å b = 6.075 Å c = 6.075 Å α = 102.93° β = 102.93° γ = 102.93°    **List type:** noninterleaving    **List:** 3D

[Separate molecules](#) [Combined 3D view](#)

#### Molecules

CIF datablock 0, entity 1 of 1

**Basis:** (1,0,0 0,1,0 0,0,1)

**Dimensionality:** 3

**Formula:** C<sub>4</sub> Mg<sub>3</sub> O<sub>12</sub> (C<sub>4</sub> Mg<sub>3</sub> O<sub>12</sub>)

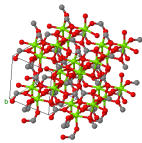
**Multiplicity:** 1

**Coordination sequences:**

- 2,7,10,30,31,75,62,142,110,224
- 1,4,19,22,49,44,113,92,189,130
- 3,6,27,24,51,54,129,92,189,144
- 2,11,13,34,33,65,71,148,111,241
- 2,11,15,30,35,69,75,156,115,245
- 6,8,20,22,64,54,114,90,198,146

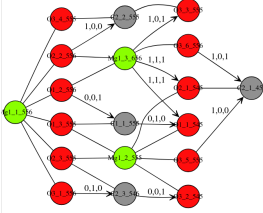
(2 2 2)

HM: P 1 E1=a,b,c  
a=6.075Å  
b=6.075Å  
c=6.075Å  
α=102.930°  
β=102.930°  
γ=102.930°



jsmol

☒ highlight edges



Work of Yaroslav Rozdobudko

Saulius Gražulis

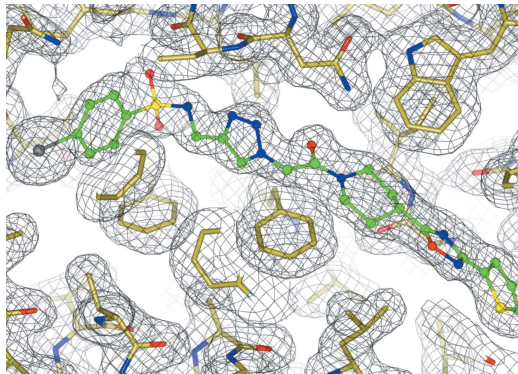
The impact of the Crystallography Open Database o

Calgary, 2026

10 / 26

# Restrain generation for macromolecular refinement

Statistics from the COD can serve as a basis for ligand geometry restrain generation: program AceDRG (Long et al. 2017a; Long et al. 2017b).

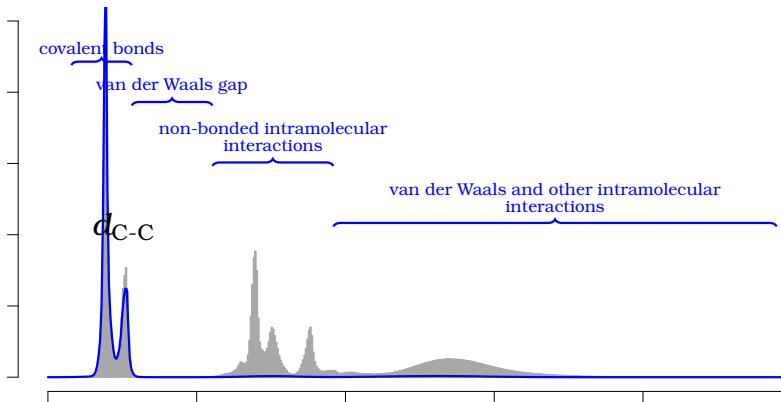


**Figure 6**

An example of a ligand refined using *AceDRG* restraints: PDB entry 2o8h with the ligand O8H (Willand *et al.*, 2010). Although the bond lengths are significantly different, the overlaid structures are visually almost identical. This figure was produced using *CCP4mg* (McNicholas *et al.*, 2011).

# Updated atomic radii from open sources

Example: carbon-carbon distances. Data presented with *full provenance*.



((Šidlauskaitė et al. 2026), in press.)



# The COD approach

## Linking tables

The COD SQL data schema + COD CIF dictionary;

data

| Field | Type            | Key |
|-------|-----------------|-----|
| file  | int(7) unsigned | PRI |
| a     | double unsigned | MUL |
| siga  | float unsigned  |     |
| b     | double unsigned | MUL |
| sigb  | float unsigned  |     |
| c     | double unsigned | MUL |
| sigc  | float unsigned  |     |
| ...   | ...             | ... |

XRDa\_x\_COD

| Field       | Type            | Key |
|-------------|-----------------|-----|
| id          | int(11)         | PRI |
| cod_id      | int(7) unsigned | MUL |
| ext_id      | int(11)         | MUL |
| relation_id | int(11)         | MUL |

relations

| Field       | Type         | Key |
|-------------|--------------|-----|
| id          | int(11)      | PRI |
| name        | varchar(255) |     |
| description | text         |     |

Similar schemas: Star schema:  
(Apeh 2024)



# Example of the COD data record raw data link

XRDa as an example

<https://www.crystallography.net/cod/3000497.html>

## 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

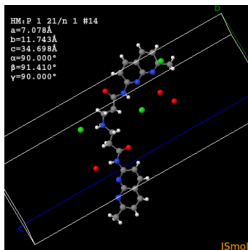
## Documentation

COD Wiki  
Obtaining COD  
License  
Privacy and GDPR  
Querying COD  
Citing COD  
COD Mirrors  
Advice to donators  
Useful links

## Information card for entry 3000497

[3000496](#) << [3000497](#) >> [3000498](#)

## Preview



Coordinates [3000497.cif](#)  
Original paper (by DOI) [HTML](#)  
External links [XRDa](#)

## ▼ Structure parameters

|                        |                                                                                                             |
|------------------------|-------------------------------------------------------------------------------------------------------------|
| Formula                | C24 H26 Cl3 N7 O5                                                                                           |
| Calculated formula     | C24 H26 Cl3 N7 O5                                                                                           |
| Authors of publication | Kawamoto, Akihiro; Kojima Chojiro; Kurisu, Genji; Nakane, Takanori; Nakatani Kazuhiko; Sakurabayashi Shuhei |
| Journal of publication | Journal of the American Chemical Society                                                                    |
| Year of publication    | 2025                                                                                                        |
| a                      | 7.078 Å                                                                                                     |
| b                      | 11.743 Å                                                                                                    |
| c                      | 34.698 Å                                                                                                    |
| α                      | 90°                                                                                                         |
| β                      | 91.41°                                                                                                      |
| γ                      | 90°                                                                                                         |

(Sakurabayashi et al. 2025)

# XRDa side

Links to the COD :)

<https://xrda.pdbj.org/entry/335>

**XRDa**  
Xtal Raw Data Archive

English 日本語

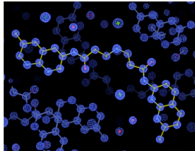
 Login using ORCID

[Help](#) [Browse](#) [Statistics](#)



### Entry information

|                                      |                                                                                                                                                                                                                                                                                                                                                                                                                         |
|--------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Entry ID:                            | XRD-00335                                                                                                                                                                                                                                                                                                                                                                                                               |
| Entry title:                         | MicroED dataset of ND                                                                                                                                                                                                                                                                                                                                                                                                   |
| Entry type:                          | Electron Diffraction                                                                                                                                                                                                                                                                                                                                                                                                    |
| XRD entry authors:                   | <a href="#">Shuhei Sakurabayashi</a> ;<br><a href="#">Kyoko Furuita</a> ;<br><a href="#">Takeshi Yamada</a> ;<br>Noriaki Sugiura;<br>Makoto Nomura;<br><a href="#">Takanori Nakane</a> ;<br><a href="#">Akihiro Kawamoto</a> ;<br><a href="#">Genji Kurisu</a> ;<br><a href="#">Yohei Miyanoiri</a> ;<br><a href="#">Toshimichi Fujiwara</a> ;<br><a href="#">Kazuhiro Nakatani</a> ;<br><a href="#">Chojiro Kojima</a> |
| Entry:                               | <a href="#">Download</a>  (141.85 GB)                                                                                                                                                                                                                                                                                                |
| XRDa entry DOI:                      | <a href="#">10.51093/xrd-00335</a>                                                                                                                                                                                                                                                                                                   |
| CCDC entry:                          | <a href="#">2350096</a>                                                                                                                                                                                                                                                                                                               |
| Crystallography Open Database entry: | <a href="#">3000497</a>                                                                                                                                                                                                                                                                                                               |
| Citation DOI:                        | <a href="#">10.1021/jacs.4c17538</a>                                                                                                                                                                                                                                                                                                 |



# A side note: links with other data sources

## Pubchem as an example

pubchem\_x\_cod

| pubchem_x_cod |           |         |
|---------------|-----------|---------|
| id            | ext_id    | cod_id  |
| 120396        | 101539687 | 7016989 |
| 99284         | 168336540 | 4135255 |
| 18081         | 168304792 | 1554891 |
| 147524        | 25198875  | 7155506 |
| 134464        | 168350383 | 7104999 |

**169593** rows  
(Vaitkus et al. 2023)



## 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

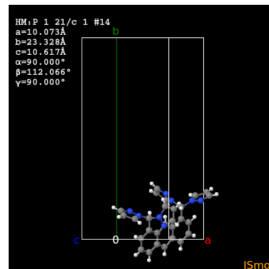
### Documentation

COD Wiki  
Obtaining COD  
License  
Privacy and GDPR  
Querying COD  
Citing COD  
COD Mirrors  
Advice to donators  
Useful links

## Information card for entry 7016989

[7016988](#) << **7016989** >> [7016990](#)

### Preview



Coordinates

[7016989.cif](#)

Original paper (by DOI)

[HTML](#)

External links

[PubChem](#)

# A side note: links with other data sources

## Pubchem as an example

pubchem\_x\_cod

| id     | ext_id    | cod_id  |
|--------|-----------|---------|
| 120396 | 101539687 | 7016989 |
| 99284  | 168336540 | 4135255 |
| 18081  | 168304792 | 1554891 |
| 147524 | 25198875  | 7155506 |
| 134464 | 168350383 | 7104999 |

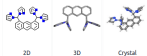
**169 593** rows  
(Vaitkus et al. 2023)

NIH National Library of Medicine  
National Center for Biotechnology Information

PubChem About Docs Submit Contact

COMPOUND SUMMARY

1-[[8-[Di(pyrazol-1-yl)methyl]anthracen-1-yl]-pyrazol-1-ylmethyl]pyrazole

|                   |                                                                                                     |
|-------------------|-----------------------------------------------------------------------------------------------------|
| PubChem CID       | 101539687                                                                                           |
| Structure         | <br>2D 3D Crystal |
| Molecular Formula | C <sub>28</sub> H <sub>22</sub> N <sub>6</sub>                                                      |
| Molecular Weight  | 470.5 g/mol<br><small>Computed by PubChem 2.2 (PubChem release 2025.04.14)</small>                  |
| Dates             | Create: 2015-12-18<br>Modify: 2025-08-23                                                            |

# A side note: links with other data sources

## Wikidata as an example

wikidata\_x\_cod

| id  | ext_id    | cod_id  |
|-----|-----------|---------|
| 145 | Q423494   | 9001039 |
| 2   | Q1067912  | 9017917 |
| 150 | Q182264   | 9001030 |
| 14  | Q3649812  | 9000094 |
| 111 | Q19861259 | 2105051 |

76 rows



## 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

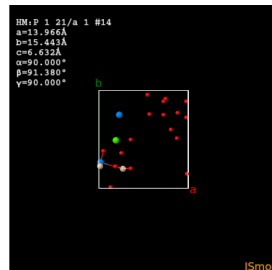
### Documentation

COD Wiki  
Obtaining COD  
License  
Privacy and GDPR  
Querying COD  
Citing COD  
COD Mirrors  
Advice to donors  
Useful links

## Information card for entry 9001039

[9001038](#) << **9001039** >> [9001040](#)

### Preview



Coordinates [9001039.cif](#)

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

# A side note: links with other data sources

## Wikidata as an example

wikidata\_x\_cod

| id  | ext_id    | cod_id  |
|-----|-----------|---------|
| 145 | Q423494   | 9001039 |
| 2   | Q1067912  | 9017917 |
| 150 | Q182264   | 9001030 |
| 14  | Q3649812  | 9000094 |
| 111 | Q19861259 | 2105051 |

**76** rows

subclass of

nesosilicates

▼ 0 references

edit

+ add reference

+ add value

image



Uranophane.jpg

1,400 × 829; 1.01 MB

media legend

Mineral de la col·lecció del departament de geologia de la Brigham Young University, Utah (Catalan)

► 1 reference

+ add value

named after

uranium

► 1 reference

edit

+ add value

Navigation icons: back, forward, search, etc.

Saulius Gražulis

The impact of the Crystallography Open Database o

Calgary, 2026

17 / 26

# Relation types

Different relations for different records

Relation “ontology”: the `relations` table;

| +-----+-----+-----+ ... |                    |                                               |
|-------------------------|--------------------|-----------------------------------------------|
| id                      | name               | description ...                               |
| +-----+-----+-----+ ... |                    |                                               |
| 1                       | RELATED            | THE CIF is related in some way to the ext ... |
| 2                       | DIRECTLY_DERIVED   | The CIF file was generated directly from ...  |
| 3                       | INDIRECTLY_DERIVED | Related by name, but not confirmed as the ... |
| 4                       | SIMILAR_TO         | It shares some characteristics but wasn't ... |
| +-----+-----+-----+ ... |                    |                                               |

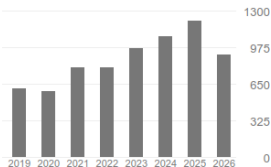


## Citations according to Google Scholar

(<https://scholar.google.lt/citations?user=BswX10sAAAAJ&hl=en>, 2026-08-08 21:00 UTC):

| TITLE                                                                                                                                                                                                                                                   | CITED BY | YEAR |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|------|
| <b>Crystallography Open Database—an open-access collection of crystal structures</b><br>S Gražulis, D Chateigner, RT Downs, AFT Yokochi, M Quirós, L Lutterotti, ...<br>Applied Crystallography 42 (4), 726-729                                         | 2349     | 2009 |
| <b>Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration</b><br>S Gražulis, A Daškevič, A Merkys, D Chateigner, L Lutterotti, M Quirós, ...<br>Nucleic Acids Res 40, D420-D427 | 1710     | 2012 |
| <b>AceDRG: a stereochemical description generator for ligands</b><br>F Long, RA Nicholls, P Emsley, S Gražulis, A Merkys, A Vaitkus, ...<br>Biological Crystallography 73 (2), 112-122                                                                  | 511      | 2017 |
| <b>Validation of the crystallography open database using the crystallographic information framework</b><br>A Vaitkus, A Merkys, S Gražulis<br>Applied Crystallography 54 (2), 661-672                                                                   | 479      | 2021 |
| <b>Using SMILES strings for the description of chemical connectivity in the Crystallography Open</b>                                                                                                                                                    | 350      | 2018 |

|           |      |      |
|-----------|------|------|
| Citations | 9612 | 5785 |
| h-index   | 39   | 24   |
| i10-index | 65   | 42   |



Public access [VIEW ALL](#)

8 articles 38 articles

# New ways of publishing science

Suggesting a new type of scholarly publication:

Peer reviewed database record!

COD data can be **cited**:

- Referencing machine readable data:

Srivastava, R. C.; Klooster, W. T.; Koetzle, T. F. "Neutron Structures of Ammonium Fluoroberyllate" (1999) *The Crystallography Open Database*, rev. 176759, the COD Advisory Board (eds.), <https://www.crystallography.net/cod/2002926.cif@176759>. [Retrieved 2016-09-21 16:48 EEST]

- Referencing a human-readable page:

Srivastava, R. C.; Klooster, W. T.; Koetzle, T. F. "Neutron Structures of Ammonium Fluoroberyllate" (1999) *The Crystallography Open Database*, rev. 176759, the COD Advisory Board (eds.), <https://www.crystallography.net/cod/2002926>. [Retrieved 2016-09-21 16:48 EEST]

# The summing up

- Open data and software allows us to provide *full* provenance to our calculations;
- Open data drive scientific innovation;
- Open data can make linked knowledge networks;
- Open data can increase awareness of general public about science, demonstrate justifications;

# Acknowledgements

## **VU LSC IBT (KICIS)**

Anastasija Kudrevceva  
Andrius Merkys  
Antanas Vaitkus  
Algirdas Grybauskas

## **VU MIF II (FMG)**

Linus Laibinis  
Karolis Petrauskas  
Irus Grinis  
Haroldas Giedra

## **QM community**

Linus Vilčiauskas  
Vytautas Žalandauskas  
Lukas Razinkovas  
Nicola Marzari  
Giovanni Pizzi  
Matthew Evans  
Rickard Armiento

## **COD Advisory board**

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

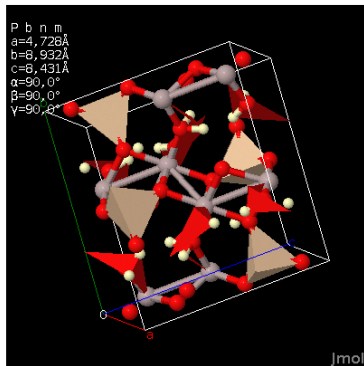
## **Funding:**

Data Center for Machine Learning and Quantum Computing in Natural and Biomedical Sciences, 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>

# References I

- Apeh, Kenneth (2024). *What is a Star Schema in Power BI*. URL:  
<https://fp20analytics.com/wp-content/uploads/2024/10/What-is-a-Star-Schema-in-Power-BI.pdf>.
- First, Eric L. et al. (2013). “MOFomics: Computational pore characterization of metal–organic frameworks”. In: *Microporous and Mesoporous Materials* 165, pp. 32–39. issn: 1387-1811. doi:  
10.1016/j.micromeso.2012.07.049. URL:  
<http://www.sciencedirect.com/science/article/pii/S1387181112004817>.
- 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>.
- Le Bail, Armel (2008). “Frontiers Between Crystal-structure Prediction and Determination by Powder Diffraction”. In: *Powder Diffraction Suppl.*, S5–S12. doi: 10.1154/1.2903488.
- Long, Fei et al. (Feb. 2017a). “ACEDRG: A stereo-chemical description generator for ligands”. In: *Acta Crystallographica Section D* 73.2, pp. 112–122. doi: 10.1107/S2059798317000067.
- (Feb. 2017b). “Validation and extraction of stereochemical information from small molecular databases”. In: *Acta Crystallographica Section D* 73.2, pp. 103–111. doi: 10.1107/S2059798317000079.

# References II

- Merkys, Andrius et al. (Nov. 15, 2017). “A posteriori metadata from automated provenance tracking: Integration of AiiDA and TCOD”. In: *Journal of Cheminformatics* 9.1, p. 56. doi: 10.1186/s13321-017-0242-y. arXiv: 1706.08704v3 [cond-mat.mtrl-sci]. URL: <https://jcheminf.springeropen.com/articles/10.1186/s13321-017-0242-y>.
- Pizzi, Giovanni (2018). “Open-Science Platform for Computational Materials Science: AiiDA and the Materials Cloud”. In: *Handbook of Materials Modeling*. Springer International Publishing, p. 1. doi: 10.1007/978-3-319-42913-7\_64-1.
- Sakurabayashi, Shuhei et al. (Apr. 2025). “NMR-based rational drug design of G:G mismatch DNA binding ligand trapping transient complex via disruption of a key allosteric interaction”. In: *Journal of the American Chemical Society*, p. 4c17538. issn: 1520-5126. doi: 10.1021/jacs.4c17538.
- Senior, Andrew W. et al. (Jan. 2020). “Improved protein structure prediction using potentials from deep learning”. In: *Nature* 577.7792, pp. 706–710. doi: 10.1038/s41586-019-1923-7. eprint: <https://doi.org/10.1038/s41586-019-1923-7>.
- Shaffer, P. A. et al. (Nov. 1946). “The use of punched cards in molecular structure determinations I. Crystal structure calculations”. In: *The Journal of Chemical Physics* 14.11, pp. 648–658. doi: 10.1063/1.1724081. URL: <https://authors.library.caltech.edu/45530/1/SCHjcp46a.pdf>.
- Šidlauskaitė, Eglė et al. (2026). “Determination of atomic radii from small-molecule crystal structures”. In: *Journal of Cheminformatics*. in press.

# References III

- 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.
- Wilkinson, Mark D. et al. (Mar. 2016). “The FAIR guiding principles for scientific data management and stewardship”. In: *Scientific Data* 3.1. doi: 10.1038/sdata.2016.18.