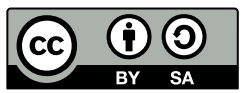
Crystallography Open Database (COD)

Saulius Gražulis

Vilnius, 2024

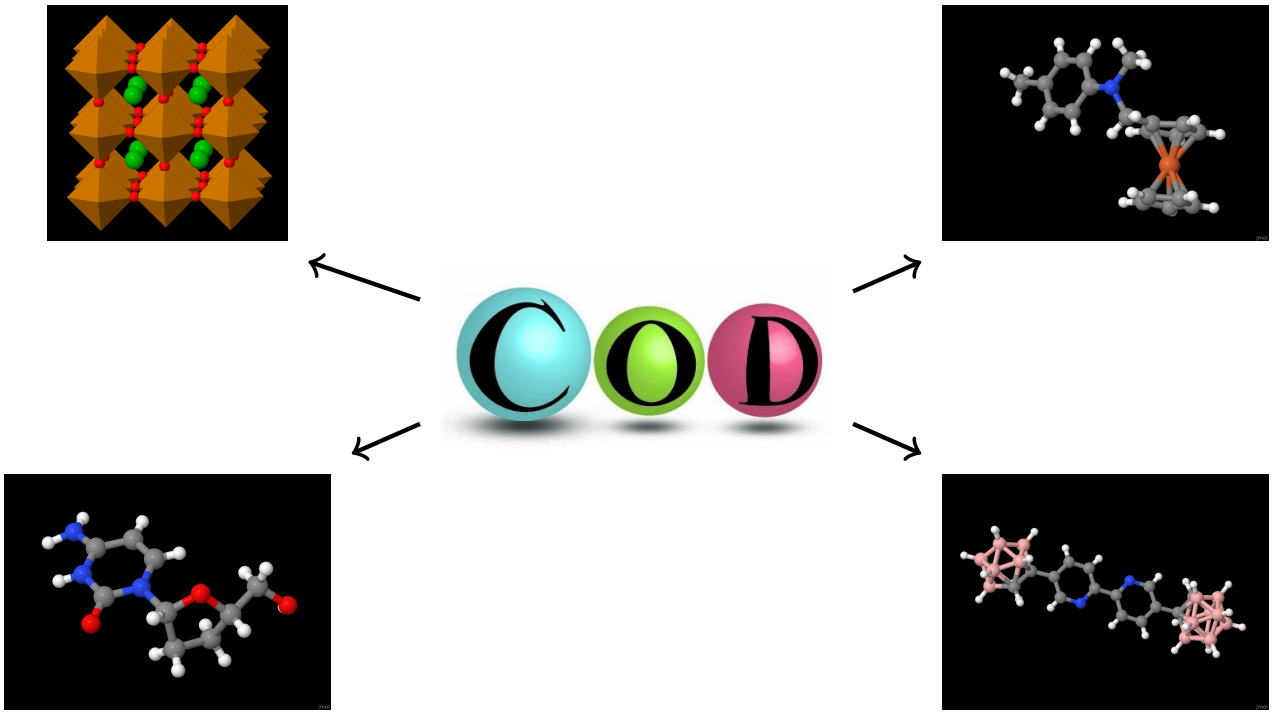
Vilnius University Institute of Biotechnology

Id: slides.tex 2572 2024-02-22 16:47:37Z saulius February 22, 2024



COD contents

https://www.crystallography.net/cod



510 995 records as of 2024-02-22, available under **CC0 License**

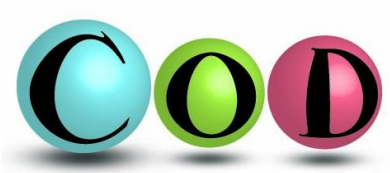
All data are presented in a standardised, machine-readable form (Gražulis et al. 2009; Gražulis et al. 2012), in CIF (Hall et al. 1991) format.

Accessing the COD

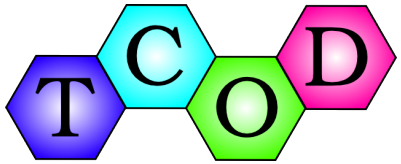
COD data can be accessed:

1. Via the Web page:
   https://www.crystallography.net/cod/7159763.html
2. Via the COD REST API:
   https://www.crystallography.net/cod/7159763.cif
   https://www.crystallography.net/cod/result?text=perovskite
3. Via the OPTIMADE API (Andersen et al. 2021):
   https://www.crystallography.net/cod/optimade/structures?filter=elements+HAS+"U"
4. Via SQL:
   ```
   mysql -u cod_reader -h sql.crystallography.net cod -e \
   'select file from data where formula = "- H2 O -"'
   ```
5. By downloading to your computer using Subversion, `rsync` or simple Web download:
   https://wiki.crystallography.net/howtoobtaincod

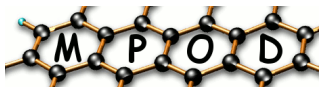
COD “sisters”



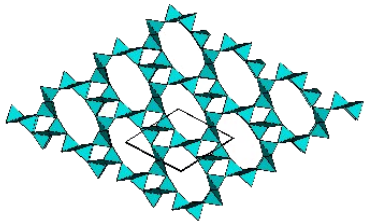
http://www.crystallography.net/cod
$>$ 500 000 entries



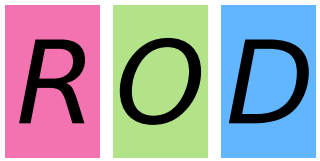
http://www.crystallography.net/tcod
$>$ 7400 entries (ready to grow to $>$ 10^7?)



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



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



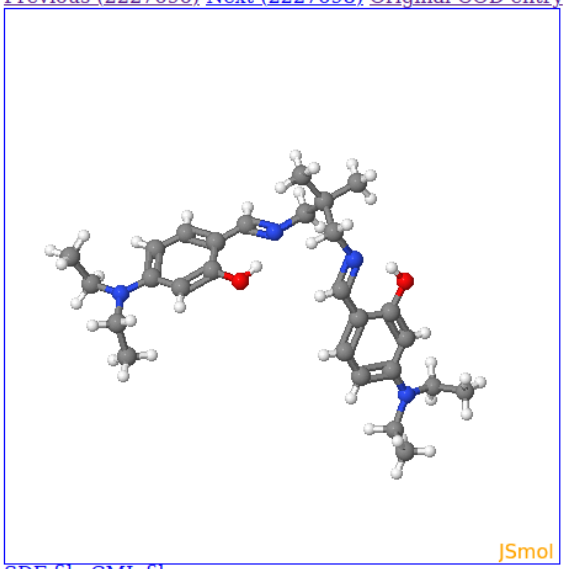
http://solsa.crystallography.net/rod/
$>$ 1100 entries

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

COD chemical repertoire

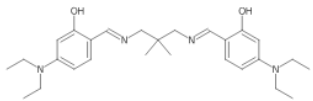
http://molecules.crystallography.net/

Previous (2227696) Next (2227698) Original COD entry



SDF file CML file

Reduced structural formula



Reduced canonical SMILES:

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

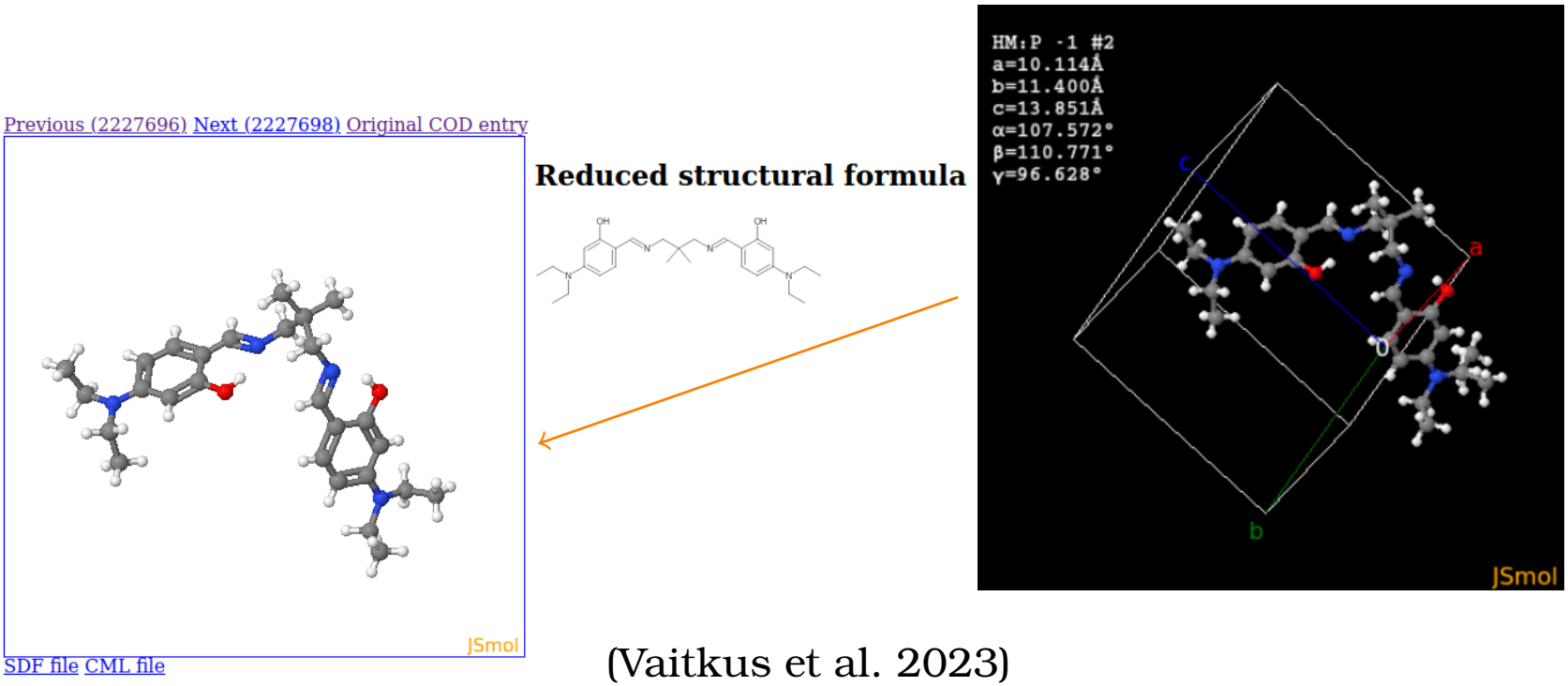
Unique components

SMILES	InChI
CCN(c1ccc(c(c1)O)/C=N/CC(C/N=C/c1ccc(cc1O)N(CC)CC)(C)C)CC	InChI=1S/C27H40N4O2/c1-7-30(8-2)23-13-11-21(25(32)15-23)17-28-19-27(5,6)2 /h11-18,32-33H,7-10,19-20H2,1-6H3/b28-17+,29-18+

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

COD chemical repertoire

http://molecules.crystallography.net/



(Vaitkus et al. 2023)

Reduced canonical SMILES:

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

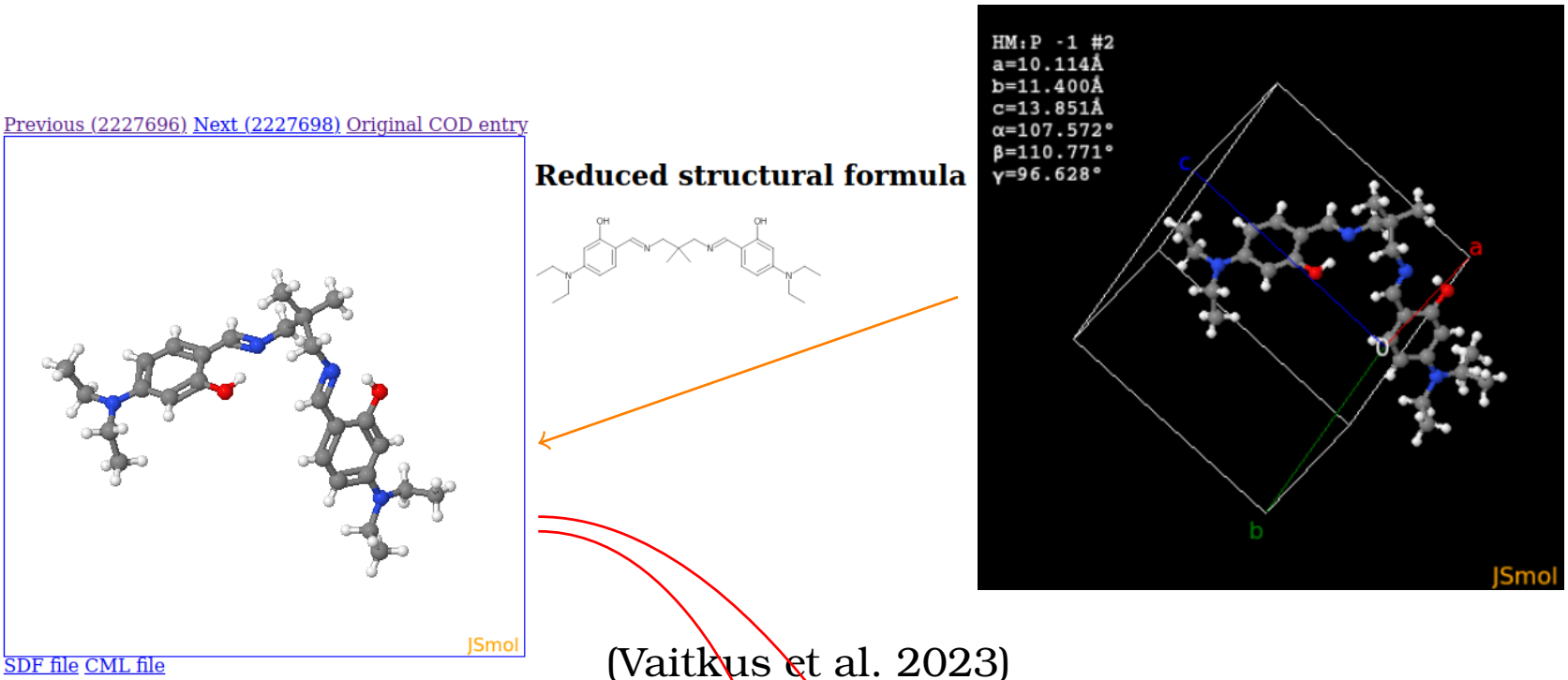
Unique components

SMILES	InChI
CCN(c1ccc(c(c1)O)/C=N/CC(C/N=C/c1ccc(cc1O)N(CC)CC)(C)C)CC	InChI=1S/C27H40N4O2/c1-7-30(8-2)23-13-11-21(25(32)15-23)17-28-19-27(5,6)2…/h11-18,32-33H,7-10,19-20H2,1-6H3/b28-17+,29-18+

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

COD chemical repertoire

http://molecules.crystallography.net/



(Vaitkus et al. 2023)

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

Acknowledgements

VU LSC IBT (KICIS)

Andrius Merkys
Antanas Vaitkus
Algirdas Grybauskas

QM community

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

VU MIF II (FMG)

Linas Laibinis
Karolis Petrauskas

COD Advisory board

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

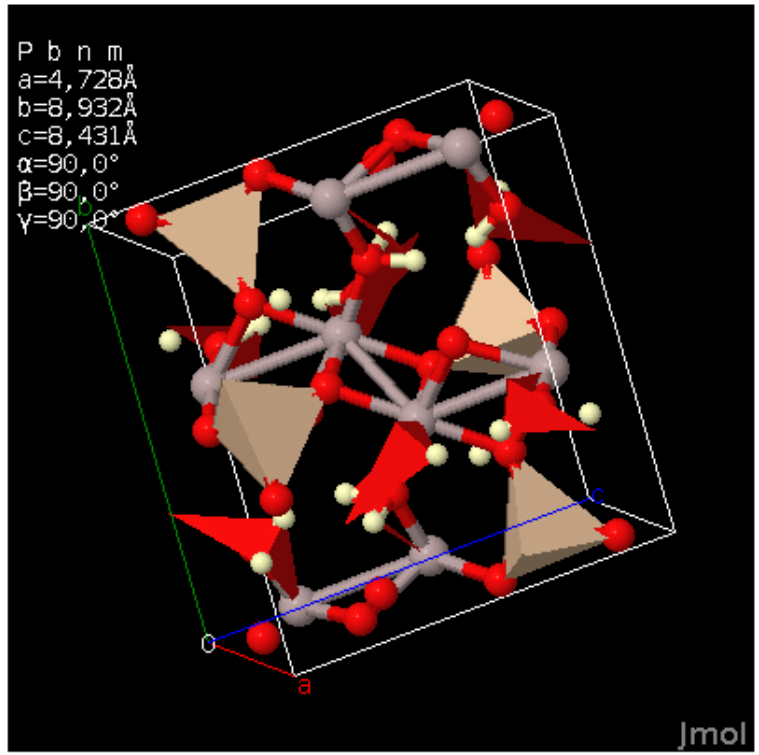
Cheminf community

Evan Bolton
Paul Thiessen
Thomas Sander

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

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

Thank you!



Coordinates	2207377.cif
Original IUCr paper	HTML

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

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

http://www.crystallography.net/cod/archives/2024/slides/MaRDA-meeting-repository-session/slides.pdf

References I


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


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


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


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


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

References II


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


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


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.