Launching the Theoretical Crystallography Open Database

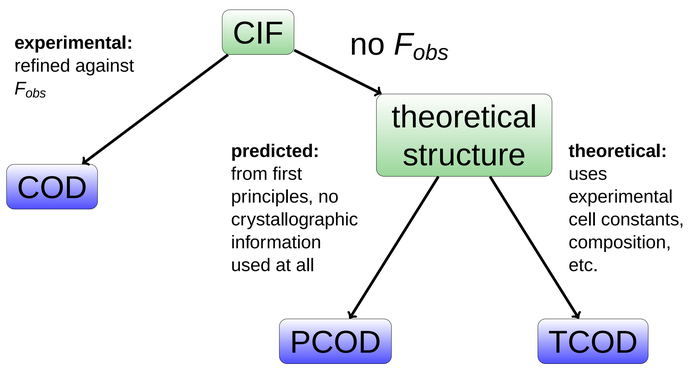
Saulius Gražulis[a], Andrius Merkys[a], Antanas Vaitkus[b], Armel Le Bail[c], Daniel Chateigner[d,e,f], Linas Vilčiauskas[g], Stefaan Cottenier[h], Torbjörn Björkman[i] and Peter Murray-Rust[j]

[a]Vilnius University, Institute of Biotechnology, Department of Protein-DNA Interactions, Vilnius, Lithuania; [b]Vilnius University, Faculty of Mathematics and Informatics, Department of Mathematical Computer Science, Vilnius, Lithuania; [c]Universite du Maine, Laboratoire des Oxydes et Fluorures, Université du Maine, Faculté des Sciences, Le Mans, France; [d]Université de Caen-Basse Normandie, UMR 6508 CRISMAT, Caen, France; [e]ENSICAEN, UMR 6508 CRISMAT, Caen, France; [f]CNRS, UMR 6508 CRISMAT, Caen, France; [g]Department of Chemistry, New York University, New York, USA; [h]Ghent University, Center for Molecular Modeling (CMM) & Department of Materials Science and Engineering (DMSE), Zwijnaarde, Belgium; [i]Aalto University School of Science, COMP/Department of Applied Physics, Espoo, Finland; [j]The 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].

Structures in *COD databases



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. http://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.

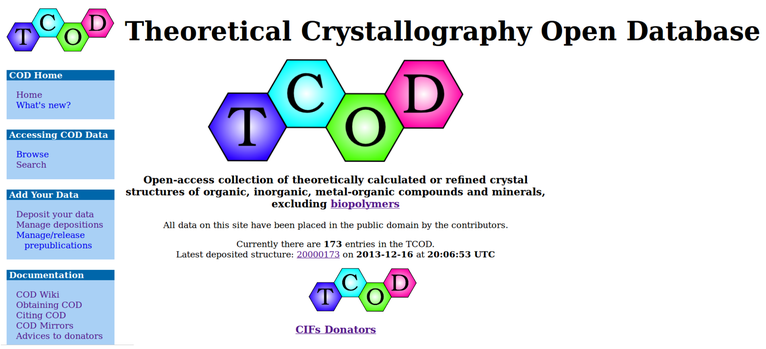
[4] Gražulis et al. Theoretical crystallography open database, retrieved: 2014-07-23. http://www.crystallography.net/tcod/.

[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.

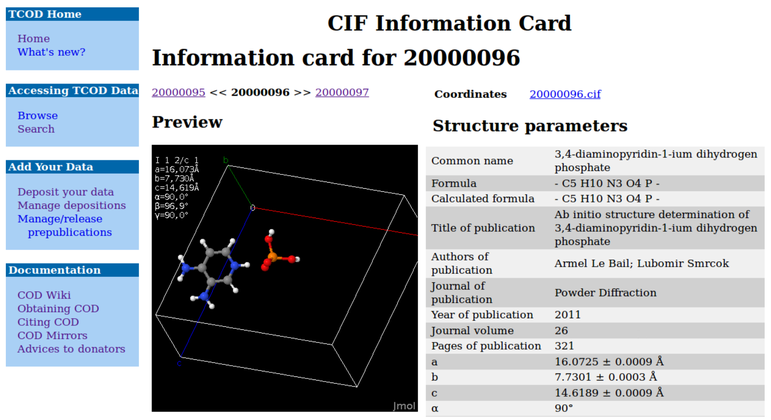
[6] Pepponi et al. Mpod: A material property open database linked to structural information. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 284(0):10 – 14, 2012. E-MRS 2011 Spring Meeting, Symposium M: X-ray techniques for materials research-from laboratory sources to free electron lasers.

TCOD Web site

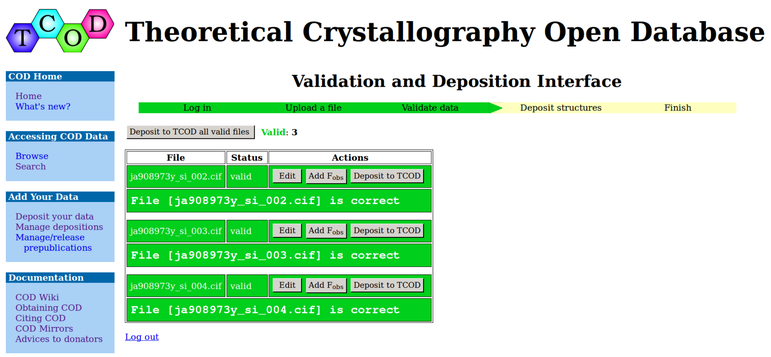
http://www.crystallography.net/tcod/



TCOD structure retrieval



TCOD data deposition



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	1. last cycle energy changes	1. input scripts and files
2. atomic coordinates	2. residual forces on atoms	2. command line
3. bibliography reference	3. code-specific convergence criteria	3. output logs of the 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

```
data_tcode_structure_type
    _name '_tcode_structure_type'
    _type char
    loop_ _enumeration
        _enumeration_detail

        ground-state
            'refined crystal structure at ground state'
```

cif_dft.dic

```
data_tcod_dft_valence_electrons
    _name '_dft_valence_electrons'
    _type numb
    _definition
; Total 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 → MySQL → R Sweave → LaTeXbeamerposter