Methods for the Determination of Covalent Bonds in Small Molecule Crystals

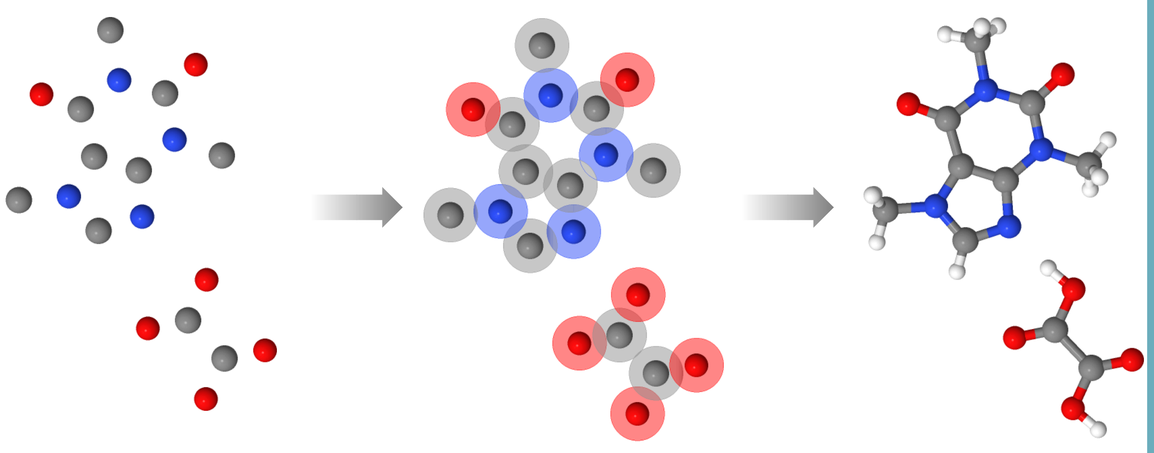
Eglė Šidlauskaitė[a], Andrius Merkys[b]

[a]Department of Mathematical Computer Science, Vilnius University, Vilnius, Lithuania;
[b]Department of Protein-DNA Interactions, Vilnius University Institute of Biotechnology, Vilnius, Lithuania.

OPEN READINGS 2021

Introduction

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

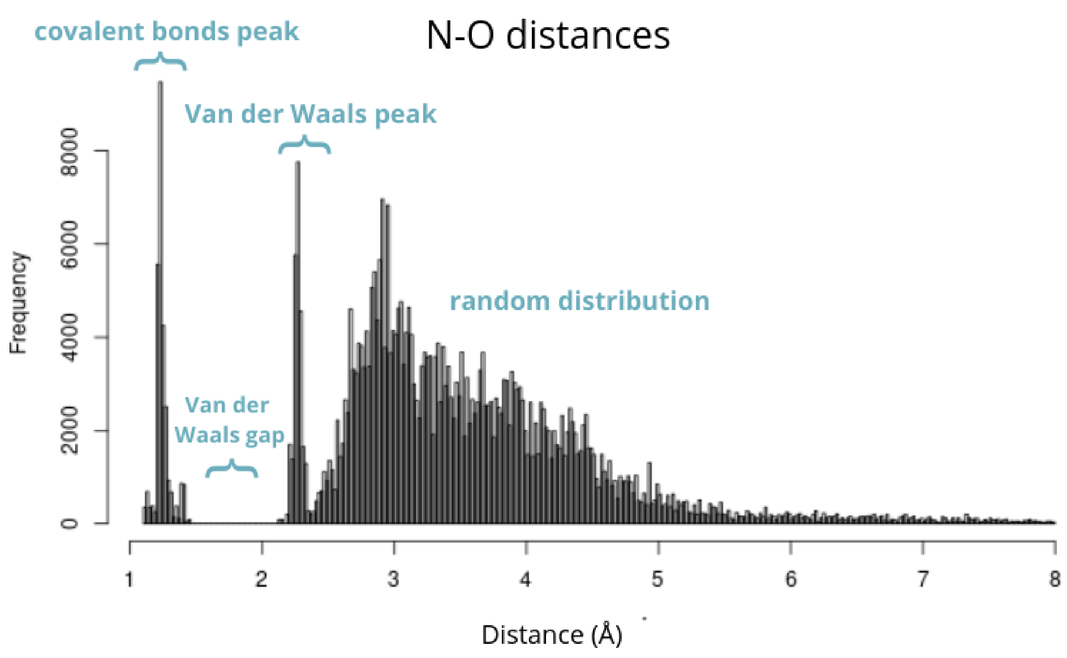


- Widely used covalent radii tables [2] are derived from data from databases that are not open-access, therefore restrict usage and spread of its data and derived results;
- This issue motivates research for a suitable method for determination of covalent bonds in order to recreate covalent radii tables based on data from an open-access database, e.g. the Crystallography Open Database (COD) [3].

Data from the COD

- There are 450 000 entries in COD;
- Atom distances from 10 000 entries were collected;
- Generated 398 atom distance classes for the research.

Van der Waals gap



- Van der Waals gap separates covalent bond lengths from distances between atoms affected by Van der Waals forces;
- Distances between atoms are not observed in this interval [4];
- Distribution of distances between atoms can be approximated by a two component Gaussian mixture model.

Methods

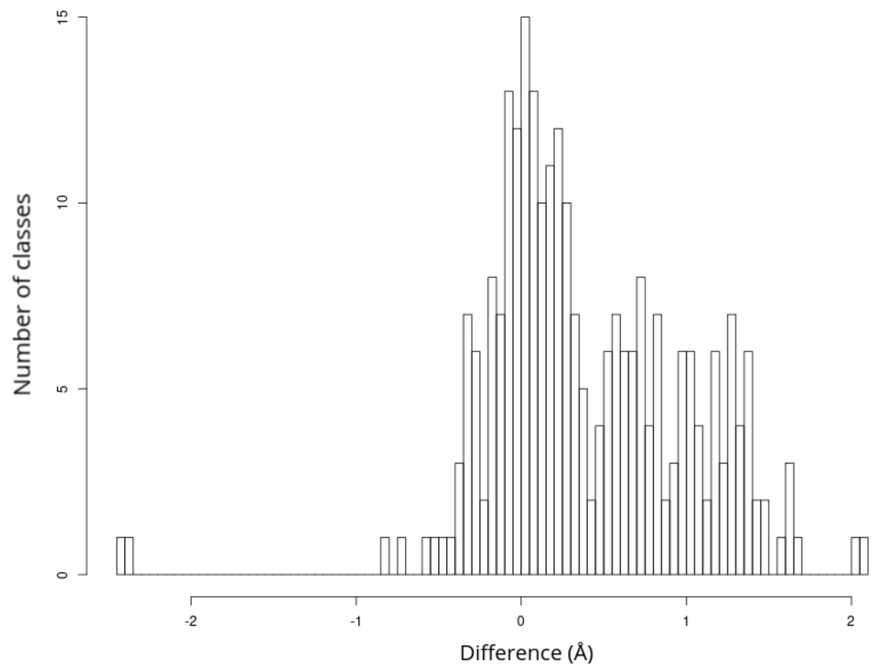
- Generate mixture models with 1 to 10 components;
- Determine the best fitting model per Bayesian Information Criterion [5];
- Determine lowest density region of the model using simplex method.

Two component models

- For 19% of classes, 2 component model was best fitting;
- For 15% of classes, 2 component model was not generated;
- Random distribution has big impact on determination of Van der Waals gap.

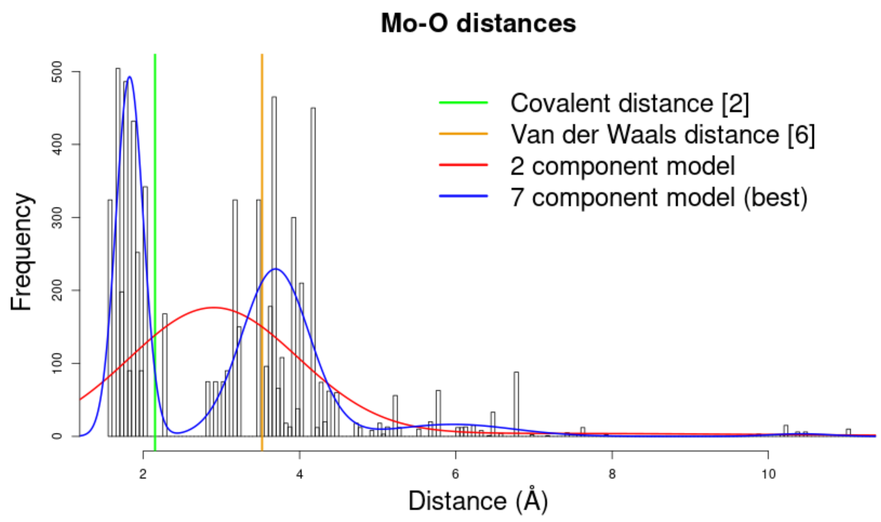
Removal of values exceeding Van der Waals radius

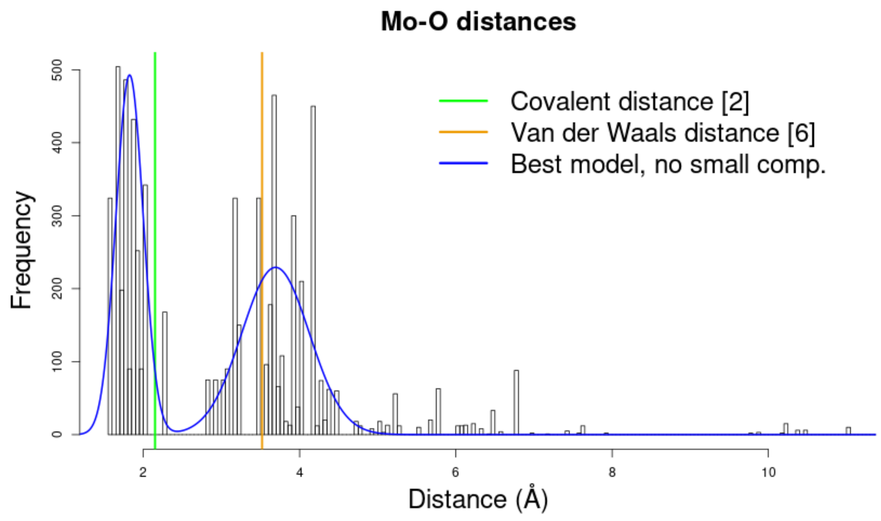
- Covalent radius of an atom is shorter than its Van der Walls radius;
- Effectiveness of the method can be evaluated by removing values that exceed Van der Walls radius [6] to reduce impact of random distribution;
- In over 50% of classes calculated covalent radii differ from values in covalent radii tables [2] in less than 0.5 Å.



Removal of small-proportioned components

- Method uses the mixture model that fits the dataset best;
- Small-proportioned model components are removed;
- Almost 50% of cases after removal only 2 components remain.





Conclusions

- Methodology for automated determination of covalent distance was created;
- The most effective method was determined to be the removal of small-proportioned model components;
- Further analysis on methods to reduce the impact of random distribution on parameters of mixture model components is required.

Bibliography

[1] Gražulis et al. Computing stoichiometric molecular composition from crystal structures.

[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.

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

[6] Bondi. van der Waals volumes and radii. *The Journal of Physical Chemistry*, 68(3):441–451, Mar 1964.

egle.sidlauskaite@mif.stud.vu.lt