Derivation of a Covalent Radii Table From Open Crystallographic Data

Eglė Šidlauskaitė[a], Antanas Vaitkus[b], Andrius Merkys[b]

[a]Faculty of Medicine, Vilnius University, Vilnius, Lithuania
[b]Sector of Crystallography and Chemical Informatics, Institute of Biotechnology, Life Sciences Center, Vilnius University, Vilnius, Lithuania

OPEN READINGS 2023

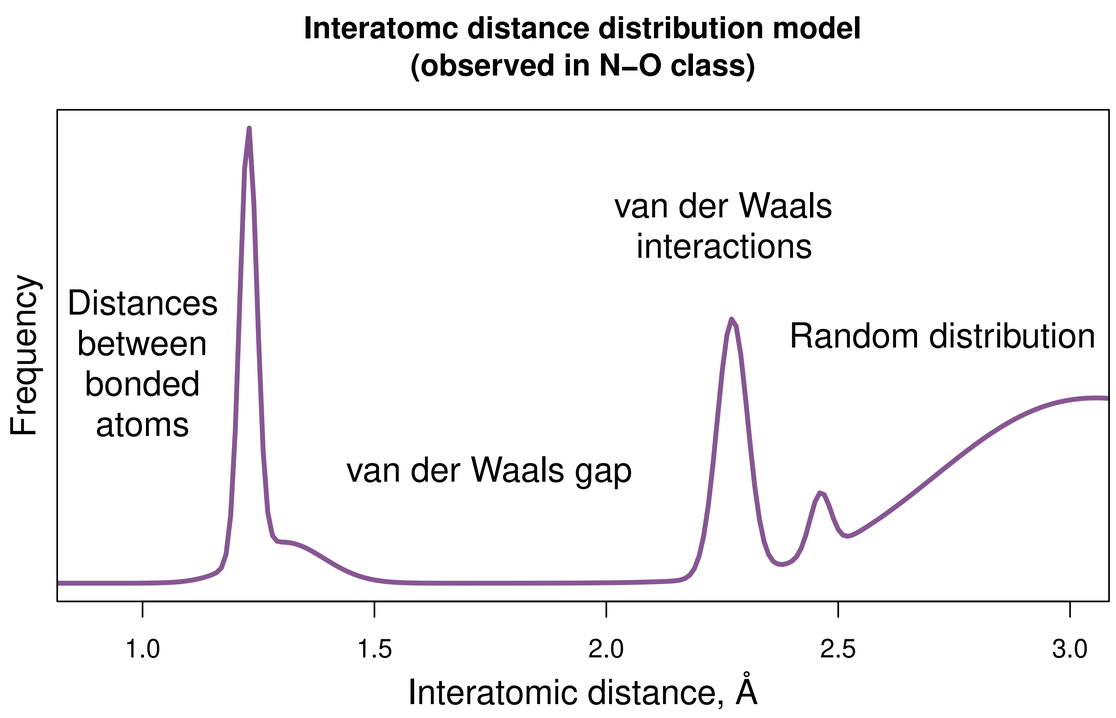
Introduction

X-ray crystallography is able to determine the exact positions of atoms that constitute crystal structures [1], but cannot capture chemical bonding. Due to this, bonding is often determined using a heuristic which considers two atoms to be connected if the calculated distance between them is less or equal to the sum of their covalent radii. This distance is called a covalent bond length.

Motivation

- Most popular covalent radii tables [2, 3, 4] are derived using data from proprietary databases that restrict the usage and spread of such derivative datasets.
- There is a need to develop an independent workflow that allows covalent radii tables to be automatically calculated and ensures reproducibility of the results.

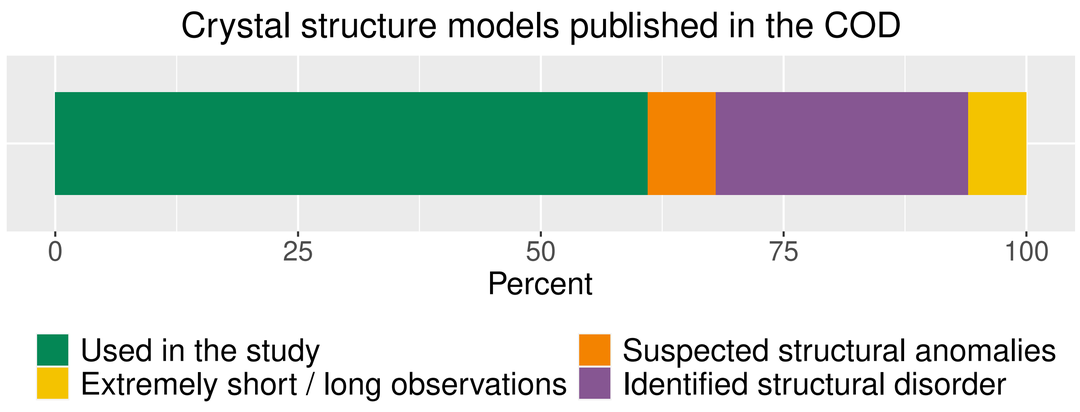
Interatomic distance distribution



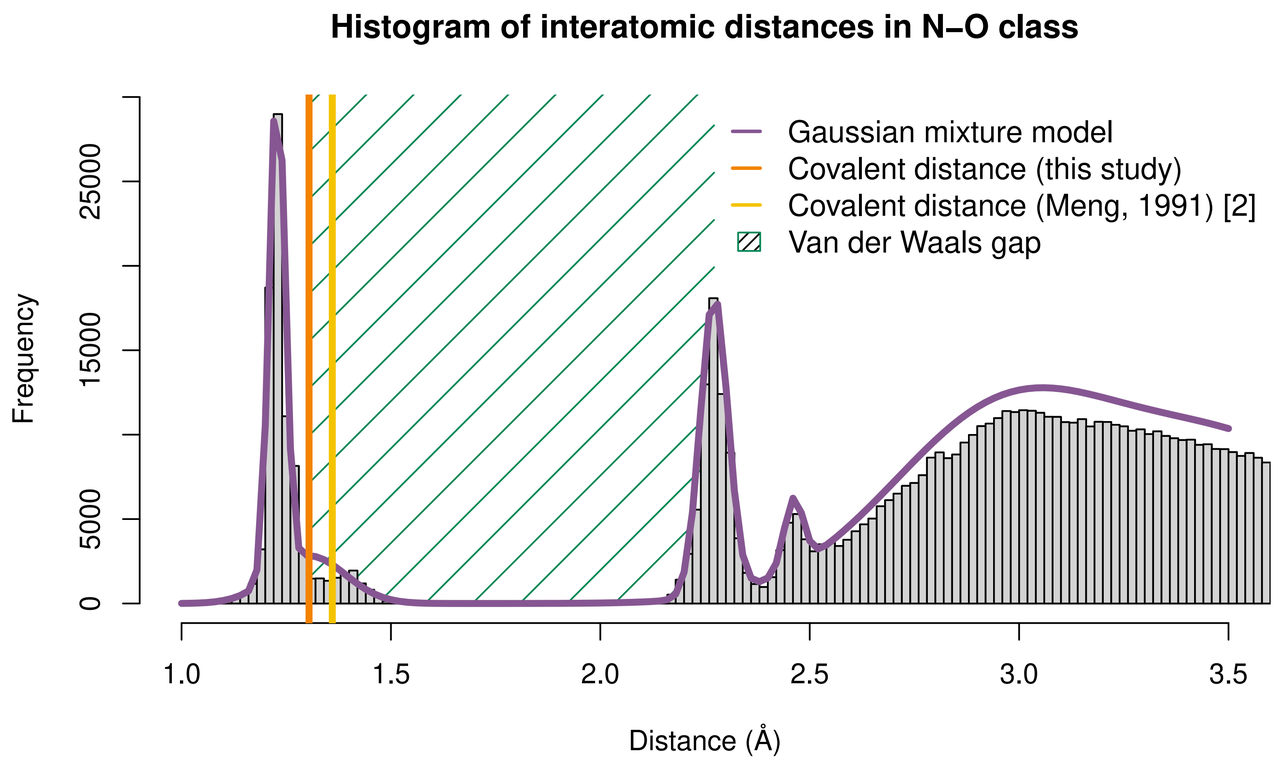
- All interatomic distance observations are classified based on chemical elements between which the distance was measured.
- Van der Waals gap is an interval that separates peaks formed by covalent interactions from those formed mainly by van der Waals interactions [5].
- Interatomic distance distribution can be approximated by a Gaussian mixture model [5].

Data and filtering

- 485 000 entries from the Crystallography Open Database (COD) [6] were analyzed.
- Crystallographic models need to be filtered to ensure that only realistic distances between atoms are used in the study.

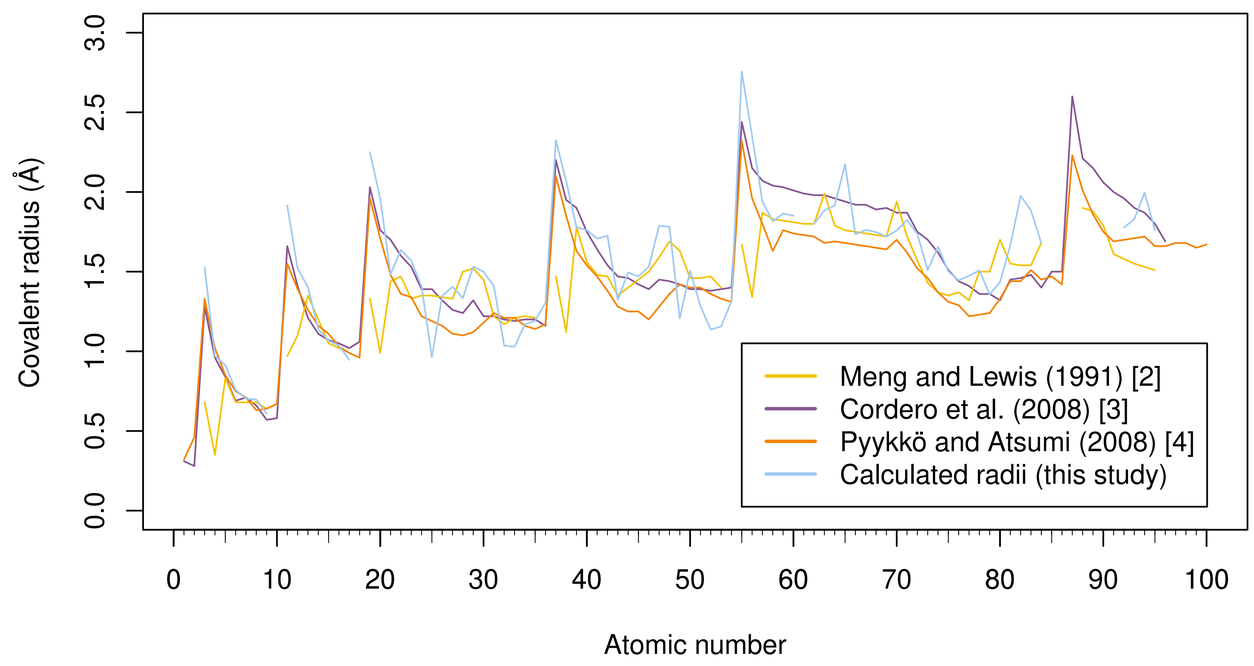


Derivation of covalent radii values



- Method uses the best fitting Gaussian mixture model per Bayesian Information Criterion [7].
- Each pair of neighbouring components is examined to find the lowest density region in the distribution, which corresponds to be the van der Waals gap.
- Mean of the component preceding the van der Waals gap is selected as the maximum covalent bond length for each atom distance class.

Derived covalent radii table



- Derived covalent radii values follow expected distribution trend based on the atomic number.

Conclusions

- Methodology for automated covalent radii derivation was developed.
- Derived covalent radii table is freely accessible online (via QR code on the left).
- Further research will be focused on testing the derived covalent radii through real molecule connectivity validation.

Bibliography

[1] Gražulis et al. Computing stoichiometric molecular composition from crystal structures. *Journal of Applied Crystallography*, 48(1):85–91, 2015.

[2] Meng et al. Determination of molecular topology and atomic hybridization states from heavy atom coordinates. *Journal of Computational Chemistry*, 12(7):891–898, sep 1991.

[3] Cordero et al. Covalent radii revisited. *Dalton Transactions*, 21:2832–2838, 2008.

[4] Pyykkö et al. Molecular Single–Bond Covalent Radii for Elements 1–118. *Chemistry—A European Journal*, 15(1):186–197, 2008.

[5] Alvarez. A cartography of the van der Waals territories. *Dalton Transactions*, 42:8617–8636, 2015.

[6] 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.

[7] Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, pages 461–464, 1978.

This project has received funding from Vilnius University Science Support Fund (agreement No. MSF-JM-15/2021) and Research Council of Lithuania (agreement No. S-MIP-23-87)

eglsidlauskaite@gmail.com