Derivation of a Covalent Radii Table From Open Crystallographic Data

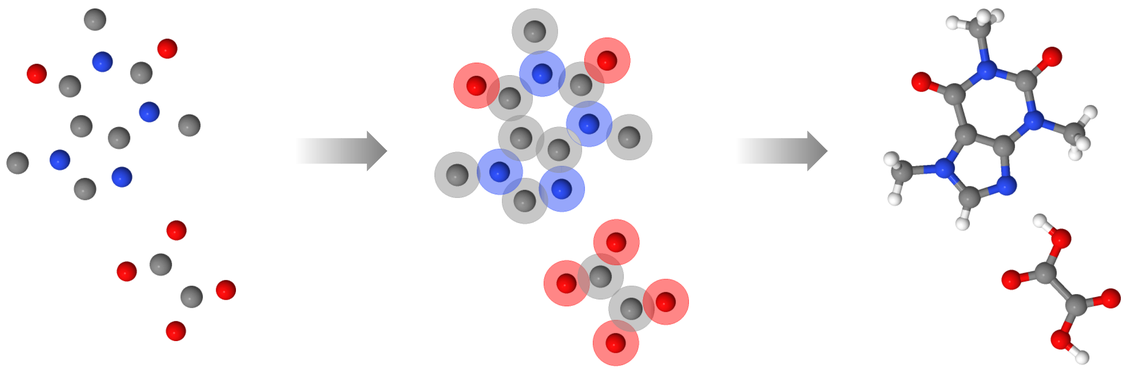
Eglė Šidlauskaitė[a], Antanas Vaitkus[a], Andrius Merkys[a]

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

OPEN READINGS 2022

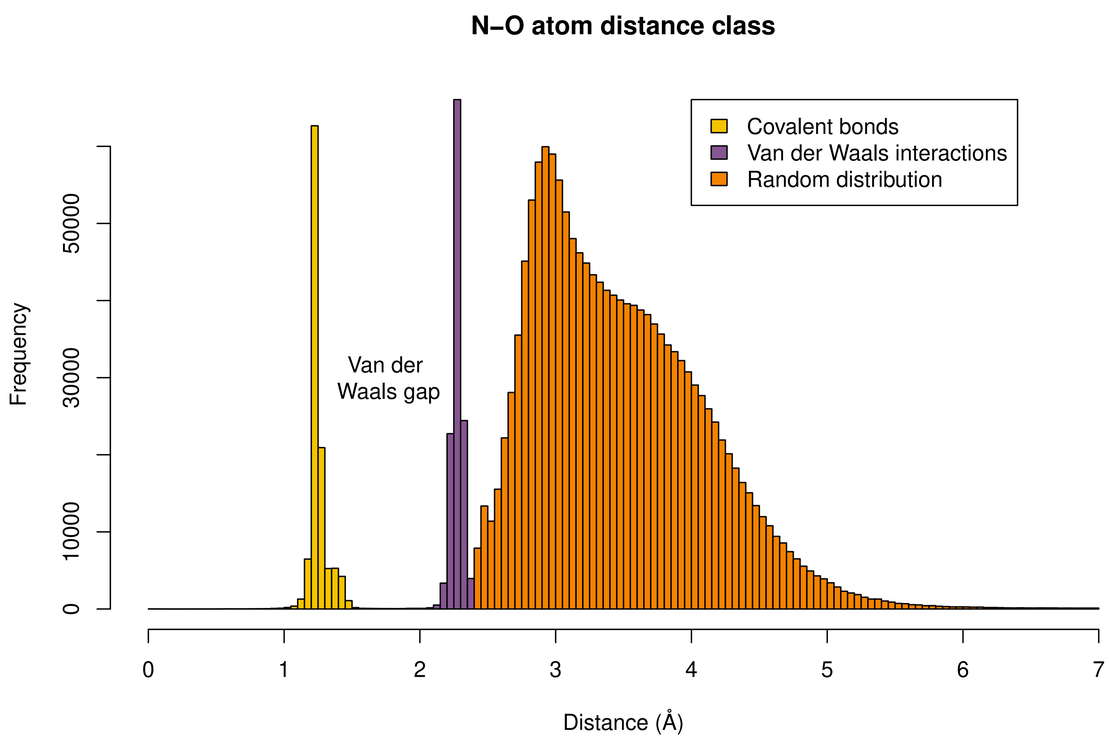
Introduction

X-ray crystallography is able to determine the exact positions of atoms that constitute crystal structures [1], but cannot capture chemical bonding. Two atoms are considered to be connected by a chemical bond if the calculated distance between them is lesser or equal to the sum of their covalent radii. This distance is called a covalent bond length.



- Most popular covalent radii tables [2] are derived using data from proprietary databases that restrict the usage and spread of such derivative datasets.
- There is a need to develop a completely independent workflow that allows covalent radii tables to be automatically calculated using open crystallographic datasets such as the Crystallography Open Database (COD) [3].

Interatomic distance distribution



- An interatomic distance distribution consists of all distances that were observed between atoms of two selected chemical types (e.g. C-O, O-O, N-O, etc.).
- Van der Waals gap is an interval within the distribution that separates peaks formed by covalent interactions from those formed mainly by van der Waals interactions. Normally, no interatomic distances are observed in this interval [4].
- Interatomic distance distribution can be approximated by a Gaussian mixture model.

Data from the COD

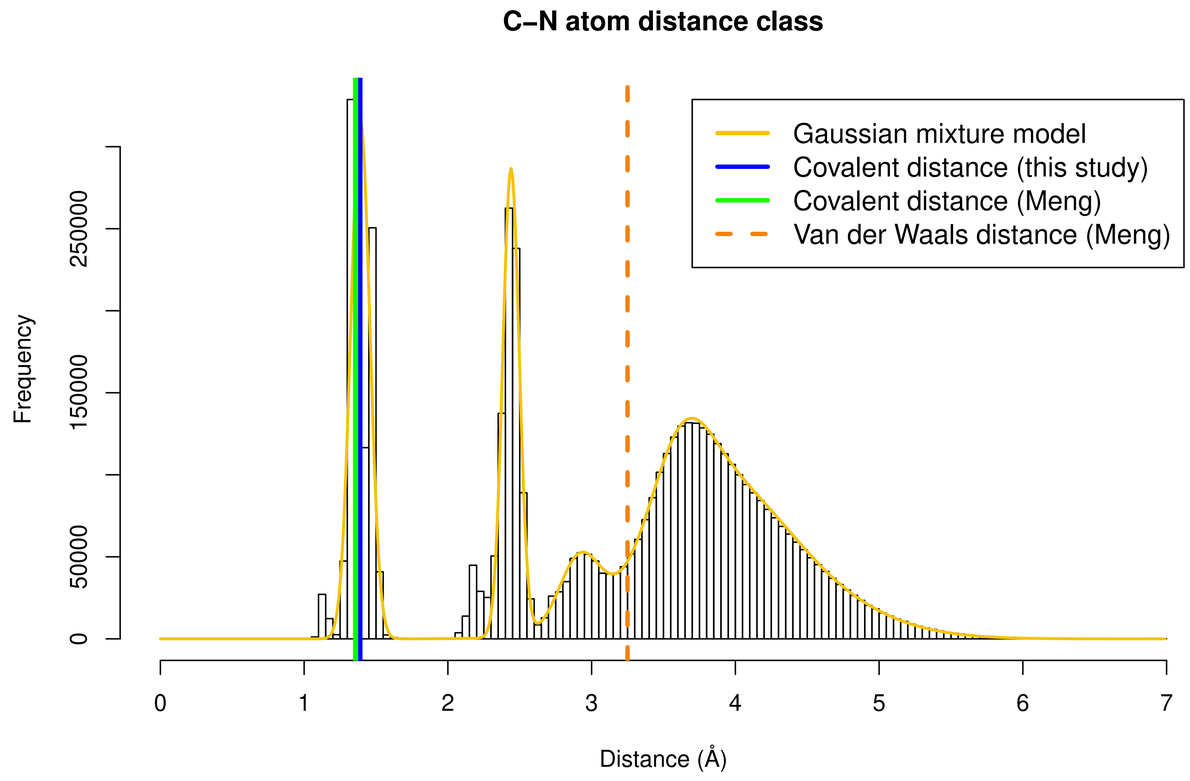
- There are 485 000 entries in COD.
- Atom distances from 361 000 entries were collected.
- Generated 3015 interatomic distance distributions for the research.

Filtering the data

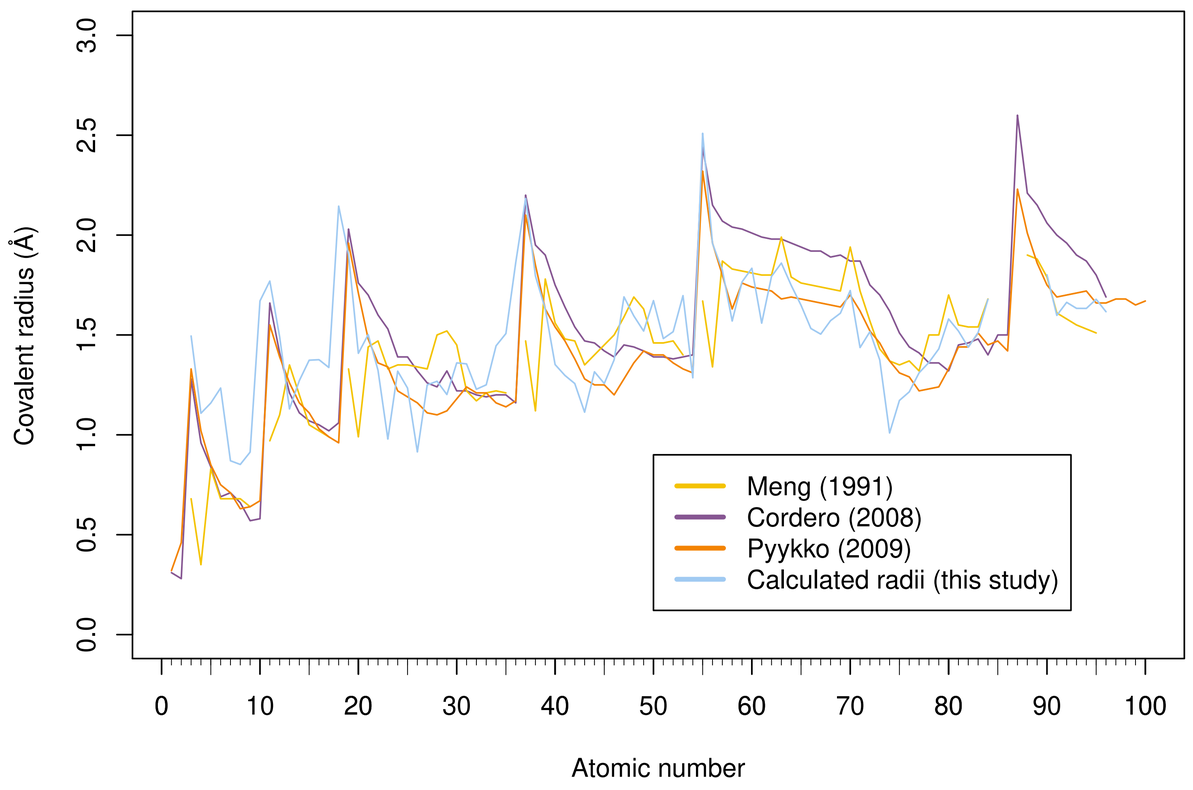
- Entries with superspace groups, unmodelled solvent molecules or calculated non-hydrogen sites were excluded.
- Entries measured under extreme temperature or pressure were excluded.
- Entries with unusually high or unusually low density were excluded.
- Filtering removed 30% of COD entries and 3% of atom distance classes.

Derivation of covalent radii values

- Method uses the best fitting mixture model per Bayesian Information Criterion [5].
- Each pair of components is examined to find the lowest density region in the distribution, which is determined to be the van der Waals gap.
- Mean of the component preceding the van der Waals gap is selected as the covalent bond length.
- Covalent radii table is derived using the least squares method.



Derived covalent radii table



- Root-mean-square deviations from other covalent radii tables:

	Meng 1991	Cordero 2008	Pyykkö 2009
RMSD (Å)	0.296	0.311	0.250

Conclusions

- Methodology for automated covalent radii derivation was developed.
- Derived covalent radii table is promising, however, requires further improvements.
- Research of methods to reduce the impact of random distribution on derived covalent radii is ongoing.

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

[5] 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.

egle.sidlauskaite@mf.stud.vu.lt