Geometrinių savybių aptikimas atviroje mažų molekulių duomenų bazėje COD

Andrius Merkys, Antanas Vaitkus,
Mykolas Okulič-Kazarinas ir Saulius Gražulis

Vilnius, 2016

Vilniaus universiteto Biotechnologijos institutas

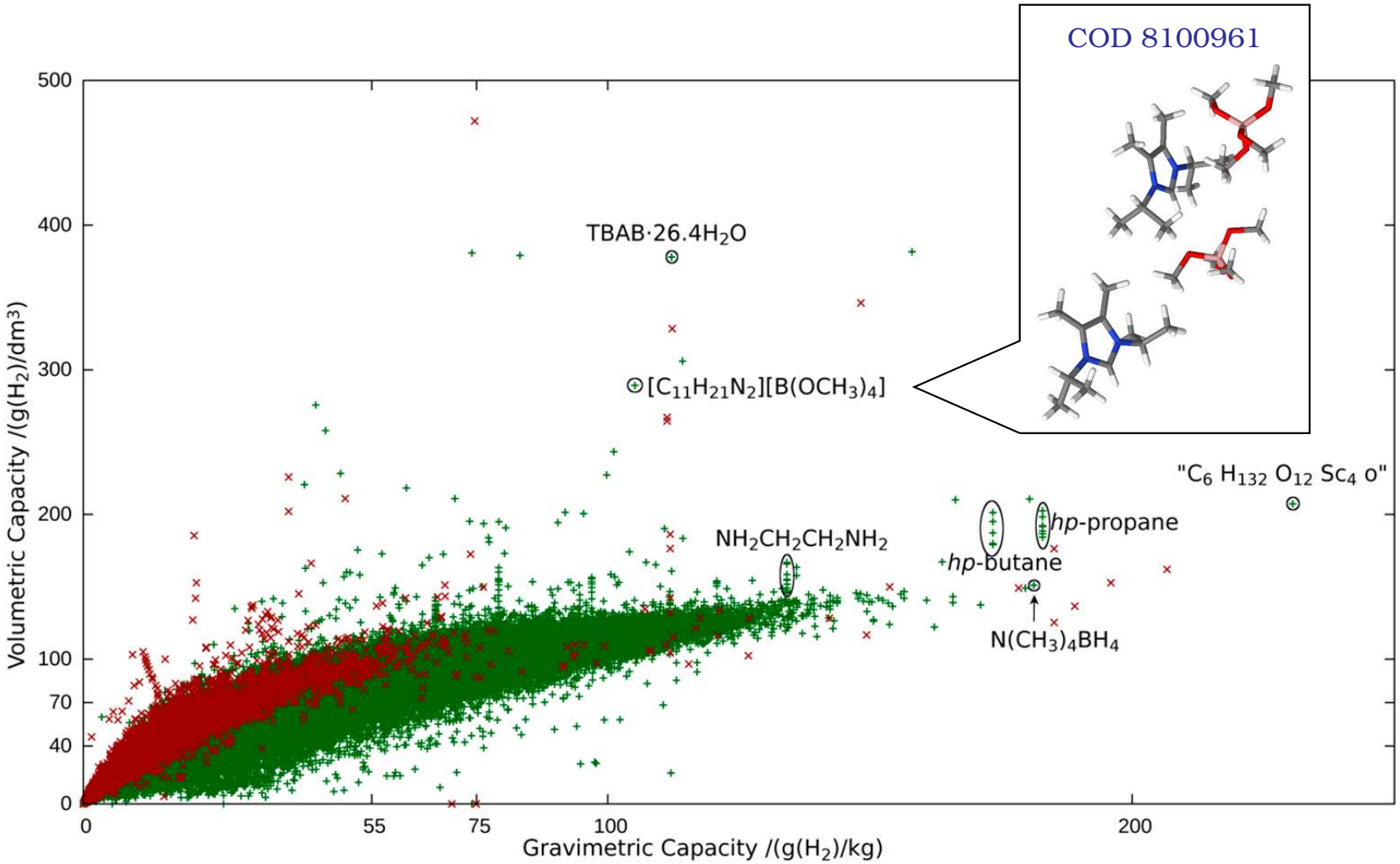
Struktūrinės mažų molekulių duomenų bazės

Crystallography Open Database (COD)	**Cambridge Structural Database (CSD)**	**Inorganic Crystal Structure Database (ICSD)**
350 000 įrašų	800 000 įrašų	180 000 įrašų
atvira	prenumeruojama	prenumeruojama

2016 m. vasario mėn. duomenys

Vandenilio talpa kristaluose

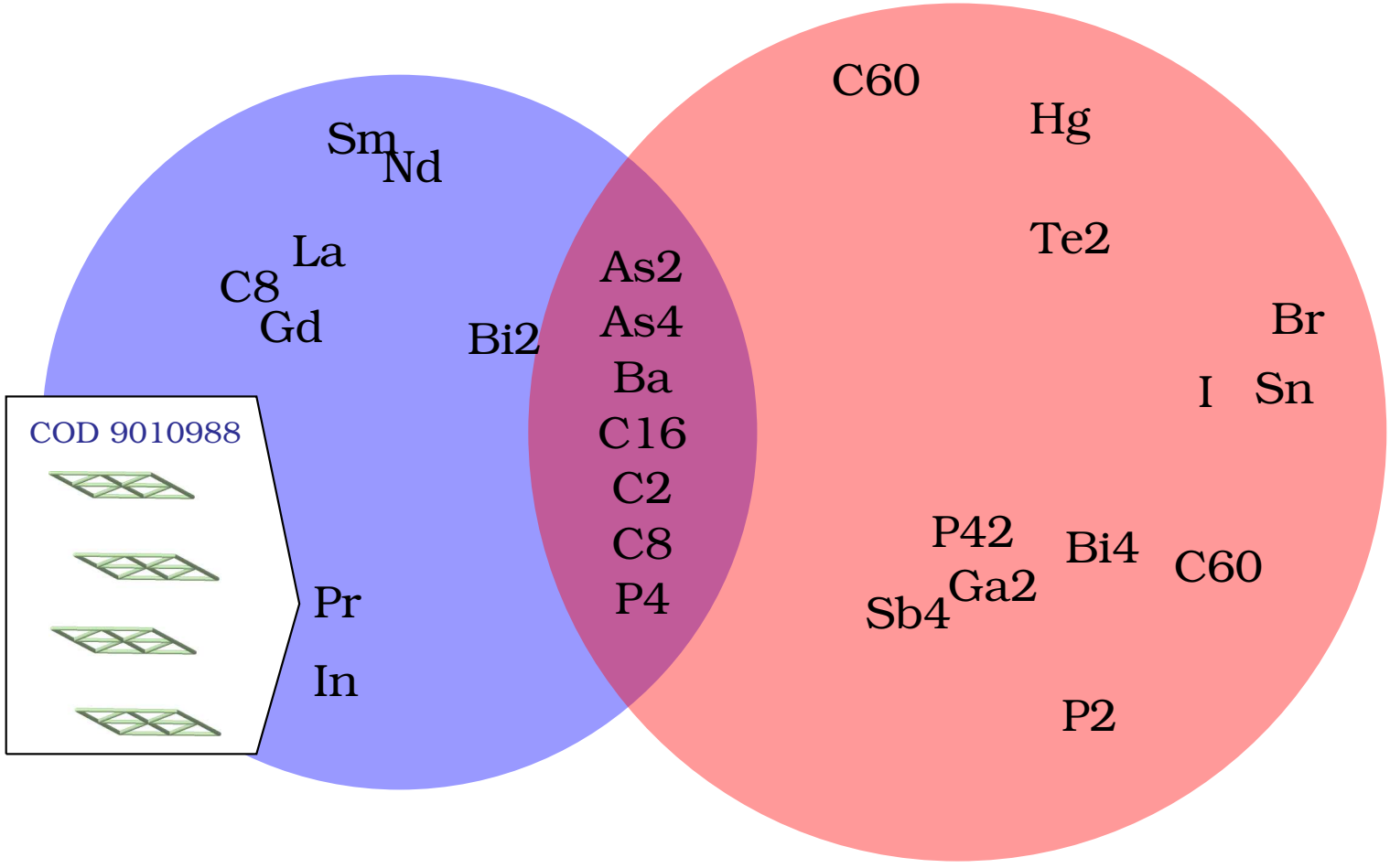
in silico tyrimas COD duomenų bazėje



Breternitz ir Gregory (2015)

Ar grafenas tik vienas toks?

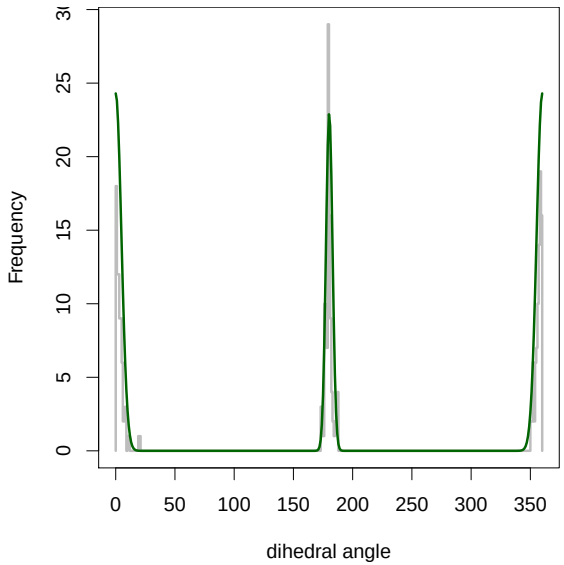
Dvimačių alotropų paieška COD ir ICSD duomenų bazėse

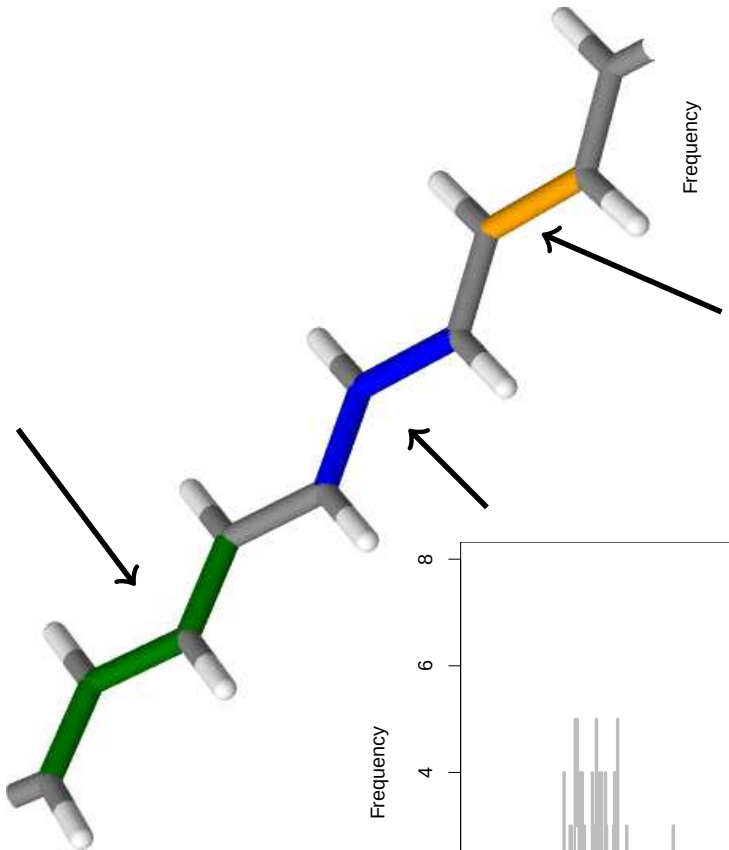


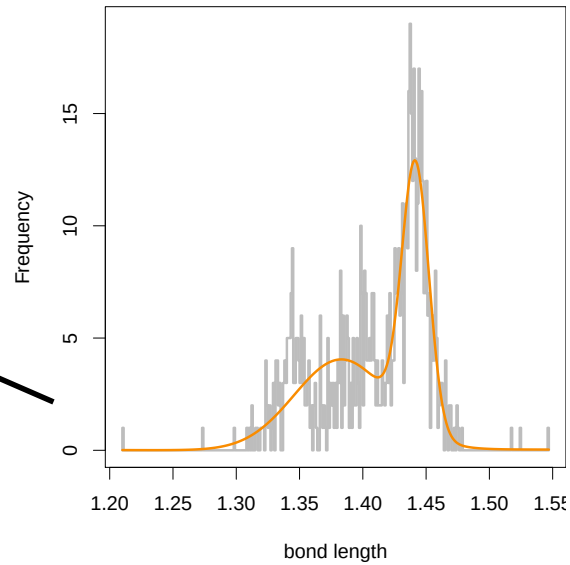
COD (15 vnt.) ICSD (19 vnt.)

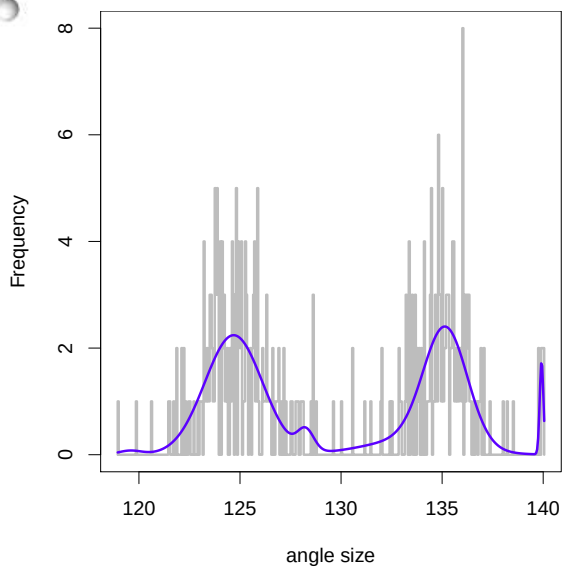
Mounet et al. (rengiama)

Molekulių geometrinės savybės



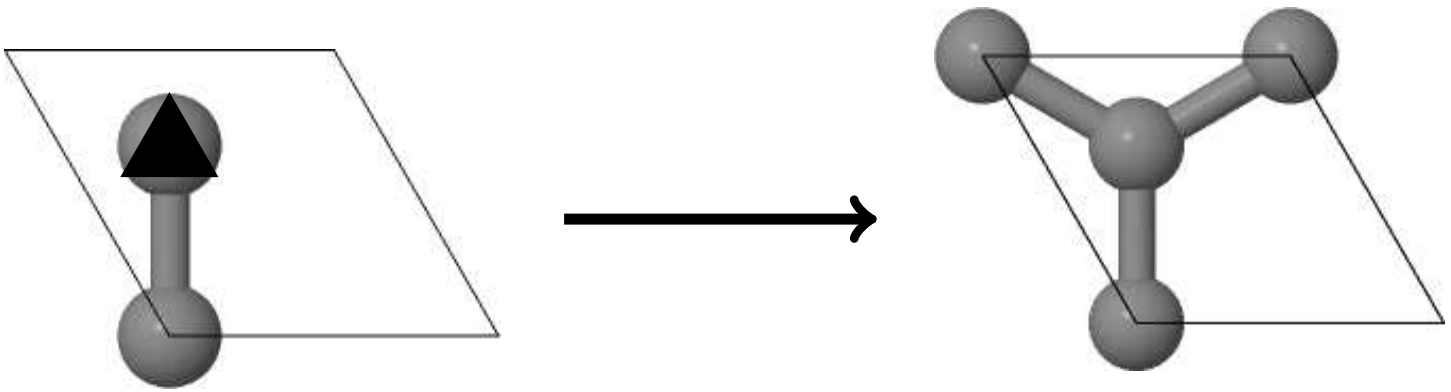






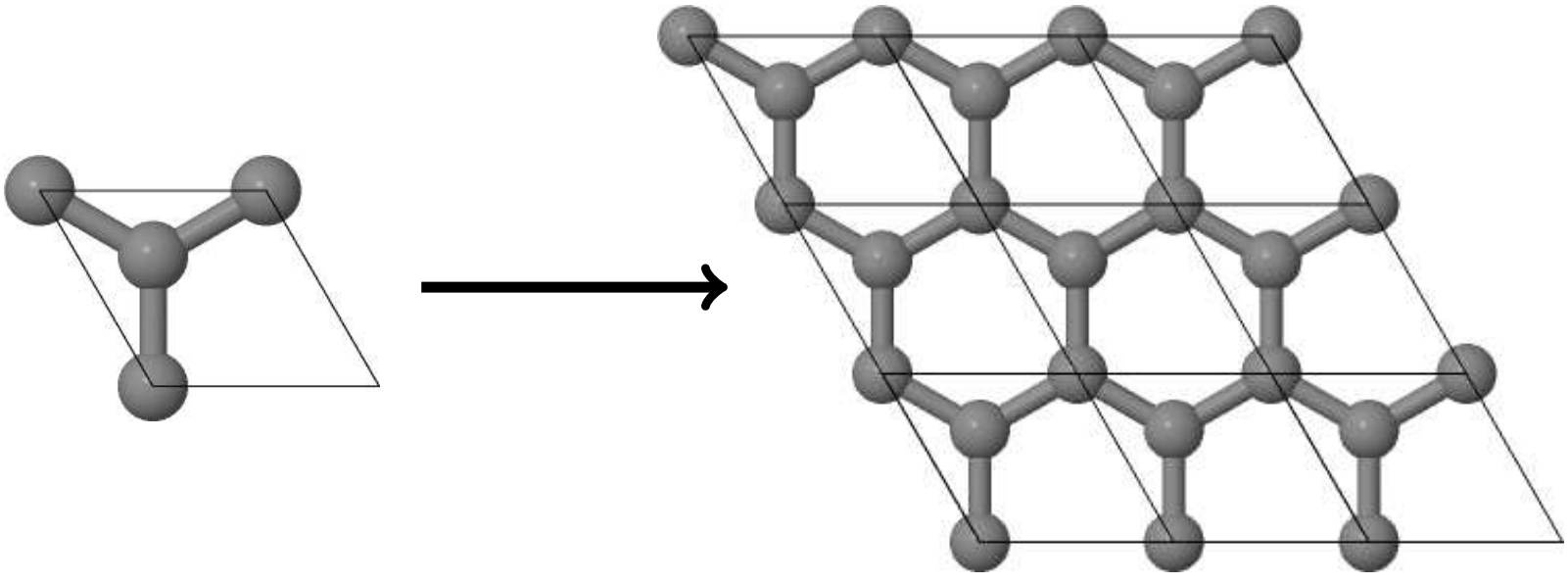
Kristalo simetrijos atstatymas

Pavyzdys: grafito sluoksnis, COD 1011060



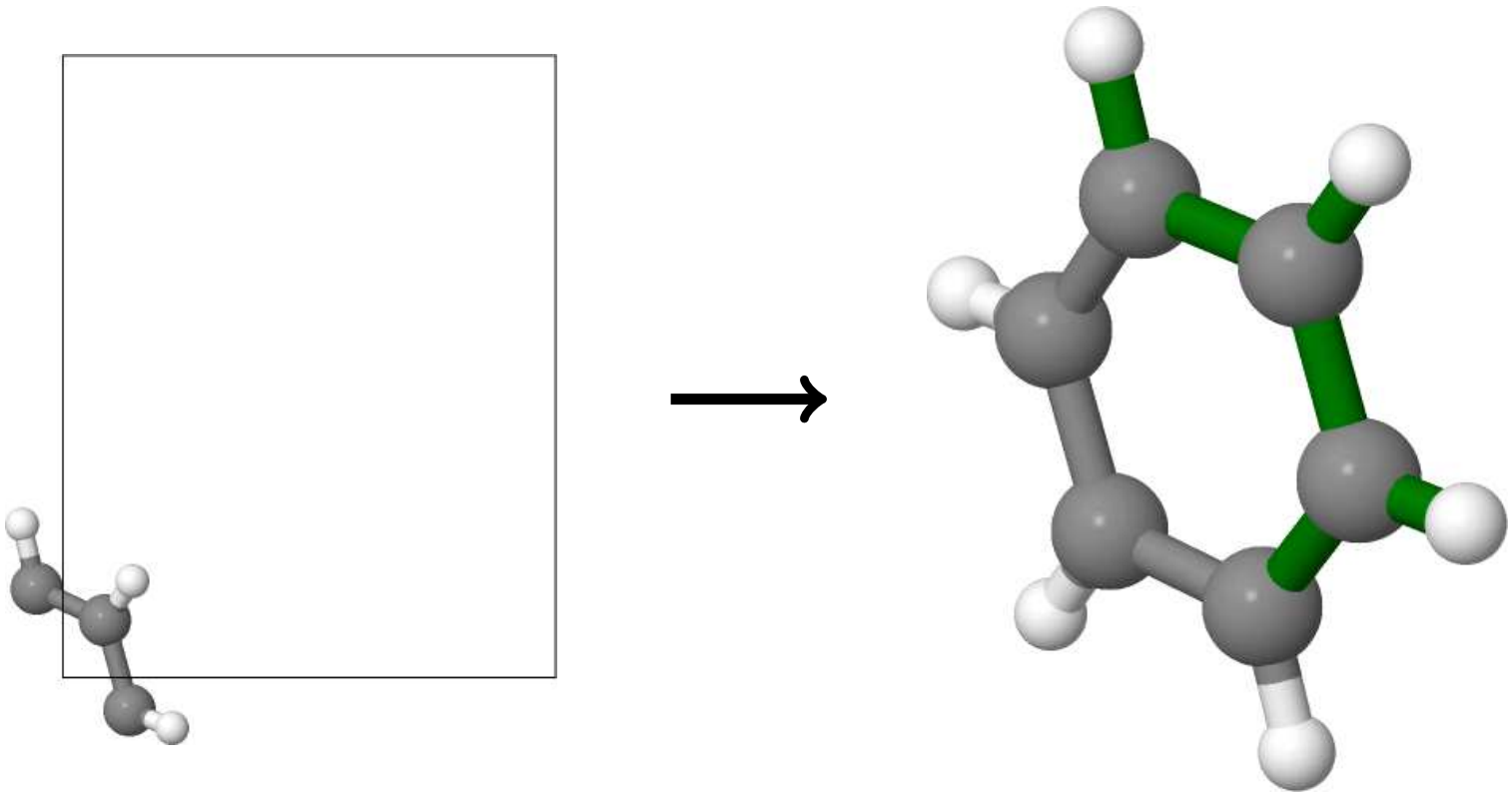
Tarpatominių jungčių paieška

Pavyzdys: grafito sluoksnis, COD 1011060

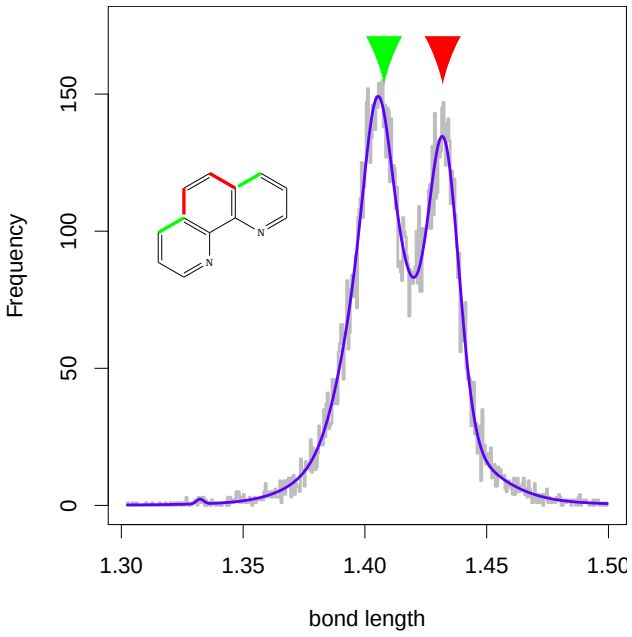


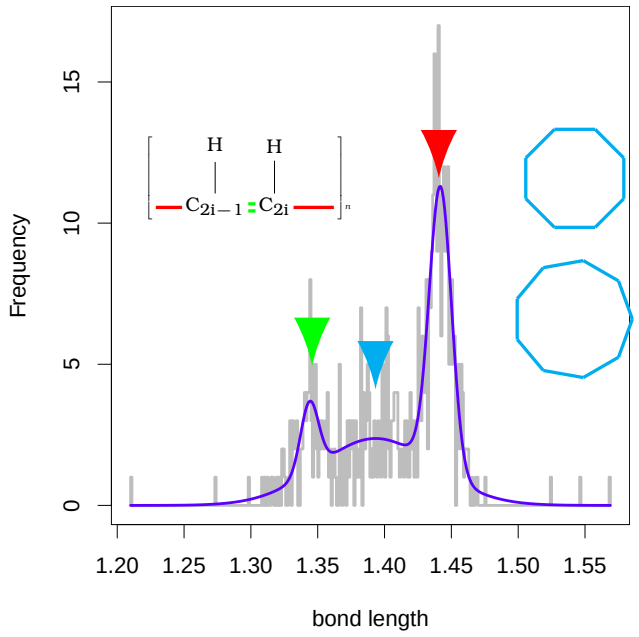
Simetriškai ekvivalenčių stebinių atmetimas

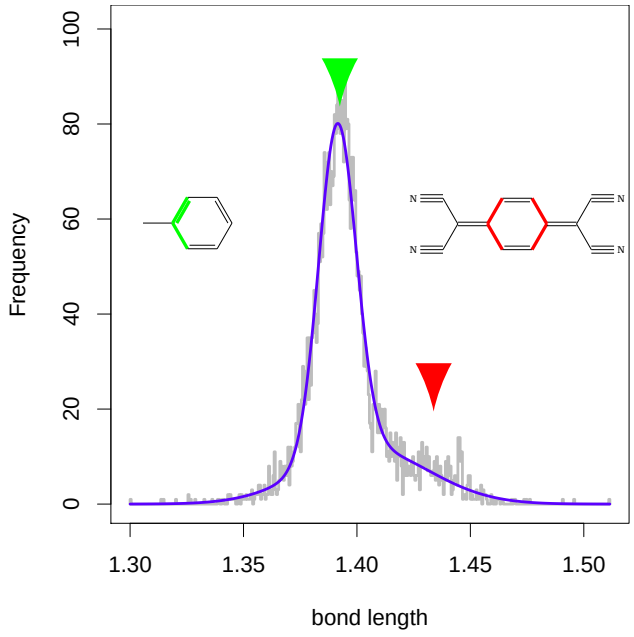
Pavyzdys: benzeno molekulė, COD COD 2100348

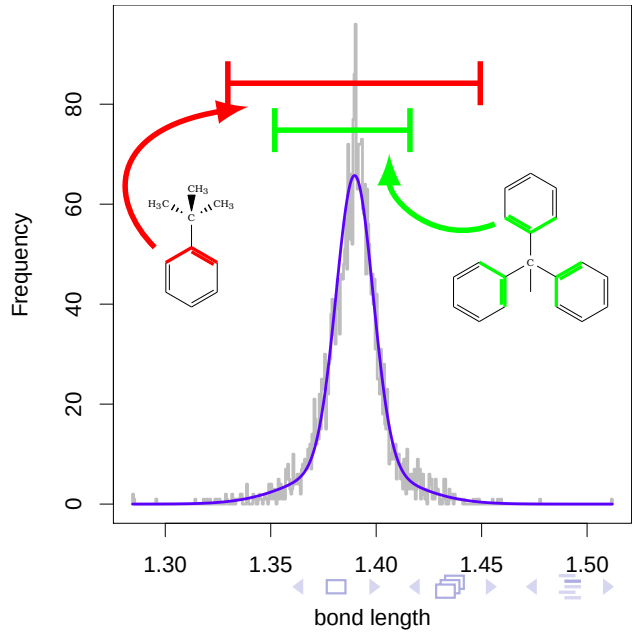


Stebinių imčių aprašymas









COD geometrijos naršyklė

http://www.crystallography.net/geometry

Search

order by class name (ascending)

order by observation count (descending)

To search for bond models, please enter one to two COD atom class names separated by ';'.

Examples:

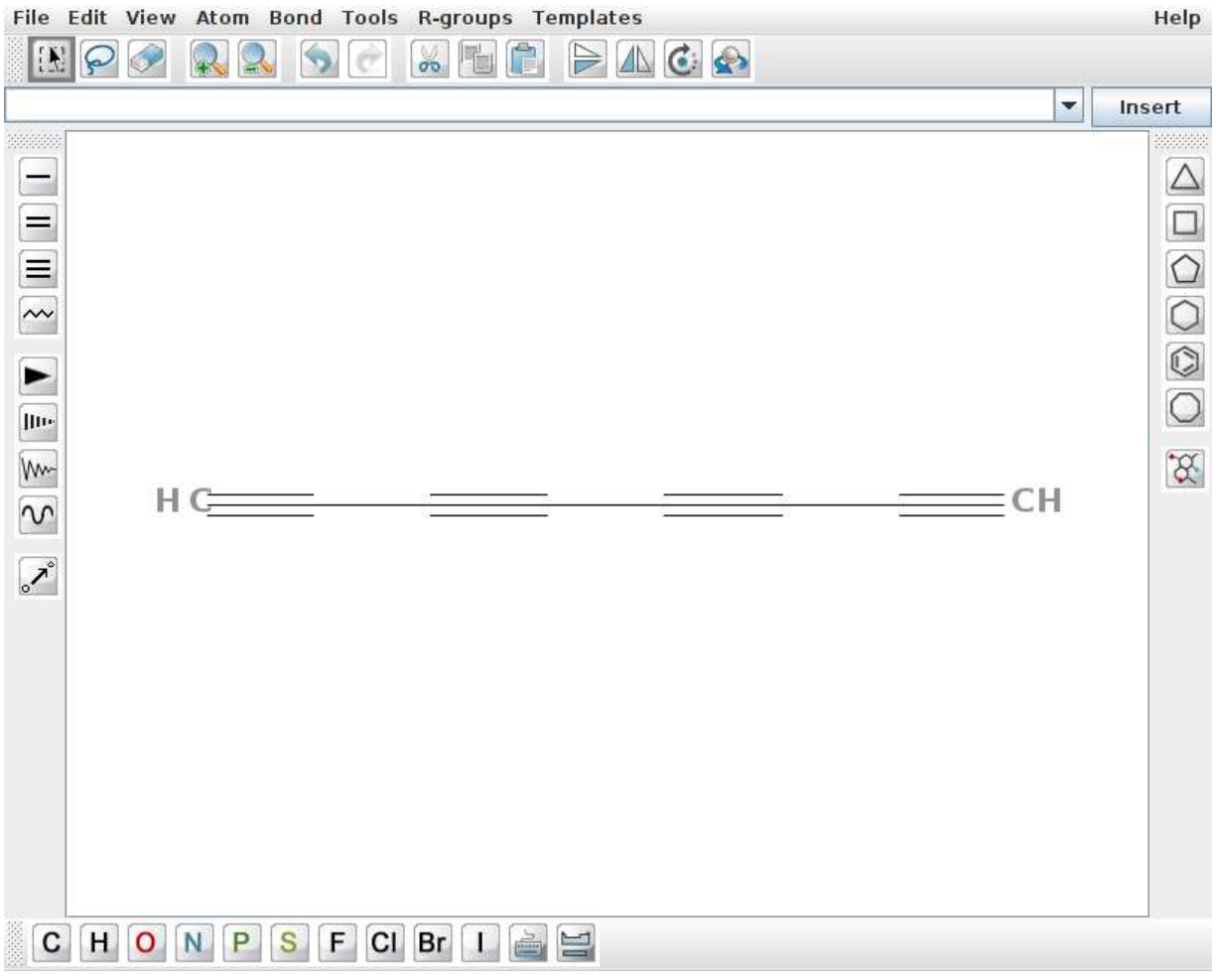
- C(CC)2;C(CC)2
- c(cCH)2(H);c(cCH)2(H)

To search for angle models, please enter three COD atom class names separated by ';'.

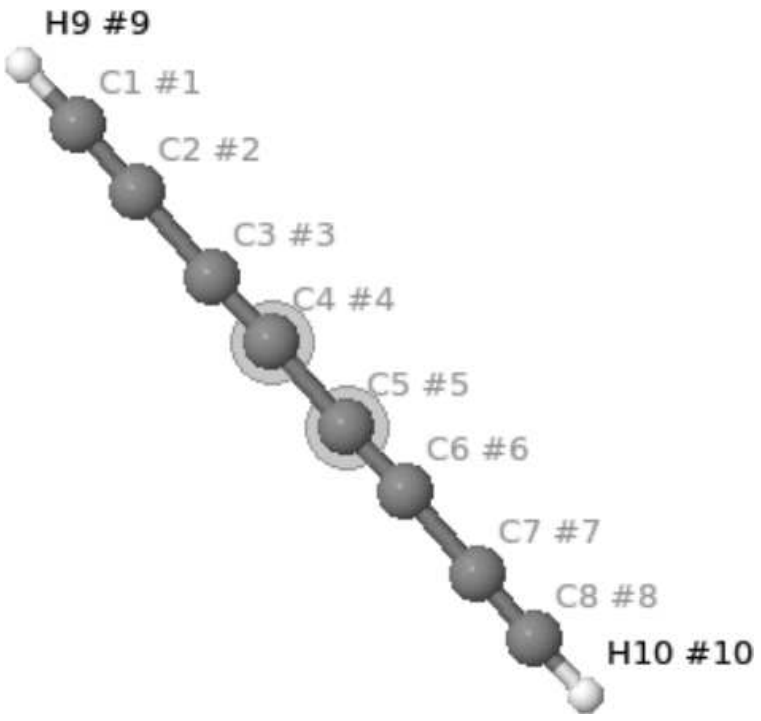
To search for dihedral angle models, please enter four COD atom class names separated by ';'.

Percent sign (%) can be used as a wildcard.

Paieška: poliino anglis-anglis jungties ilgis



Dõminančių deskriptorių pasirinkimas



Download structure

- [ ] C1, C(CC)(H)
- [ ] C2, C(CC)(CH)
- [ ] C3, C(CC)2
- [x] C4, C(CC)2
- [x] C5, C(CC)2
- [ ] C6, C(CC)2
- [ ] C7, C(CC)(CH)
- [ ] C8, C(CC)(H)
- [ ] H9, H(CC)
- [ ] H10, H(CC)

Submit Back

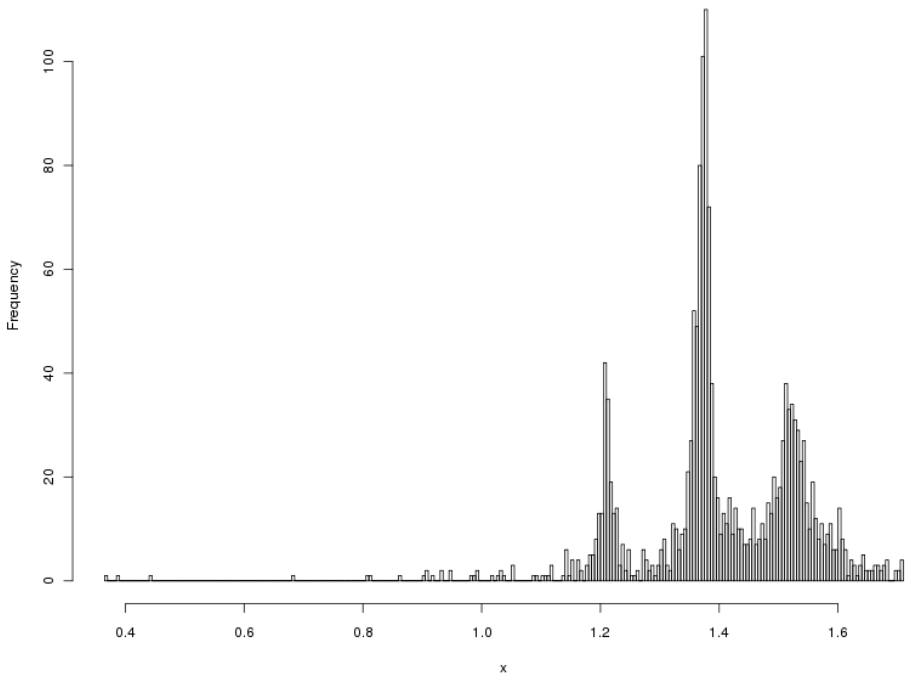
Poliinų jungtys COD duomenų bazėje

Search

order by class name (ascending)

order by observation count (descending)

Distribution of C(CC)2 - C(CC)2 bonds



Breaks: 269 Min: Max: Histogram width: 800 Display

Warning: models for this class are scheduled for recalculation, out-of-date models are presented here for reference only.

	Distribution	Peaks	BIC	Log-likelihood
○	cauchy	3	-2983.63250734474	1521.29969433851
○	cauchy	4	-2972.23756189253	1526.65851186221
○	cauchy	5	-2954.17602788954	1528.68403511053
○	cauchy	6	-2945.47851368168	1535.3915682564
○	cauchy	7	-2931.49848709219	1539.45784521146
○	cauchy	8	-2931.13696008285	1550.3333719566
○	cauchy	9	-2911.97656549846	1551.80946491421
○	cauchy	10	-2903.18311848377	1558.46903165667
○	cauchy	2	-2306.60210414815	1171.72820249042
○	gaussian	4	-2174.55991783404	1127.81968983297
○	gaussian	5	-2164.18853401055	1133.69028817103
○	gaussian	2	-1955.22813260451	996.041216718596
○	gaussian	3	-1947.4479559367	1003.2074186345
○	gaussian	1	-1773.73018669565	894.235953514359
○	cauchy	1	-1556.82925243875	785.785486385914
○	gaussian	6	NaN	NaN
○	gaussian	7	NaN	NaN
○	gaussian	8	NaN	NaN
○	gaussian	9	NaN	NaN
○	gaussian	10	NaN	NaN

☐ Show individual mixture components

Show Sort by BIC Sort by Log-likelihood

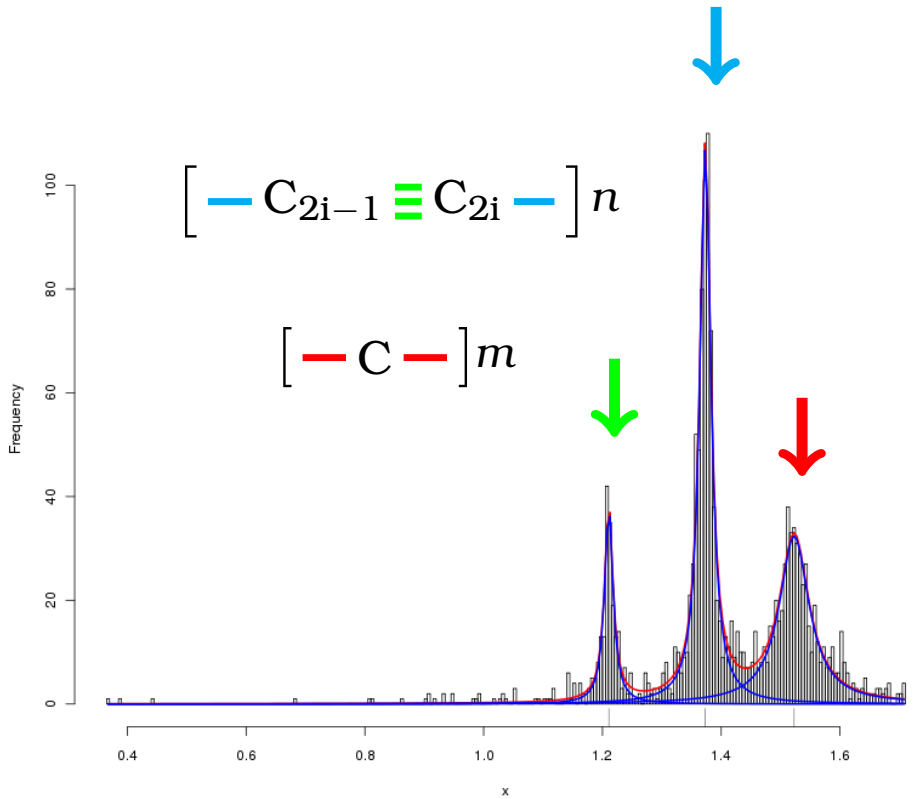
Poliinų jungtys COD duomenų bazėje

Search

order by class name (ascending)

order by observation count (descending)

Distribution of C(CC)2 - C(CC)2 bonds



Breaks: 269 Min: Max: Histogram width: 800 Display

Warning: models for this class are scheduled for recalculation, out-of-date models are presented here for reference only.

Distribution	Peaks	BIC	Log-likelihood
cauchy	3	-2983.63250734474	1521.29969433851
cauchy	4	-2972.23756189253	1526.65851186221
cauchy	5	-2954.17602788954	1528.68403511053
cauchy	6	-2945.47851368168	1535.3915682564
cauchy	7	-2931.49848709219	1539.45784521146
cauchy	8	-2931.13696008285	1550.3333719566
cauchy	9	-2911.97656549846	1551.80946491421
cauchy	10	-2903.18311848377	1558.46903165667
cauchy	2	-2306.60210414815	1171.72820249042
gaussian	4	-2174.55991783404	1127.81968983297
gaussian	5	-2164.18853401055	1133.69028817103
gaussian	2	-1955.22813260451	996.041216718596
gaussian	3	-1947.4479559367	1003.2074186345
gaussian	1	-1773.73018669565	894.235953514359
cauchy	1	-1556.82925243875	785.785486385914
gaussian	6	NaN	NaN
gaussian	7	NaN	NaN
gaussian	8	NaN	NaN
gaussian	9	NaN	NaN
gaussian	10	NaN	NaN

Show individual mixture components

Show Sort by BIC Sort by Log-likelihood

Tolimesni planai

- Atvira prieiga prie duomenų MySQL protokolu
- Periodiškas duomenų bazės atnaujinimas
- Struktūrų-išskirčių pašalinimas iš imties

Padėkos

VU Biotechnologijos institutas

Virginijus Šikšnys
(*skyriaus vedėjas*)

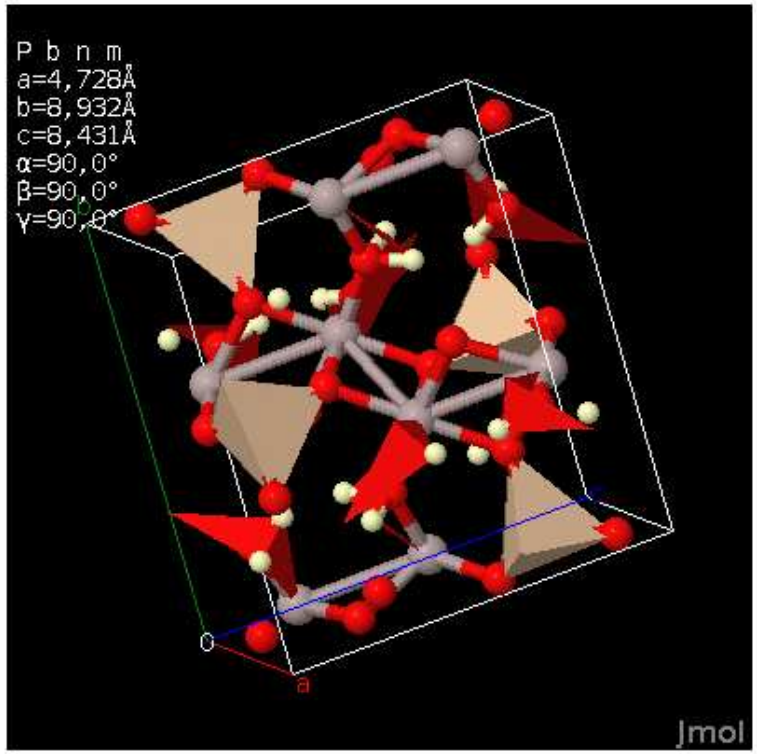
Saulius Gražulis
Antanas Vaitkus
Mykolas Okulič-Kazarinas

COD Advisory Board

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

Ačiū už dėmesį!

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



Coordinates	2207377.cif
Original IUCr paper	HTML

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