

Launching the Theoretical Crystallography Open Database Saulius Gražulis^a, Andrius Merkys^a, Antanas Vaitkus^b, Armel Le Bail^c, Daniel Chateigner^{d,e,f}, Linas Vilčiauskas^g, Stefaan Cottenier^h,

Torbjörn Björkman⁷ and Peter Murray-Rust⁷

^aVilnius University, Institute of Biotechnology, Department of Protein-DNA Interactions, Vilnius University, Faculty of Mathematics and Informatics, Department of Mathematical Computer Science, Vilnius, Lithuania; ^cUniversite du Maine, Laboratoire des Oxydes et Fluorures, Université de Sciences, Le Mans, France; ^dUniversité de Caen-Basse Normandie, UMR 6508 CRISMAT, Caen, France; ^eENSICAEN, UMR 6508 CRISMAT, Caen, France; ^fCNRS, UMR 6508 CRISMAT, Caen, France; ^gDepartment of Chemistry, New York, USA; ^hGhent University, Center for Molecular Modeling (CMM) & Department of Materials Science and Engineering (DMSE), Zwijnaarde, Belgium; ⁱAalto University School of Science, COMP/Department of Applied Physics, Espoo, Finland; ^jThe University of Cambridge, Department of Chemistry, Cambridge, UK

Abstract

As computational chemistry methods enjoy unprecedented growth, computer power increases and price/performance ratio drops, a large number of crystal structures can today be refined and their properties computed using modern theoretical approaches (DFT, post-HF, QM/MM, etc.). We thus increasingly feel that an open collection of theoretically computed chemical structures would be a valuable resource for the scientific community. To address this need, we have launched a Theoretical Crystallography Open Database (TCOD, [4]). The TCOD sets a goal to collect a comprehensive set of computed crystal structures that would be made available under an Open Data license and invites all scientists to deposit their published results or pre-publication data. Accompanied with a large set of experimental structures in the COD database [3], the TCOD opens immediate possibilities for experimental and theoretical data cross-validation. The property results can now be tested against the Material Properties Open Database [6, 1].



Bibliography

- [1] Chateigner et al. Material properties open database, retrieved: 2014-07-23. http://www.materialproperties.org/.
- [2] Gražulis et al. Crystallography open database, retrieved: 2014-07-24. //www.crystallography.net/cod
- [3] Gražulis et al. Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. Nucleic Acids Research, 40(D1):D420–D427, Jan 2012.
- [4] Gražulis et al. Theoretical crystallography open database, retrieved: 2014-07-23. http://www.crystallography.net/tcod/.
- [5] Hall et al. The crystallographic information file (CIF): a new standard archive file for crystallography. Acta Crystallographica Section A, 47(6):655-685, Nov 1991.
- [6] Pepponi et al. Mpod: A material property open database linked to structural information. *Nuclear Instruments and* Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, 284(0):10 – 14, 2012. E-MRS 2011 Spring Meeting, Symposium M: X-ray techniques for materials research-from laboratory sources to free electron lasers.

TCOD Web site

http://www.crystallography.net/tcod/

Theoretical Crystallography Open Database



dd Your Data

posit your data

lanage deposition

anage/release prepublications

COD Wiki Obtaining COD

Citing COD COD Mirrors

Advices to donato



Open-access collection of theoretically calculated or refined crystal ructures of organic, inorganic, metal-organic compounds and minerals, excluding biopolymers

All data on this site have been placed in the public domain by the contributors Currently there are **173** entries in the TCOD Latest deposited structure: <u>20000173</u> on **2013-12-16** at **20:06:53 UTC**



TCOD structure retrieval

TCOD Home Home

Accessing TCOD Data

What's new?

Browse

Search

Add Your Data

Deposit your data

Manage depositions

/anage/release prepublications

Documentation

Obtaining COD

Advices to donators

COD Wiki

Citing COD

COD Mirrors

CIF Information Card Information card for 2000096

<u>0000095</u> << 20000096 >> <u>20000097</u>

Preview

7,730Å 14,619Å

Structure parameters 3,4-diaminopyridin-1-ium dihydrogen ommon name phosphate C5 H10 N3 O4 P alculated formula - C5 H10 N3 O4 P Ab initio structure determination of Title of publication 3.4-diaminopyridin-1-ium dihydroge phosphate Authors of Armel Le Bail; Lubomir Smrcok publication Journal of

20000096.cif

publication Year of publication Journal volume Pages of publication

Coordinates

Powder Diffraction 2011 26 321 $16.0725 \pm 0.0009 \text{ Å}$ 7.7301 ± 0.0003 Å

 14.6189 ± 0.0009 Å 90°

TCOD data deposition



Theoretical Crystallography Open Database

What's new?

Accessing COD Data Search

Add Your Data Deposit your data Manage depositions Manage/release prepublications

Documentation COD Wiki Obtaining COD Citing COD COD Mirrors

Validation and Deposition Interface Deposit to TCOD all valid files **Valid**: 3 File Status Actions a908973y_si_002.cif_valid____Edit___Add F_{obs}___Deposit to TCOD File [ja908973y_si_002.cif] is correct 908973y_si_003.cif valid Edit Add F_{obs} Deposit to TCOD File [ja908973y_si_003.cif] is correct 008973y_si_004.cif valid Edit Add F_{obs} Deposit to TCOD File [ja908973y_si_004.cif] is correct <u>Log out</u> Advices to donators

7	

:
 C
F
 C



Structure description levels

Structures may be described at different level of detail in TCOD:

Level 0	Level 1	Level 2
	Level 0, plus:	Level 1, plus:
1. lattice and symmetry	 last cycle energy changes 	 input scripts files
2. atomic coordinates	 residual forces on atoms 	2. command lin
 bibliography reference 	3. code-specific convergence criteria	 output logs c code

Dictionaries

To ensure high quality of deposited data, TCOD offers ontologies in a form of CIF [5] dictionaries and an automated pipeline that checks each submitted structure against a set of community-specified criteria for convergence, computation quality and reproducibility. Dictionaries are available at: http://www.crystallography.net/tcod/cif/dictionaries/:

cif_tcod.dic	cif_dft
data_tcode_structure_type name '_tcode_structure_type' type char loopenumeration enumeration_detail	data r t c : Tot
ground-state 'refined crystal structure at ground state'	

tcod dft valence electrons me ' dft valence electrons' pe numb al number of electrons in calculation

Conclusions

- The CIF [5] appears to be very well suitable to describe results of computational chemistry and computational crystallography.
- The existing COD [2] software permitted very fast implementation of a database for theoretically computed structures.
- It remains to be seen if the community will endorse this format of data exchange.

Acknowledgements

This research was funded by a grant (No. KEL-051/2014) from the Research Council of Lithuania.

Created with Ubuntu \rightarrow MySQL \rightarrow R Sweave \rightarrow PT_{F} Xbeamerposter

Finish