Atvira kristalografinė duomenų bazė COD

kūrimas ir pritaikymai

Saulius Gražulis ir COD komanda

GMC seminaras
Vilnius, 2019

Vilniaus universitetas, GMC

Biotechnologijos institutas

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Turinys

1. Motyvacija – kodėl duomenų bazės?
2. COD kūrimo istorija
3. COD turinys
4. COD pritaikymai
5. COD nauda mums :)
6. Pamąstymai apie atvirus duomenis
7. Tolimesni planai

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Turinys

1. Motyvacija – kodėl duomenų bazės?
2. COD kūrimo istorija
3. COD turinys
4. COD pritaikymai
5. COD nauda mums :)
6. Pamąstymai apie atvirus duomenis
7. Tolimesni planai

This project has received funding from the European Union's Horizon 2020
research and innovation program under grant agreement No 689868.

Kur rasti reikalingą informaciją?

Klausimas: kaip suskaičiuoti vidurkį ir standartinį nuokrypį duomenų sraute?

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Kur rasti reikalingą informaciją?

Klausimas: kaip suskaičiuoti vidurkį ir standartinį nuokrypį duomenų sraute?

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \, ; \quad s_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\overline{x}_n - x_i)^2}$$

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Kur rasti reikalingą informaciją?

Klausimas: kaip suskaičiuoti vidurkį ir standartinį nuokrypį duomenų sraute?

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i\,; \quad s_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\overline{x}_n - x_i)^2}$$

Per daug RAM!

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Kur rasti reikalingą informaciją?

Klausimas: kaip suskaičiuoti vidurkį ir standartinį nuokrypį duomenų sraute?

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i\,;\quad s_n = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right)}$$

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Kur rasti reikalingą informaciją?

Klausimas: kaip suskaičiuoti vidurkį ir standartinį nuokrypį duomenų sraute?

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i\,; \quad s_n = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right)}$$

Netikslu!

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Kur rasti reikalingą informaciją?

Klausimas: kaip suskaičiuoti vidurkį ir standartinį nuokrypį duomenų sraute?

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i\,;\;\; s_n = \ldots ?$$

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Kur rasti reikalingą informaciją?

Klausimas: kaip suskaičiuoti vidurkį ir standartinį nuokrypį duomenų sraute?

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \,;\; S_n = S_{n-1} + \frac{n}{n-1}\left(\overline{x}_n - x_n\right)^2 \;;\; s_n = \sqrt{\frac{1}{n-1}S_n}$$

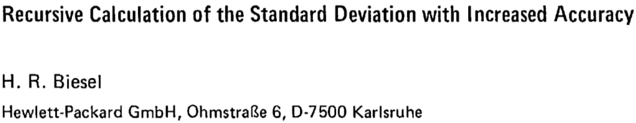
OK :)

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Kur rasti reikalingą informaciją?

Klausimas: kaip suskaičiuoti vidurkį ir standartinį nuokrypį duomenų sraute?

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \; ; \; S_n = S_{n-1} + \frac{n}{n-1}\left(\overline{x}_n - x_n\right)^2 \; ; \; s_n = \sqrt{\frac{1}{n-1} S_n}$$



Recursive Calculation of the Standard Deviation with Increased Accuracy

H. R. Biesel

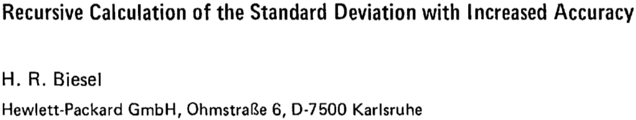
Hewlett-Packard GmbH, Ohmstraße 6, D-7500 Karlsruhe

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Kur rasti reikalingą informaciją?

Klausimas: kaip suskaičiuoti vidurkį ir standartinį nuokrypį duomenų sraute?

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \, ; \; S_n = S_{n-1} + \frac{n}{n-1}\left(\overline{x}_n - x_n\right)^2 \; ; \; s_n = \sqrt{\frac{1}{n-1} S_n}$$



Recursive Calculation of the Standard Deviation with Increased Accuracy

H. R. Biesel

Hewlett-Packard GmbH, Ohmstraße 6, D-7500 Karlsruhe

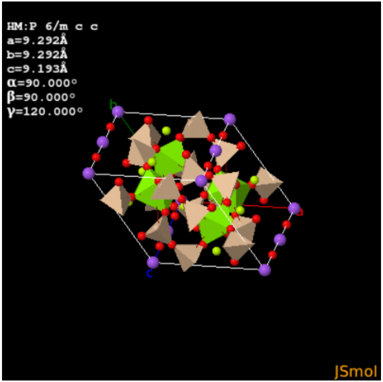
Chromatographia, Vol. 10, No. 4, April 1977, p. 173

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

A question to answer

http://en.wikipedia.org/wiki/Emerald

$\overset{?}{\Rightarrow}$



http://www.crystallography.net/5000095.html

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Sprendimas – skaitmeniniai kompiuteriai

Stefan Kögl [CC BY-SA 3.0]

Dave McGuire [Public Domain]

This project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 689868.

Turinys

1. Motyvacija – kodėl duomenų bazės?
2. COD kūrimo istorija
3. COD turinys
4. COD pritaikymai
5. COD nauda mums :)
6. Pamąstymai apie atvirus duomenis
7. Tolimesni planai

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

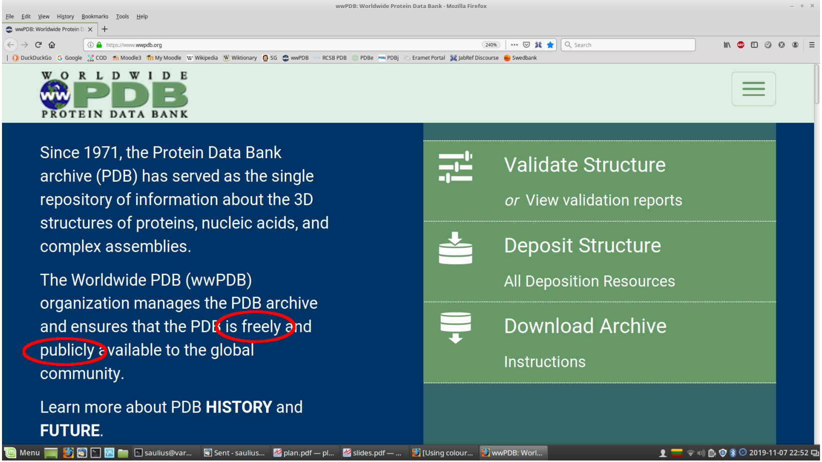
Data Sharing in Crystallography

Started quite early

- **1948 Acta Cryst. (IUCr)** The Acta Crystallographica journal was launched, *all coordinates were printed in journal articles, and Acta Crystallographica published the structure factors as well*
- **1965 CSD (CCDC)** *The CCDC was established at the Department of Chemistry, Cambridge University /. . . / about 2000 structures published before 1965 were gradually incorporated into the developing database*
- **1971 PDB** *In June 1971, the two communities attended the Cold Spring Harbor Symposium on Quantitative Biology (Cold Spring Laboratory Press, 1972)*

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Su PDB viskas gerai


wwPDB: Worldwide Protein Data Bank - Mozilla Firefox
File Edit View History Bookmarks Tools Help
wwPDB: Worldwide Protein D
https://www.wwpdb.org
240%
Search
DuckDuckGo Google COD Moodle3 My Moodle Wikipedia Wiktionary SG wwPDB RCSB PDB PDBe PDBj Eramet Portal JabRef Discourse Swedbank
WORLDWIDE PDB PROTEIN DATA BANK
Menu saulius@var... Sent - saulius... plan.pdf — pl... slides.pdf — ... [Using colour... wwPDB: Worl...
2019-11-07 22:52

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

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

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://summary.ccdc.cam.ac.uk/about-this-service, last accessed 2016-11-24)

- “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]

This project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 689868.

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.

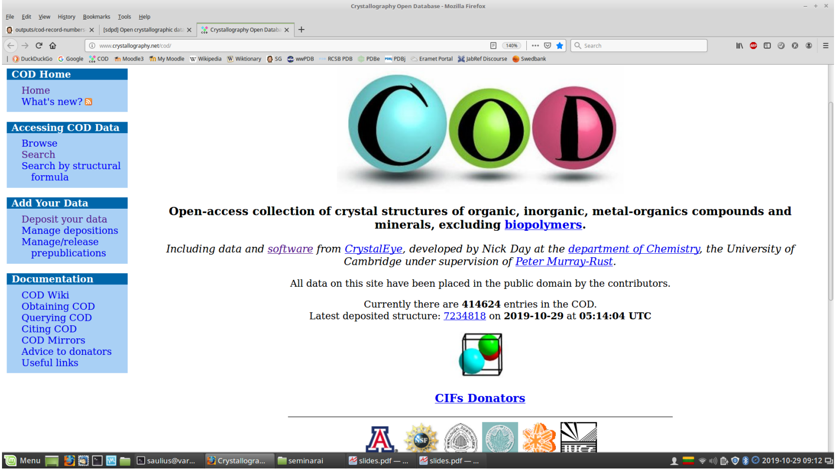
gemstonede (Dr. Michael BERNDT) Fri Feb 14, 2003 1:26 pm

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

16 years later … :)

The Crystallography Open Database

http://www.crystallography.net/cod



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
Advice to donators
Useful links

Open-access collection of crystal structures of organic, inorganic, metal-organics 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.

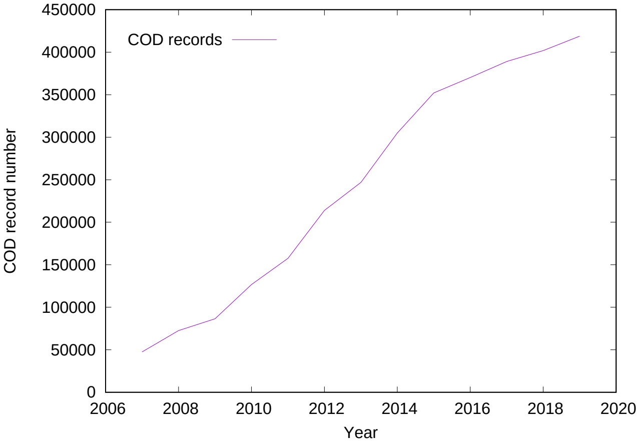
Currently there are **414624** entries in the COD.
Latest deposited structure: 7234818 on **2019-10-29** at **05:14:04 UTC**

CIFs Donators

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

COD persistence

COD is on-line for 16 years, increased 8-fold over the last 10 years; currently contains over 410 000 records (2019):



This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Turinys

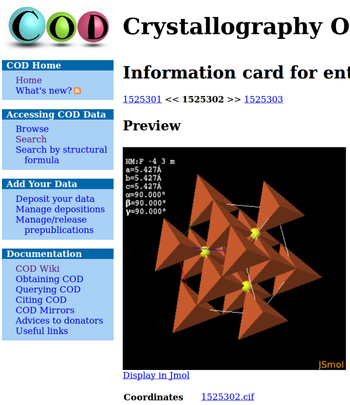
1. Motyvacija – kodėl duomenų bazės?
2. COD kūrimo istorija
3. COD turinys
4. COD pritaikymai
5. COD nauda mums :)
6. Pamąstymai apie atvirus duomenis
7. Tolimesni planai

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

A COD crystal structure page example

Sphalerite

http://www.crystallography.net/cod/1525302.html



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 donators
- Useful links

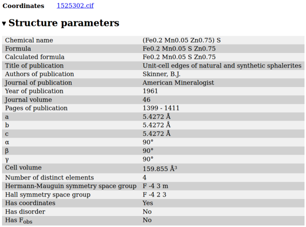
Information card for entry 1525302

1525301 << **1525302** >> 1525303

Preview

Display in Jmol

Coordinates 1525302.cif



Coordinates 1525302.cif

▾ Structure parameters

Chemical name	(Fe0.2 Mn0.05 Zn0.75) S
Formula	Fe0.2 Mn0.05 S Zn0.75
Calculated formula	Fe0.2 Mn0.05 S Zn0.75
Title of publication	Unit-cell edges of natural and synthetic sphalerites
Authors of publication	Skinner, B.J.
Journal of publication	American Mineralogist
Year of publication	1961
Journal volume	46
Pages of publication	1399 - 1411
a	5.4272 Å
b	5.4272 Å
c	5.4272 Å
α	90°
β	90°
γ	90°
Cell volume	159.855 Å^3
Number of distinct elements	4
Hermann-Mauguin symmetry space group	F -4 3 m
Hall symmetry space group	F -4 2 3
Has coordinates	Yes
Has disorder	No
Has F_{obs}	No

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

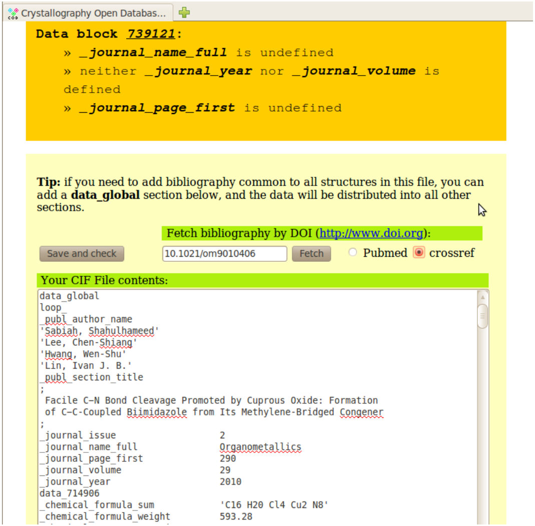
COD data validation

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
   ```

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

COD validavimo ir deponavimo svetainė



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):

10.1021/om9010406

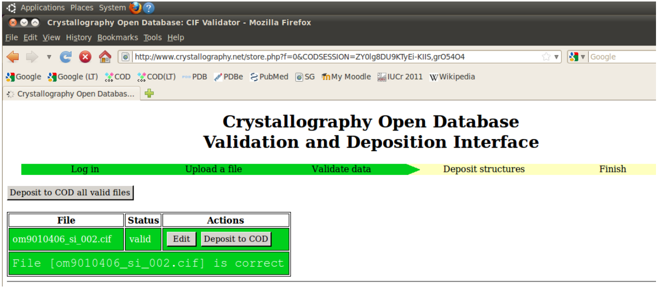
○ Pubmed ● crossref

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 Biimidazole 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
```

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

COD validavimo ir deponavimo svetainė


Applications Places System
Crystallography Open Database: CIF Validator - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://www.crystallography.net/store.php?f=0&CODSESSION=ZY0lg8DU9KTyEi-KIIS,grO54O4
Google
Google Google (LT) COD COD(LT) PDB PDBe PubMed SG My Moodle IUCr 2011 Wikipedia
Crystallography Open Databas...

Crystallography Open Database Validation and Deposition Interface

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

Deposit to COD all valid files

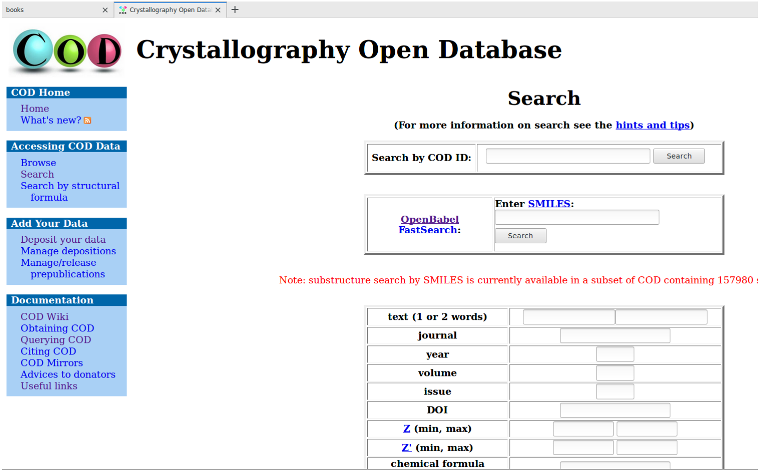
File	Status	Actions
om9010406_si_002.cif	valid	Edit Deposit to COD

```
File [om9010406_si_002.cif] is correct
```

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

COD Search Form

Data can be queried on-line using basic crystallographic parameters or metadata (http://www.crystallography.net/cod/search.html)



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 donators
- Useful links

Search

(For more information on search see the hints and tips)

Search by COD ID:	Search

OpenBabel FastSearch:	Enter SMILES: Search

Note: substructure search by SMILES is currently available in a subset of COD containing 157980

text (1 or 2 words)	
journal	
year	
volume	
issue	
DOI	
Z (min, max)	
Z' (min, max)	
chemical formula	

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

COD query examples

Web, REST, SQL

- Via the WWW interface – go for “search” in:
  - http://www.crystallography.net/cod
  - http://solsa.crystallography.net/rod
  - http://solsa.crystallography.net/hod
- Via the **stable** URLs (REST):
  - http://www.crystallography.net/cod/2000000.cif
  - http://solsa.crystallography.net/rod/3500021.rod
  - http://solsa.crystallography.net/rod/3500021.html
  - http://www.crystallography.net/cod/result?text=perovskite
- Via the **views** of the SQL database:

```
mysql -u cod_reader cod -h www.crystallography.net\
    -e 'select file, a, b, c, vol, formula
        from data where
            year between 2013 and
                         2014 and
            formula regexp " C[0-9]* "
            order by vol desc limit 10'
```

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

COD accessibility

COD is a **fully open-access database**. All records are available under public domain designation.

Provided access methods are:

- Web search
- URLs constructed from stable identifiers
- RESTful interfaces
- Full data download

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

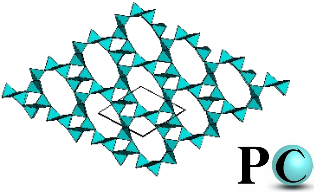
Open Crystallographic Databases

COD. TCOD. PCOD. MPOD. ...

http://www.crystallography.net/cod
> 410 000 entries (ready to grow > 10^6?)

http://www.crystallography.net/tcod
> 2000 entries (ready to grow to > 350 000?)

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



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

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

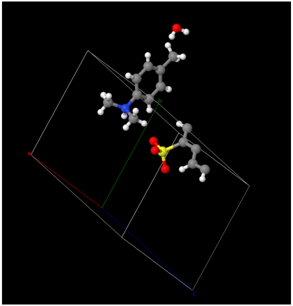
Turinys

1. Motyvacija – kodėl duomenų bazės?
2. COD kūrimo istorija
3. COD turinys
4. COD pritaikymai
5. COD nauda mums :)
6. Pamąstymai apie atvirus duomenis
7. Tolimesni planai

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Kristalografinio failo turinys

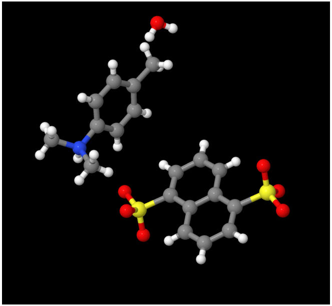
http://www.crystallography.net/cod/2231955.html



This project has received funding from the European Union's Horizon 2020
research and innovation program under grant agreement No 689868.

Kristalografinio failo turinys

http://www.crystallography.net/cod/2231955.html

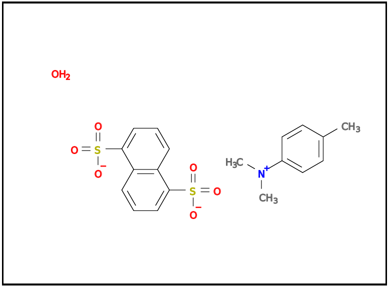


This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

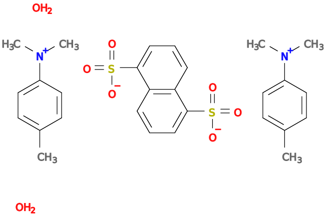
COD molekulių atstatymas

http://www.crystallography.net/cod/2231955.html

Įprasti algoritmai:



Naujas algoritmas:

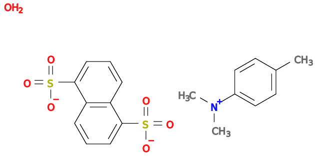


This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

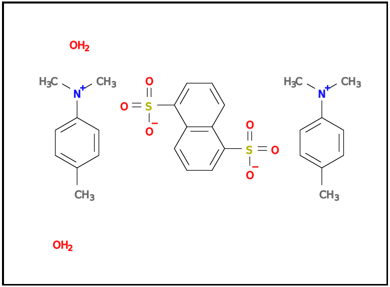
COD molekulių atstatymas

http://www.crystallography.net/cod/2231955.html

Įprasti algoritmai:



Naujas algoritmas:



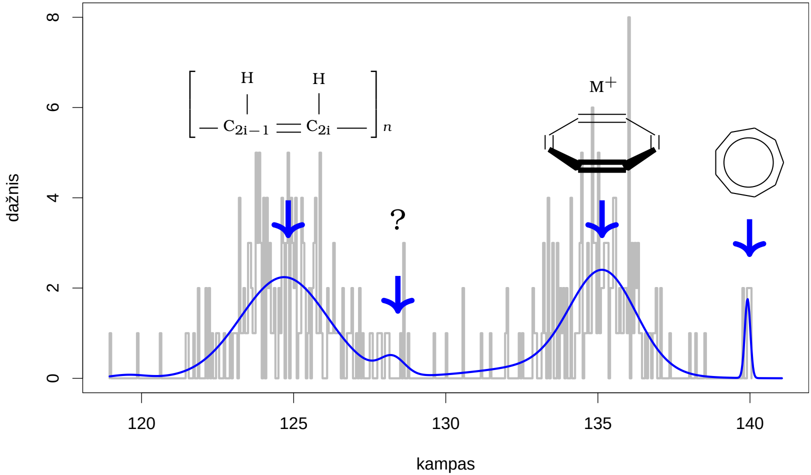
[Gražulis et al., 2015] 10.1107/S1600576714025904

A. Vaitkus, cif2molecule + OpenChemLib

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Recognition of complex features

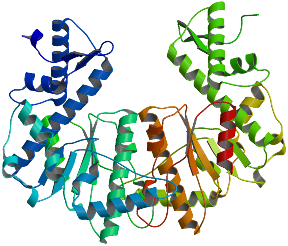
Bond angles of a polyene chain from COD



results of A. Merkys, PhD thesis

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Macromolecular structure refinement needs



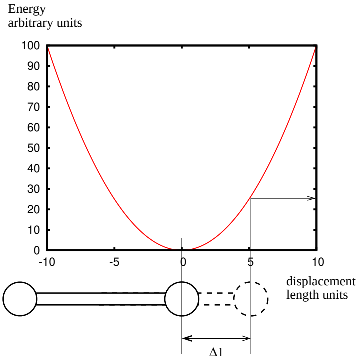
N_{atom}	N_{param}	N_{obs}	$\frac{N_{obs}}{N_{param}}$
PDB ID **1KNV**			
5070	20280	42686	**2.1**
COD ID **2002915**[1]			
38	342	10189	**30**

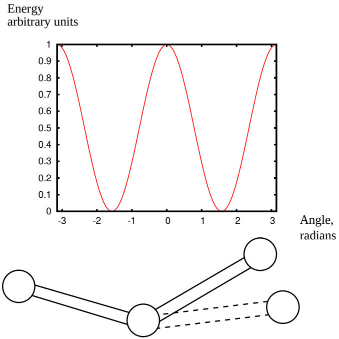
- When refining macromolecular structures, we have unfavourable parameter to observation ratio (we would like to have >5. . .)

[1] Dahaoui et al., Acta Cryst. B (1999)

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Solution – restrains



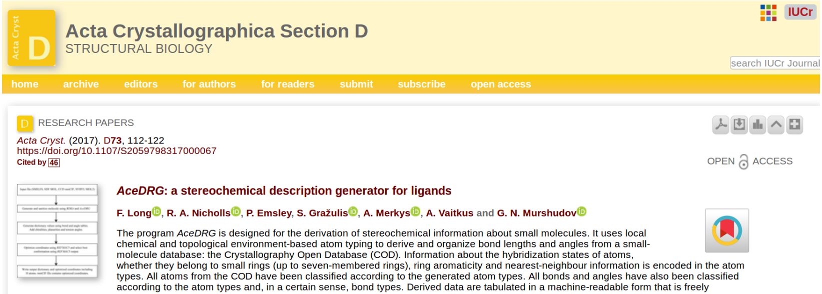


$$E = \frac{1}{2}k\Delta l^2 \quad p(\Delta l) \sim e^{-\frac{E}{k_B T}} = e^{-\frac{\Delta l^2}{2\sigma^2}} \qquad E = k(1+\cos(n\Delta\varphi)), \text{if}\, n > 0$$

- Use prior knowledge about bond lengths, angles, dihedrals, VdW radii, obtained from high-resolution organic molecule crystal structures.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Naujos apribojimų bibliotekos



Acta Crystallographica Section D
STRUCTURAL BIOLOGY

RESEARCH PAPERS

Acta Cryst. (2017). D**73**, 112-122
https://doi.org/10.1107/S2059798317000067
Cited by 46

OPEN ACCESS

AceDRG: a stereochemical description generator for ligands

F. Long, R. A. Nicholls, P. Emsley, S. Gražulis, A. Merkys, A. Vaitkus and G. N. Murshudov

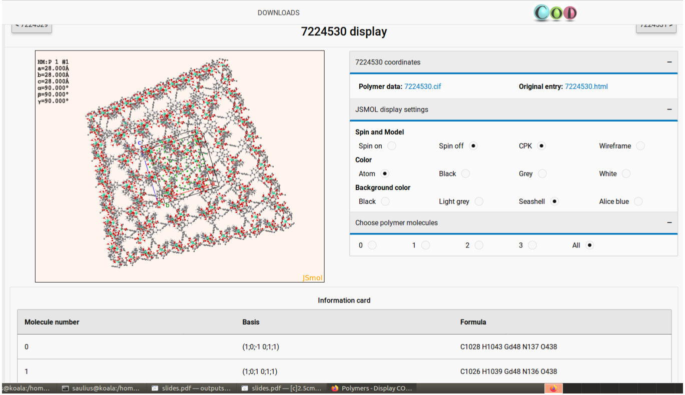
The program *AceDRG* is designed for the derivation of stereochemical information about small molecules. It uses local chemical and topological environment-based atom typing to derive and organize bond lengths and angles from a small-molecule database: the Crystallography Open Database (COD). Information about the hybridization states of atoms, whether they belong to small rings (up to seven-membered rings), ring aromaticity and nearest-neighbour information is encoded in the atom types. All atoms from the COD have been classified according to the generated atom types. All bonds and angles have also been classified according to the atom types and, in a certain sense, bond types. Derived data are tabulated in a machine-readable form that is freely

[Long et al., 2017a, Long et al., 2017b]

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

COD data applications: polymer search

- polymers-in-COD: ≈ 400 000 COD records processed
- polymers of different dimensionality (1D, 2D, 3D, 1D-2D and so on) detected, ≈ 93 000 polymer records in total.



7224530 display

7224530 coordinates −

Polymer data: 7224530.cif **Original entry:** 7224530.html

JSMOL display settings −

Spin and Model

Spin on ○ Spin off ● CPK ● Wireframe ○

Color

Atom ● Black ○ Grey ○ White ○

Background color

Black ○ Light grey ○ Seashell ● Alice blue ○

Choose polymer molecules −

0 ○ 1 ○ 2 ○ 3 ○ All ●

Information card

Molecule number	Basis	Formula
0	(1;0;-1 0;1;1)	C1028 H1043 Gd48 N137 O438
1	(1;0;1 0;1;1)	C1026 H1039 Gd48 N136 O438

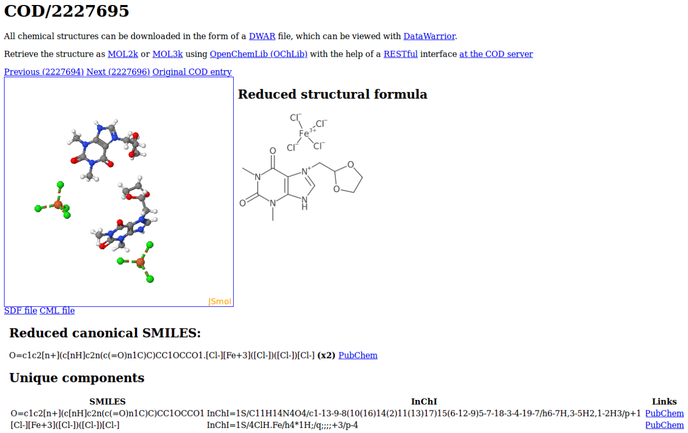
http://crystallography.net/cod/7224530.html

results of A. Belova

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

COD data applications: chemical comprehension

- molecules-in-COD: ≈ 380 000 COD records processed
- chemical structure derived automatically for ≈ 200 000 COD records.



COD/2227695

All chemical structures can be downloaded in the form of a DWAR file, which can be viewed with DataWarrior.

Retrieve the structure as MOL2k or MOL3k using OpenChemLib (OChLib) with the help of a RESTful interface at the COD server

Previous (2227694) Next (2227696) Original COD entry

SDF file CML file

Reduced structural formula

Reduced canonical SMILES:

O=c1c2[n+](c[nH]c2n(c(=O)n1C)C)CC1OCCO1.[Cl-][Fe+3]([Cl-])([Cl-])[Cl-] **(x2)** PubChem

Unique components

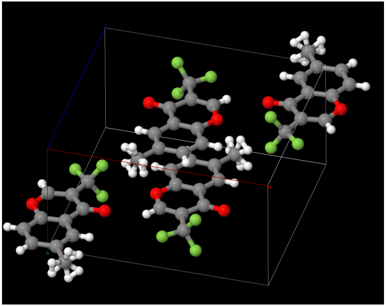
SMILES	InChI	Links
O=c1c2[n+](c[nH]c2n(c(=O)n1C)C)CC1OCCO1	InChI=1S/C11H14N4O4/c1-13-9-8(10(16)14(2)11(13)17)15(6-12-9)5-7-18-3-4-19-7/h6-7H,3-5H2,1-2H3/p+1	PubChem
[Cl-][Fe+3]([Cl-])([Cl-])[Cl-]	InChI=1S/4ClH.Fe/h4*1H;/q;;;;+3/p-4	PubChem

http://crystallography.net/cod/2227695.html results of A. Vaitkus

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Netvarka aplink specialiąją poziciją

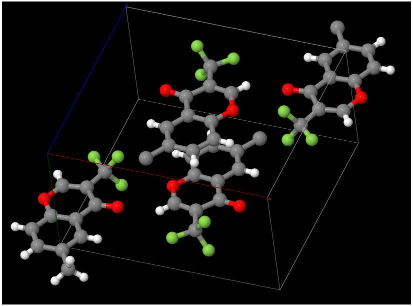
https://www.crystallography.net/cod/1544968.html



This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Netvarka aplink specialiąją poziciją

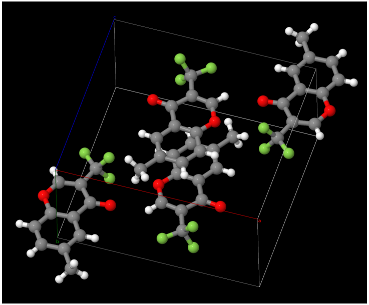
https://www.crystallography.net/cod/1544968.html



This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

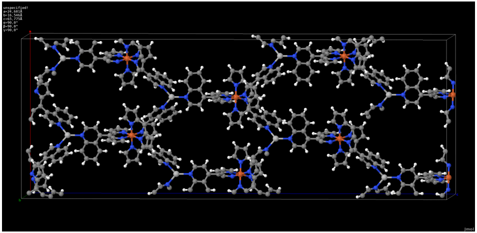
Netvarka aplink specialiąją poziciją

https://www.crystallography.net/cod/1544968.html

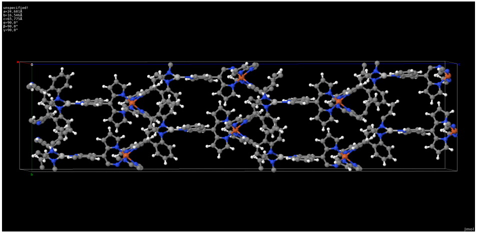


This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Netvarka aplink specialiąją poziciją polimeruose



video

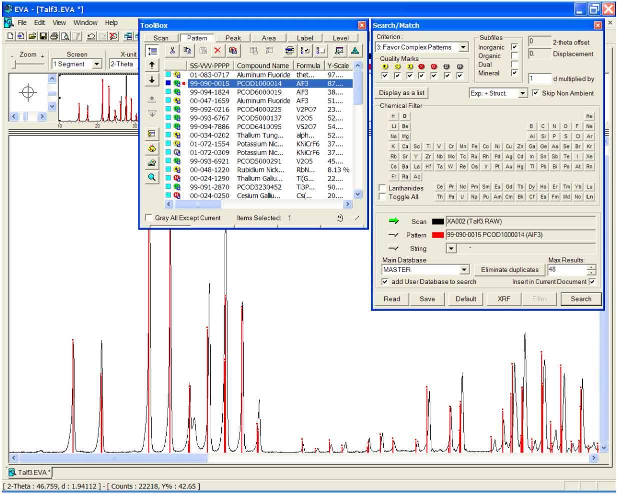


video

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

COD ir PCOD duomenų bazių panaudojimas

Kristalinės medžiagos identifikavimas



	SS-VVV-PPPP	Compound Name	Formula	Y-Scale
🟦	01-083-0717	Aluminum Fluoride	thet...	97....
🟦	99-090-0015	PCOD1000014	AlF3	87....
🟦	99-094-1824	PCOD6000019	AlF3	38....
🟦	00-047-1659	Aluminum Fluoride	AlF3	51....
🟦	99-092-0216	PCOD4000225	V2PO7	23....
🟦	99-093-6767	PCOD5000137	V2O5	52....
🟦	99-094-7886	PCOD6410095	VS2O7	54....
🟦	00-034-0202	Thallium Tung...	alph...	52....
🟦	01-072-1554	Potassium Nic...	KNiCrF6	37....
🟦	01-072-0309	Potassium Nic...	KNiCrF6	37....
🟦	99-093-6921	PCOD5000291	V2O5	45....
🟦	00-048-1220	Rubidium Nick...	RbN...	8.13 %
🟦	00-024-1290	Thallium Galliu...	Tl(G...	22....
🟦	99-091-2870	PCOD3230452	Ti3P...	90....
🟦	00-024-0250	Cesium Galliu...	Cs(...	20....

□ Gray All Except Current Items Selected: 1

Criterion : 3: Favor Complex Patterns

Subfiles: Inorganic ☑ Organic □ Dual □ Mineral ☑

0 2-theta offset

0. Displacement

1 d multiplied by

Quality Marks: ☑ ☑ ☑ ☑ ☑ ☑ ☑

Display as a list Exp. + Struct ☑ Skip Non Ambient

Chemical Filter

□ Lanthanides
□ Toggle All

Scan XA002 (Talf3.RAW)

Pattern 99-090-0015 PCOD1000014 (AlF3)

String -

Main Database MASTER Eliminate duplicates Max Results: 48

☑ add User Database to search Insert in Current Document ☑

Read Save Default XRF Filter Search

[2-Theta : 46.759, d : 1.94112] - [Counts : 22218, Y% : 42.65]

Medžiagos identifikavimas pagal Rentgeno spindulių sklaidymo intensyvumus.

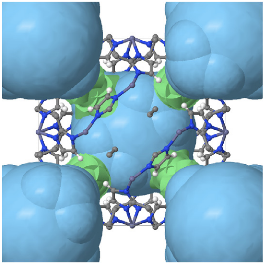
Paveiksliuką parengė Armelis Le Bail ([Le Bail, 2008])

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

MOF identifikavimas COD

Užsienio mokslininkai naudoja COD savo tyrimuose

First & Floudas (2013) „MOFomics: Computational pore characterization of metal–organic frameworks“:



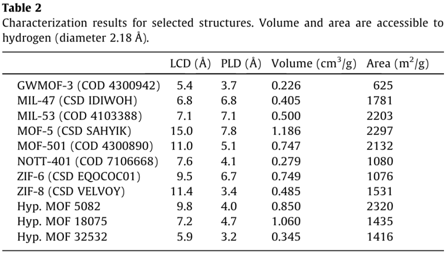


Table 2
Characterization results for selected structures. Volume and area are accessible to hydrogen (diameter 2.18 Å).

	LCD (Å)	PLD (Å)	Volume (cm^3/g)	Area (m^2/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

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

COD mokymo tikslams

Šešiavalentė anglis (!?)

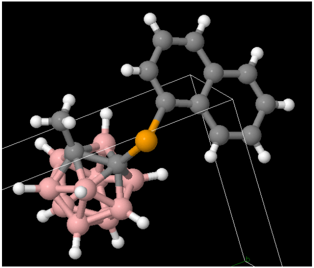
Duomenų bazėje stebime anglies („C“) atomus, nutolusius per kovalentinės jungties atstumą nuo **6** kaimynų.(?)

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

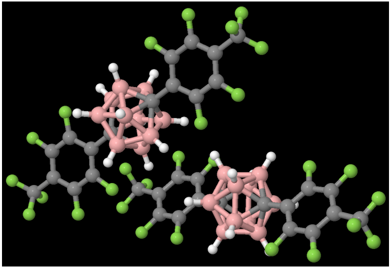
COD mokymo tikslams

Šešiavalentė anglis (!?)

Duomenų bazėje stebime anglies („C“) atomus, nutolusius per kovalentinės jungties atstumą nuo **6** kaimynų.(!)



COD 7015488



COD 7015654

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

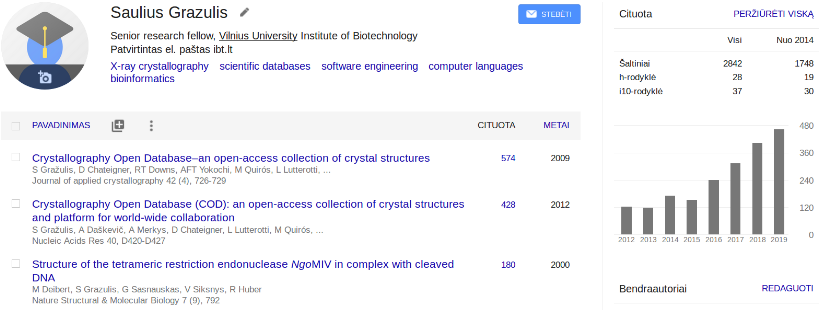
Turinys

1. Motyvacija – kodėl duomenų bazės?
2. COD kūrimo istorija
3. COD turinys
4. COD pritaikymai
5. COD nauda mums :)
6. Pamąstymai apie atvirus duomenis
7. Tolimesni planai

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

COD citavimai

Dvi pradinės COD publikacijos surinko virš 1000 citavimų!



Saulius Grazulis

Senior research fellow, Vilnius University Institute of Biotechnology
Patvirtintas el. paštas ibt.lt

X-ray crystallography scientific databases software engineering computer languages bioinformatics

□ PAVADINIMAS	CITUOTA	METAI
□ Crystallography Open Database–an open-access collection of crystal structures S Gražulis, D Chateigner, RT Downs, AFT Yokochi, M Quirós, L Lutterotti, ... Journal of applied crystallography 42 (4), 726-729	574	2009
□ Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration S Gražulis, A Daškevič, A Merkys, D Chateigner, L Lutterotti, M Quirós, ... Nucleic Acids Res 40, D420-D427	428	2012
□ Structure of the tetrameric restriction endonuclease *Ngo*MIV in complex with cleaved DNA M Deibert, S Grazulis, G Sasnauskas, V Siksnys, R Huber Nature Structural & Molecular Biology 7 (9), 792	180	2000

Cituota PERŽIŪRĖTI VISKĄ

	Visi	Nuo 2014
Šaltiniai	2842	1748
h-rodyklė	28	19
i10-rodyklė	37	30

Bendraautoriai REDAGUOTI

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

COD grantai

- MIP-124/2010 „Atviros prieigos mažų molekulių kristalografinė duomenų bazė COD“
- MIP-025/2013 „Statistinė struktūrų analizė atviroje kristalografinėje duomenų bazėje COD ir jos plėtimas“
- H2020 SOLSA „Akustinis gręžimas sujungtas su automatine mineralų analize: ištisiniame procese, gręžimo vietoje, realiu laiku“

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

The SOLSA project

Sonic drilling in Ni laterites
courtesy of Eijkelkamp SonicSampDrill

Discover SOLSA

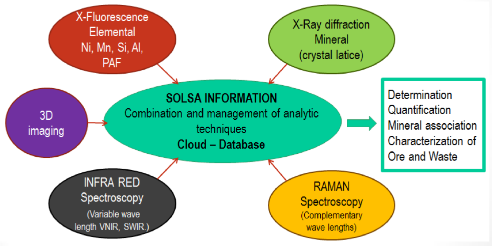
http://solsa-mining.eu/

- Crystal structures (COD)
- Raman spectra (ROD)
- Hyperspectral spectra (HOD)

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

SOLSA

The SOLSA project, COD and ROD



X-Fluorescence
Elemental
Ni, Mn, Si, Al,
PAF

X-Ray diffraction
Mineral
(crystal latice)

3D
imaging

SOLSA INFORMATION
Combination and management of analytic techniques
Cloud – Database

INFRA RED
Spectroscopy
(Variable wave length VNIR, SWIR.)

RAMAN
Spectroscopy
(Complementary wave lengths)

Determination
Quantification
Mineral association
Characterization of Ore and Waste

⇔

COD

ROD

COD will be used in SOLSA for:

- mineral identification;
- subsequent data dissemination.

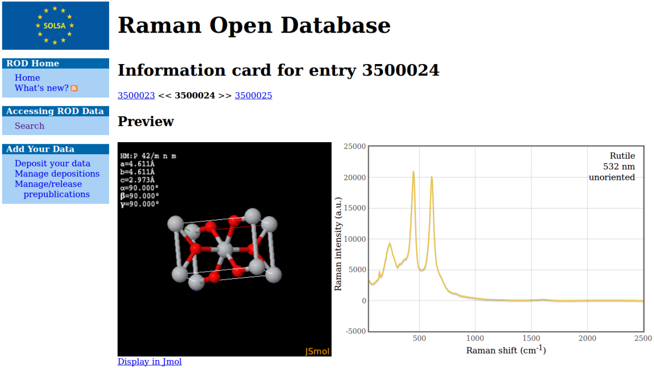
SOLSA data flow diagram courtesy Monique Le Guen, ERAMET.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Raman spectroscopy data ROD

The Raman Open Database

http://solsa.crystallography.net/rod



ROD Home

Home

What's new?

Accessing ROD Data

Search

Add Your Data

Deposit your data

Manage depositions

Manage/release prepublications

Raman Open Database

Information card for entry 3500024

3500023 << **3500024** >> 3500025

Preview

Display in Jmol

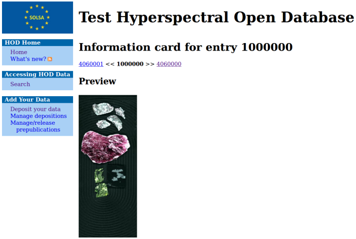
Data records contributed to the ROD by Yassine El Mendili

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

HOD record example

examples/hod/1000000-head.cif:

```
data_1000000
loop_
_[local]_description
'ENVI File'
'Created [Wed Jun 08 12:34:07 2016]'
_[local]_wavelength_units          Nanometers
loop_
_hyper_bands.default
220
227
253
_hyper_bands.lines                  937
_hyper_bands.number                 288
_hyper_bands.samples                384
_hyper_file.byte_order              0
_hyper_file.data_type               4
_hyper_file.type                    ENVI_Standard
_hyper_header.offset                0
_hyper_header_file.contents
;ENVI
description = {
  ENVI File, Created [Wed Jun 08 12:34:07 2016]}
samples = 384
lines   = 937
```

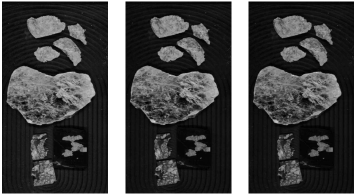


Test Hyperspectral Open Database

Information card for entry 1000000

4060001 << 1000000 >> 4060000

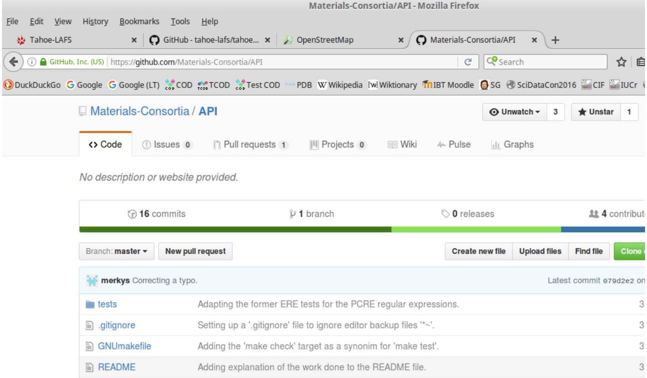
Preview



This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Common REST API

- Agreed upon in the 2016 Leiden CECAM workshop;
- Suitable for all structural and QM databases.


Materials-Consortia/API - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Tahoe-LAFS
GitHub - tahoe-lafs/tahoe...
OpenStreetMap
Materials-Consortia/API
GitHub, Inc. (US) https://github.com/Materials-Consortia/API
Search
DuckDuckGo Google Google (LT) COD TCOD Test COD PDB Wikipedia Wiktionary IBT Moodle SG SciDataCon2016 CIF IUCr
Materials-Consortia / API
Unwatch 3
Unstar 1
Code
Issues 0
Pull requests 1
Projects 0
Wiki
Pulse
Graphs
Branch: master
New pull request
Create new file
Upload files
Find file
Clone

No description or website provided.

16 commits | 1 branch | 0 releases | 4 contribut

merkys Correcting a typo. Latest commit 079d2e2 on

tests	Adapting the former ERE tests for the PCRE regular expressions.	3
.gitignore	Setting up a '.gitignore' file to ignore editor backup files '*~'.	3
GNUmakefile	Adding the 'make check' target as a synonim for 'make test'.	3
README	Adding explanation of the work done to the README file.	3

https://github.com/Materials-Consortia/API

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Turinys

1. Motyvacija – kodėl duomenų bazės?
2. COD kūrimo istorija
3. COD turinys
4. COD pritaikymai
5. COD nauda mums :)
6. Pamąstymai apie atvirus duomenis
7. Tolimesni planai

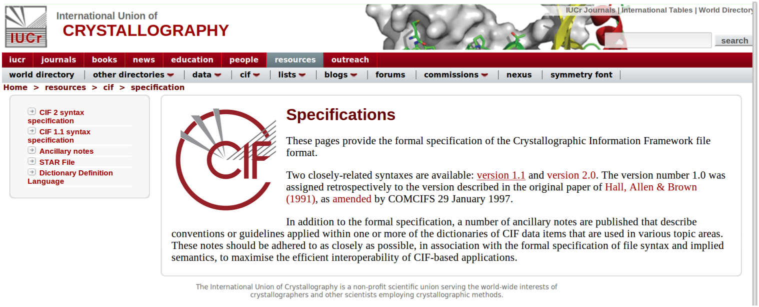
This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Requirements for long-term data archiving and reuse

- Platform independence
    - Text-based formats (ASCII, UTF-8)
- Software independence
- Network-transparency
    - Standard, open protocols (W3C http)
    - Standard, open data carrier formats (JSON, XML, CIF).
    - RESTful servers
- Machine-readable semantics
    - Dictionaries, schemas
- Durability
    - Persistent identifiers
    - Open data principles
    - FAIR principles

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Data exchange in crystallography


IUCr
International Union of CRYSTALLOGRAPHY
IUCr Journals | International Tables | World Directory
search
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
CIF

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]

The Crystallographic Interchange File/Framework (CIF):

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

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

CIF for scientific data

examples/data/2100858-head.cif:

```
data_2100858
loop_
_publ_author_name
’Buttner, R. H.’
’Maslen, E. N.’
_publ_section_title
;
 Structural parameters and electron difference density in BaTiO~3~
;
_journal_issue                    6
_journal_name_full                ’Acta Crystallographica Section B’
_journal_page_first               764
_journal_page_last                769
_journal_volume                   48
_journal_year                     1992
_chemical_compound_source         ’synthetic, from a mixture of KF:KMoO4:BaTiO3’
_chemical_formula_sum             ’Ba O3 Ti’
_chemical_formula_weight          233.24
_symmetry_cell_setting            tetragonal
_symmetry_space_group_name_Hall   ’P 4 -2’
_symmetry_space_group_name_H-M    ’P 4 m m’
_cell_angle_alpha                 90.0
_cell_angle_beta                  90.0
_cell_angle_gamma                 90.0
_cell_formula_units_Z             1
_cell_length_a                    3.9998(8)
_cell_length_b                    3.9998(8)
_cell_length_c                    4.0180(8)
```

This project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 689868.

Turinys

1. Motyvacija – kodėl duomenų bazės?
2. COD kūrimo istorija
3. COD turinys
4. COD pritaikymai
5. COD nauda mums :)
6. Pamąstymai apie atvirus duomenis
7. Tolimesni planai

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

COD perspektyvos

- COD sąsajos su kitomis DB (PDB, PubChem, WikiData, ...)
- COD „kristalografo darbo vieta“ – bendruomeninis tinklalapis kristalų struktūroms laikyti ir analizuoti.
- Daugiacentris COD serveris (*angl.* multi-master replication)
- Visaapimanti struktūrų duomenų bazė
- DI (Loginio programavimo, DNN) taikymai žinioms iš COD išgauti

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Acknowledgements

VU Institute of Biotechnology

Virginijus Siksnys
(*head of the dept.*)

Andrius Merkys
Antanas Vaitkus
Algirdas Grybauskas
Alina Belova
Erikas Raginis

The SOLSA team

Monique Le Guen
Beate Orberger
Daniel Chateigner
Henry Pilliere
and all the team working on the project!

COD Advisory board

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

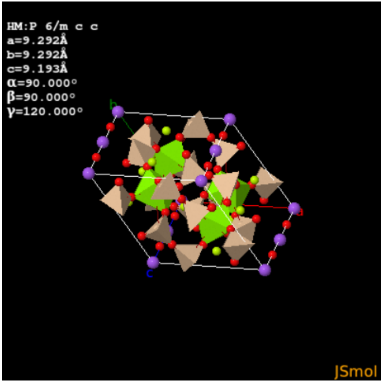
This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

SOLSA

Thank you!

http://en.wikipedia.org/wiki/Emerald



HM:P 6/m c c
a=9.292Å
b=9.292Å
c=9.193Å
α=90.000°
β=90.000°
γ=120.000°

http://www.crystallography.net/5000095.html

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

References I

- Baldi, P. (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.
  Journal of chemical information and modeling, 51:3029.

- Gražulis, S., Merkys, A., Vaitkus, A., and Okulič-Kazarinas, M. (2015).
  Computing stoichiometric molecular composition from crystal structures.
  Journal of Applied Crystallography, 48:85–91.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

References II

- Hall, S. R., Allen, F. H., and Brown, I. D. (1991).
  The crystallographic information file (CIF): a new standard archive file for crystallography.
  Acta Crystallographica Section A, 47:655–685.

- Le Bail, A. (2008).
  Frontiers between crystal-structure prediction and determination by powder diffractometry.
  Powder Diffraction Suppl., pages S5–S12.

- Long, F., Nicholls, R. A., Emsley, P., Gražulis, S., Merkys, A., Vaitkus, A., and Murshudov, G. N. (2017a).
  ACEDRG: A stereo-chemical description generator for ligands.
  Acta Crystallographica Section D, 73(2):112–122.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

References III

Long, F., Nicholls, R. A., Emsley, P., Gražulis, S., Merkys, A., Vaitkus, A., and Murshudov, G. N. (2017b).
Validation and extraction of stereochemical information from small molecular databases.
Acta Crystallographica Section D, 73(2):103–111.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

API query examples

http://crystallography.net/cod/optimade/structures?filter=elements="Si,O"ANDnelements=2&limit=1

```
{
  "resource": {
    "base_url": "http://www.crystallography.net/cod/optimade/v1.0.0-alpha.1/"
  },
  "query": {
    "api_version": "v1.0.0-alpha.1",
    "data_returned": 1,
    "representation": "/structures?filter=elements=\"Si,O\"ANDnelements=2&limit=1",
    "last_id": "1010921",
    "time_stamp": "2017-04-06T05:46:50Z",
    "implementation": {
      "maintainer": {
        "email": "cod-bugs@ibt.lt"
      },
      "title": "Crystallography␣Open␣Database",
      "version": "v1.0.0-alpha.11",
      "source_url": "svn://crystallography.net/cod/trunk/cod/cgi-bin/optimade.pl@194653"
    },
    "data_available": 344
  },
  "data": [
    {
      "last_modified": "2017-02-28T05:33:56Z",
      "properties": {
        "formula": "O2␣Si"
      },
      "url": "http://www.crystallography.net/cod/1010921.cif",
      "immutable_id": "http://www.crystallography.net/cod/1010921.cif@130149",
```

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.