Sector of Crystallography and Chemical Informatics

Saulius Gražulis

Dubingiai, 2023

Vilnius University Life Sciences Center Institute of Biotechnology

LSC Conference

Id: slides.tex 2120 2023-04-05 06:53:44Z saulius April 5, 2023

Flagship: the Crystallography Open Database (COD)

https://www.crystallography.net

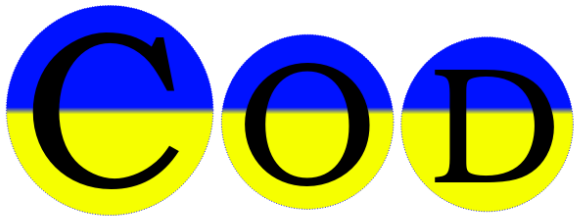
Crystallography Open Database

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



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

News

2023-03-27 Malvern Panalytical publishes a new release of their free, COD-derived search-match database. The new COD database file is meant to be used with all versions 4.x and 5.x of the PANalytical HighScore (Plus) software packages. It can be downloaded as one (7.3 GB) database file in .HSRDB format from the archive and is ready for use. Read more

Currently there are **500531** entry in the COD.
Latest deposited structure: 4038956 on **2023-04-02** at **09:49:38 UTC**

CIFs Donators

Advisory Board

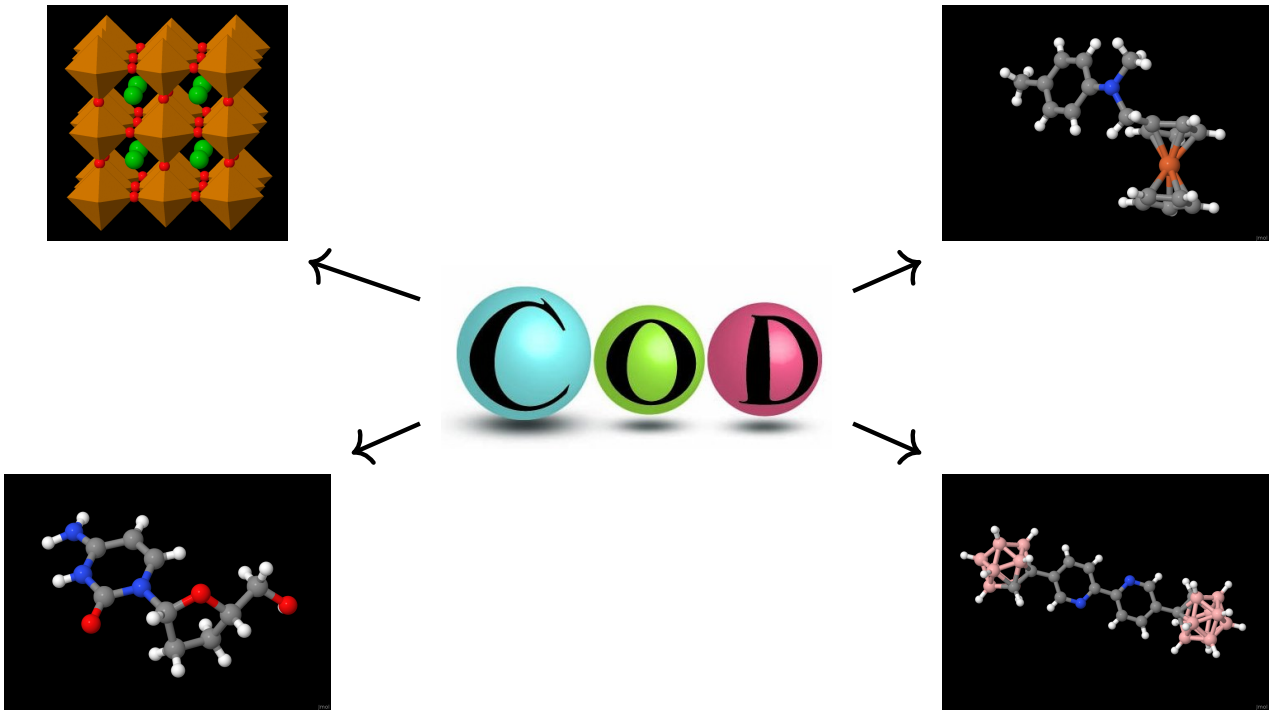
Daniel Chateigner, Xiaolong Chen, Marco Ciriotti,
Robert T. Downs, Saulius Gražulis, Werner Kaminsky, Armel Le Bail, Luca Lutterotti,
Yoshitaka Matsushita, Andrius Merkys, Peter Moeck, Peter Murray-Rust, Miguel Quirós Olozábal,
Hareesh Rajan, Antanas Vaitkus, Alexandre F.T. Yokochi

If you find bugs in the COD or have any feedback, please contact us at
cod-bugs@ibt.lt

Top of the page

COD contents

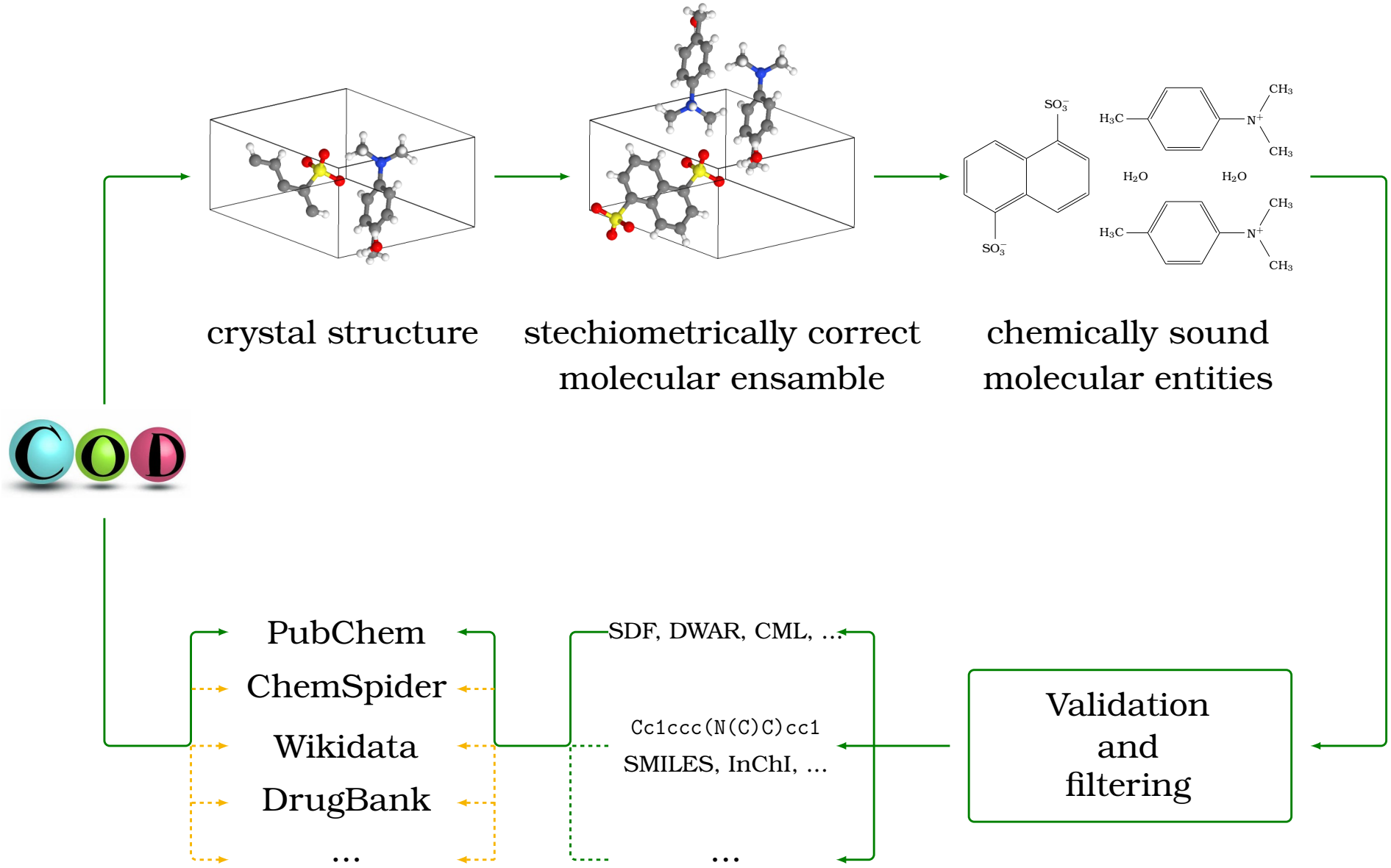
The COD covers organic, metal organic, inorganic compounds and minerals.



500592 entries as of 2023-04-04 01:51:06 UTC, under the CC0 License

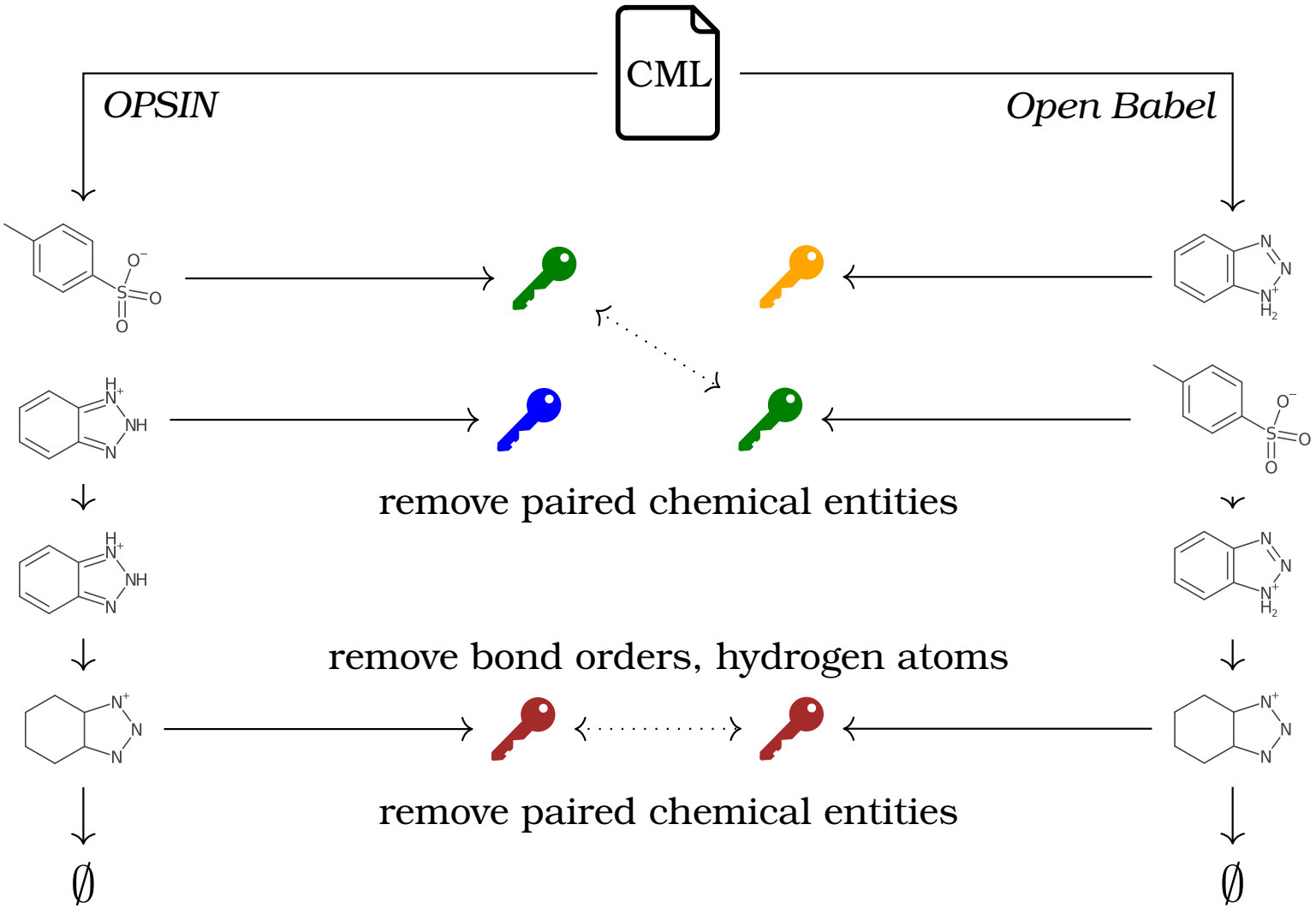
COD uses CIF for data ingestion (Grazulis et al. 2009), validation (Vaitkus et al. 2021) and transformation (Grazulis et al. 2015).

Chemical information extraction pipeline



Antanas Vaitkus, manuscript in preparation.

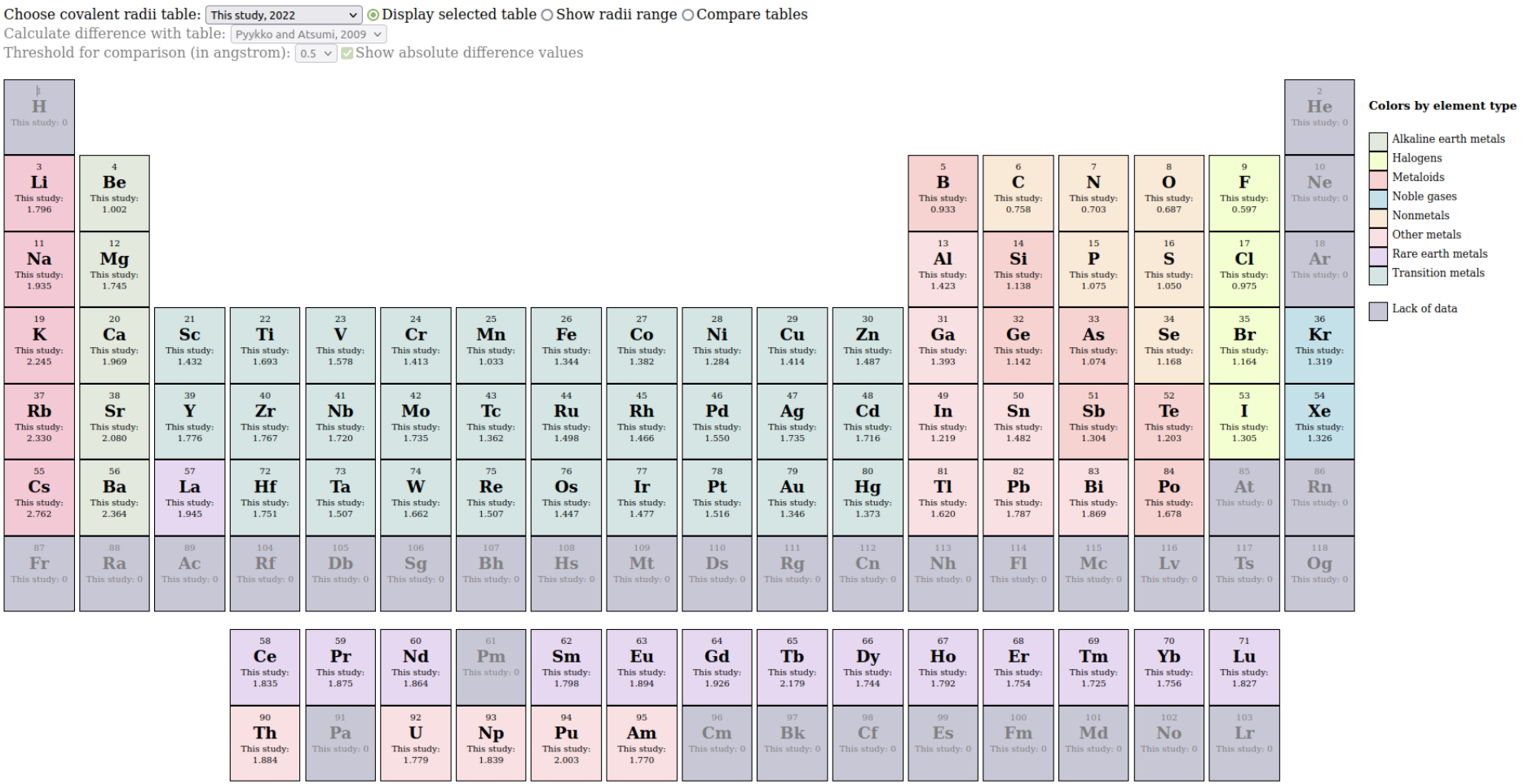
Comparison of chemical descriptions



[Merkys et al. (2023)], 2×RCoL grant, 2020–2022 and 2023–2025

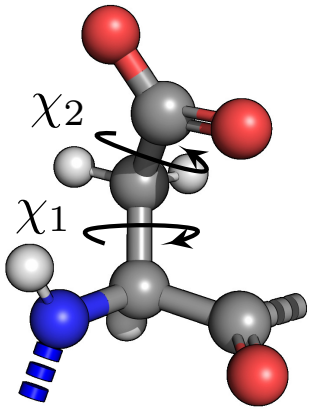
Atomic radii determination

http://databases.crystallography.lt:8080/contacts/website/cgi-bin/cov_radii_table.pl

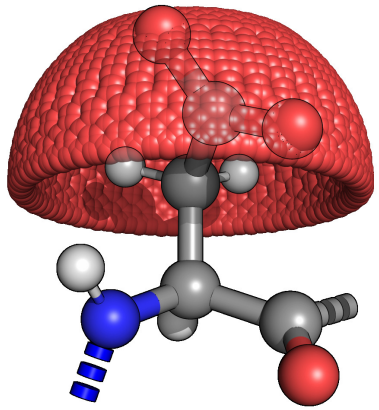


Merkys, Vaitkus & Šidlauskaitė, VU MSF grant, 2021–2022

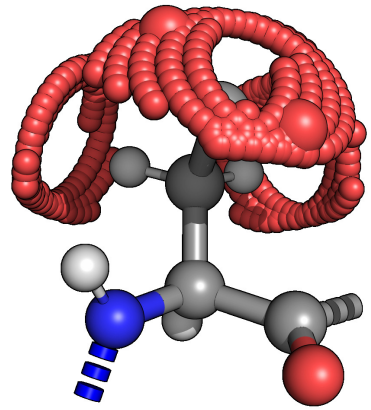
Physics-based rotamer library models

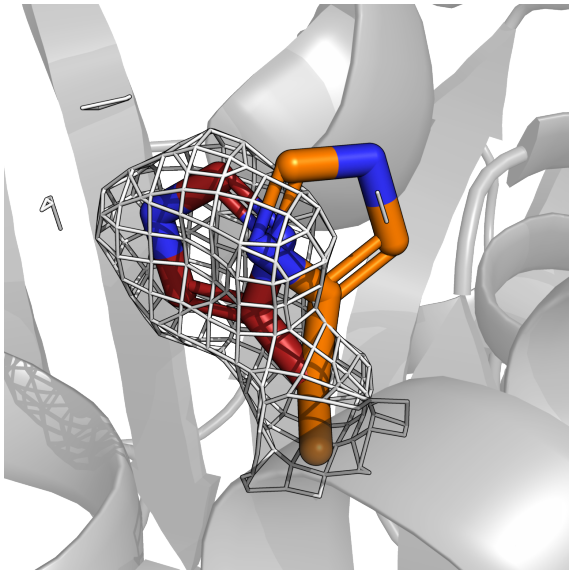


$R_{\chi i}$ $\longrightarrow$

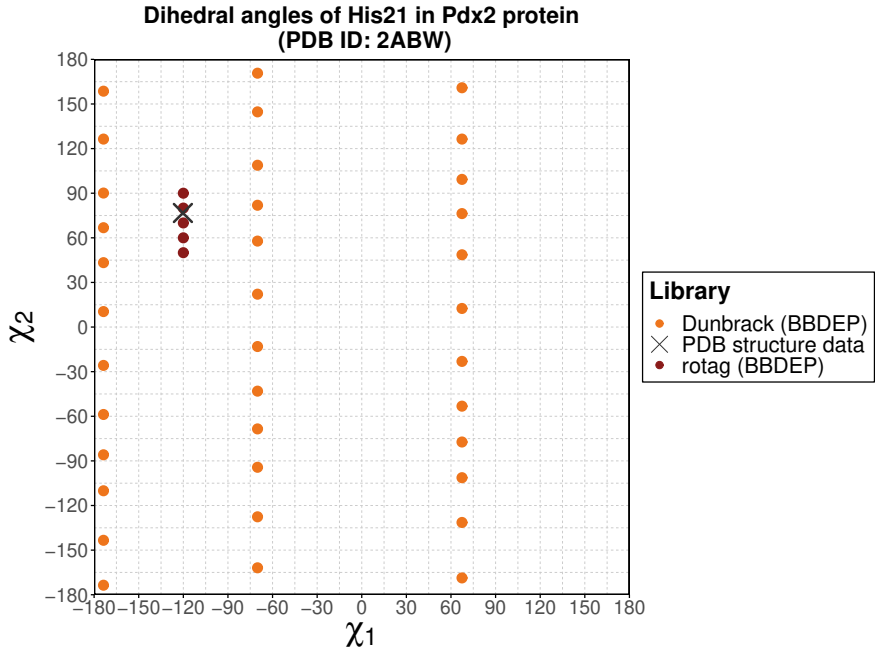


DEE $\longrightarrow$





$\longleftrightarrow$



Algirdas Grybauskas, manuscript under review

Formal proofs of the the algorithms in use

[Petrauskas et al. (2022)]

Require: H – a subgroup of a finite group G
Require: g – an element of the finite group G, $g \in G$
Ensure: The list L of the operators of a subgroup $L \leq G$ without duplicates
Ensure: L contains both g and the elements of H

1: **procedure** SIMPLEBUILDER(H, g)
▷ Build a space group generated by H and g
2: $L \leftarrow [e, h_1, h_2, \ldots, h_n]$, where $\forall i.\ h_i \in H$
3: $L_{\text{new}} \leftarrow [g]$
4: **while** L_{new} is not empty **do**
5: $g' \leftarrow \text{head}(L_{\text{new}})$
6: $L_{\text{new}} \leftarrow \text{tail}(L_{\text{new}})$
7: $L \leftarrow \text{append}(L,\ g')$
8: **for all** $h' \in L$ **do**
9: $g'' \leftarrow h' \otimes g'$
10: **if** $g'' \notin L \cup L_{\text{new}}$ **then**
11: $L_{\text{new}} \leftarrow \text{append}(L_{\text{new}},\ g'')$
12: **end if**
13: **end for**
14: **end while**
15: **return** L
16: **end procedure**

Figure 2
The optimized simple space-group-builder (core) algorithm.

```
1: have "subgroup R G"
2: proof -
3:     have R_subset: "R ⊆ carrier G" sorry
4:     moreover have R_m_closed: "⋀x y. ⟦x ∈ R; y ∈ R⟧ ⟹ x ⊗ y ∈ R" sorry
5:     moreover have R_one_closed: "1 ∈ R" sorry
6:     moreover have R_m_inv_closed: "⋀x. x ∈ R ⟹ inv x ∈ R" sorry
7:     ultimately show "subgroup R G" by (simp add: subgroup_def)
8: qed
```

Group theory in Ada/SPARK

examples/group_theory.ads

```
pragma Ada_2022;
pragma Spark_Mode (On);

generic
   type Element is private;
   Identity : Element;
   with function "*" (E, F: Element) return Element is <>;
```

```
   function Is_Closed_On_Multiplication (G : Group) return Boolean
   is (for all E of G =>
         (for all F of G => (Belongs_To (E*F, G))))
     with Ghost;

   function All_Elements_Have_Inverses (G : Group) return Boolean
   is (for all E of G => Has_Inverse (E, G))
     with Ghost;

   function Is_Group (G : Group) return Boolean
   is (Has_Identity (G) and then
         All_Elements_Have_Inverses (G) and then
         Is_Closed_On_Multiplication (G)
      )
     with Ghost;
```

Automatic compilation of proven code

Ada and SPARK

examples/make_group.ads

```
8     type Ring_Element is mod 37;
```

```
29    function Build_Group (E : Ring_Element) return Group
30      with
31      Post => Is_Group (Build_Group'Result);
```

gnatprove -P main.gpr --report=all make_group.adb

```
make_group.ads:23:14: info: postcondition proved
make_group.ads:27:14: info: postcondition proved
make_group.ads:31:14: info: postcondition proved
group_theory.ads:16:15: info: postcondition proved, in instantiation at make_group.ads:16
```

```
saulius@tasmanijos-velnias spacegroups/ $ ./run_make_group 8
(1, 8, 27, 31, 26, 23, 36, 29, 10, 6, 11, 14)

saulius@tasmanijos-velnias spacegroups/ $ ./run_make_group 7
(1, 7, 12, 10, 33, 9, 26, 34, 16)
```

Protein-Ligand Binding Database

Collaboration with BVTS (D. Matulis)

https://plbd.ibt.lt/

Crystal structures | main | Home

▸ Show table explanation

id | = | new | append | within | Filter

▶ Data download and upload panel

127 records. Up to 100 records per page | Prev | Page 1 of 2 | Clear filter | Next

ID	Label	PDB ID	Performed by	Devi...	Cmpd batch	Protein batch	Resolution [Å]	R_{cryst}	R_{free}	Compl. [%]	$N_{me...}$	Reliabili...	Notes	DB Revision
1	XS-SG037	5OGO	250		CB-00510 (VD10-39a; VD10-39b) (id = 2297)	PB-AM0010 (Carbonic anhydrase 1; chCA I; chCAI) (id = 64)	0.99	0.124	0.148			70	Generated automatically by Id: crystal_structure_table.com 16561 2022-05-19 16:25:43Z saulius	2022-11-16 (id = 3095)

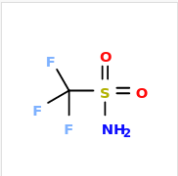
Intrinsic parameters | main | Home

▸ Show table explanation

id | = | new | append | within | Filter

▶ Data download and upload panel

1832 records. Up to 100 records per page | Prev | Page 1 of 19 | Clear filter | Next

IDs	Compound	Structure	Protein	Primary group	$K_{b,int}$ [M^{-1}]	$\Delta G_{b,int}$ [kJ/mol]	$\sigma_{\Delta G_{b,int}}$ [kJ/mol]	$\Delta H_{b,int}$ [kJ/mol]	$\sigma_{\Delta H_{b,int}}$ [kJ/mol]	FTSA source	ITC source	T [°C]	$Q_{ITC_{par}...}$	$Q_{i...}$	$N_{i...}$	DB Revision
13962-13962--16839-18154	TFMSA; TFS (id = 2258)		Carbonic anhydrase 13; CA XIII (id = 19)	PG-0011	3e+07	-44	0	-42	0.042	13962	16839 18154	37.0	90	90	3	2022-05-16 (id = 2933)

Coordinate and unit cell relaxation using `Quantum Espresso`:

COD ID	Formula	Conv.	Problem	Lat. diff. %			Bandgap		
				a	b	c	eV	Matgen	type
1527735	BaO	No	Elec. loop					2.773	Indirect
1528545	LiSbSr	Yes		0.69	0.39	0.37	0.671	0.681	Direct
1530141	Br5PbTl3	Yes		-0.10	-0.10	0.23	2.874	2.907	Direct
1535922	Ca2N	Yes		1.21	1.21	4.53	0.045	0.000	Metal
1537335	K3P	Yes		0.43	0.43	0.78	0.235	0.214	Indirect
1539138	LiORb	Yes		0.53	0.22	2.56	2.289	2.257	Indirect
...									
2013551	I2Mg	Yes		1.21	1.21	7.82	3.620	3.677	Indirect
2207375	Na3P	Yes		-0.76	-0.76	-0.57	0.501	0.403	Direct
4124784	AlO	Yes		-26.68	-26.68	-26.68	0.052	0.000	Metal
4320809	ClNa	Yes		-0.20	-0.20	-0.20	5.078	5.145	Direct
4344366	Na2S	Yes		-0.59	-0.59	-0.59	2.531	2.440	Direct
7200689	OSr	Yes		0.64	0.64	0.64	3.321	3.449	Indirect

A. Vaitkus (structure selection), V. Žalandauskas (QM scripts, computations)

Papers, funding and other activity

Publications:

- 4 manuscripts in preparation (2 with KICIS as the main contributor);
- 5 manuscripts published (2 with KICIS as the main contributor);

Grant applications:

- 6 applications attempted;
- 4 grants received ("Gilibert" with S. Grudinin (S.G.); CECAM OPTIMADE w/s funding (S.G.); RCoL S-MIP-23-87 and VU Young Scientist Grant (A.M.);

Other activities:

- Vilnius University Open Science Policy Workgroup (S.G.);
- Debian package maintainer (A.M.)

Acknowledgements

VU LSC IBT (KICIS)

Andrius Merkys
Antanas Vaitkus
Algirdas Grybauskas

VU LSC IBT (BVTS)

Daumantas Matulis
Vytautas Petrauskas
Darius Lingė
Marius Gedgaudas

VU LSC IBT (BNSTS)

Mindaugas Zaremba
Elena Manakova

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

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

References I

Balandis B, Šimkūnas T, Paketurytė-Latvė V, Michailovienė V, Mickevičiūtė A, Manakova E, et al. (2022) Beta and gamma amino acid-substituted benzenesulfonamides as inhibitors of human carbonic anhydrases. Pharmaceuticals 15(4):477, DOI 10.3390/ph15040477, URL https://doi.org/10.3390/ph15040477

Jozeliūnaitė A, Rahmanudin A, Gražulis S, Baudat E, Sivula K, Fazzi D, et al. (2022) Light-responsive oligothiophenes incorporating photochromic torsional switches. Chemistry – A European Journal DOI 10.1002/chem.202202698, URL https://doi.org/10.1002/chem.202202698

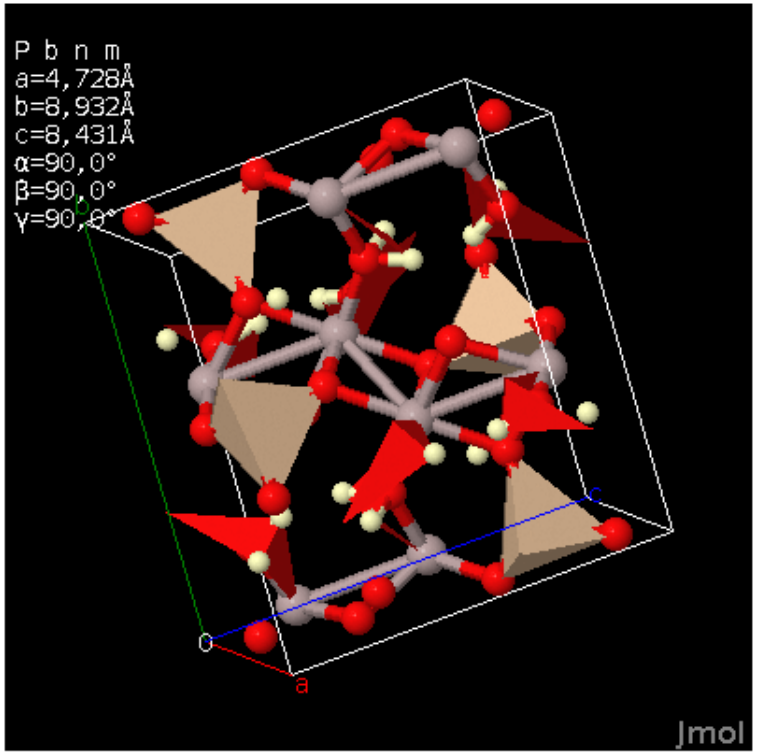
Manakova E, Golovinas E, Pocevičiūtė R, Sasnauskas G, Grybauskas A, Gražulis S, et al. (2022) Structural basis for sequence-specific recognition of guide and target strands by the archaeoglobus fulgidus argonaute protein. Research Square pp 1–17, DOI 10.21203/rs.3.rs-2305454/v1, URL https://doi.org/10.21203/rs.3.rs-2305454/v1

References II

Merkys A, Vaitkus A, Grybauskas A, Konovalovas A, Quirós M, Gražulis S (2023) Graph isomorphism-based algorithm for cross-checking chemical and crystallographic descriptions. Journal of Cheminformatics 15(1), DOI 10.1186/s13321-023-00692-1, URL https://doi.org/10.1186/s13321-023-00692-1

Petrauskas K, Merkys A, Vaitkus A, Laibinis L, Gražulis S (2022) Proving the correctness of the algorithm for building a crystallographic space group. Journal of Applied Crystallography 55(3):515–525, DOI 10.1107/s1600576722003107, URL https://doi.org/10.1107/s1600576722003107

Thank you!



Coordinates	2207377.cif
Original IUCr paper	HTML

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

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

https://www.crystallography.net/cod/archives/2023/talks/LSC/slides.pdf