VILNIUS UNIVERSITY
FACULTY OF MATHEMATICS AND INFORMATICS
DEPARTMENT OF MATHEMATICAL COMPUTER SCIENCE

Eglė Šidlauskaitė
Bioinformatics study programme

Determination of Covalent Bonds in Small Molecule Crystals

Kovalentinės jungties nustatymas mažų molekulių kristaluose

Bachelor's thesis

Supervisor **Andrius Merkys**

Vilnius, 2021

Summary Covalent radii are widely used to determine covalent bonds in small molecule crystals. However, there is no univocal published covalent radii table and most of these tables are derived based on data from databases that restrict usage of their data or derived results. Therefore, during this research five different methods of covalent radii derivation from open-access data were analysed.

The methods are based on determination of the covalent bond length by locating the Van der Waals gap in distributions of distances between atoms. This is achieved by fitting a Gaussian mixture model to distribution of distances between atoms and using simplex method to determine the lowest density region between chosen components. As covalent bond length is equal to the sum of covalent radii of atoms that create the bond, covalent radii are derived by creating an overdetermined system of equations and solving it using the least squares method.

During the research, automated tool for covalent radii derivation was created and covalent radii were derived using data from Crystallography Open Database, which is an open-source database. Calculated covalent radii were evaluated by comparing the results with published covalent radii tables.

Keywords Covalent radii derivation, Van der Waals gap, Gaussian mixture model, Simplex method, Crystallography Open Database

Santrauka Kovalentiniai spinduliai plačiai naudojami kovalentinių jungčių nustatymui mažų molekulių kristaluose. Deja, nėra vienos vienareikšmės publikuotos kovalentinių spindulių lentelės. Taip pat, dauguma šių lentelių yra išvestos naudojant duomenis iš duomenų bazių, kurios nėra atviros prieigos. Taigi, šiame tyrime ištirti penki metodai skirti kovalentinių spindulių nustatymui.

Analizuoti metodai yra paremti kovalentinės jungties ilgio apskaičiavimu remiantis Van der Waals plyšio nustatymu atomų atstumų klasėse. Van der Waals plyšys nustatomas atomų atstumų klasėms pritaikant normaliųjų skirstinių mišinio modelį ir panaudojant simplekso metodą mažiausio tankio taško nustatymui tarp pasirinktų mišinio komponentų. Kadangi kovalentinės jungties ilgis yra lygus jungtį sudarančių atomų kovalentinių spindulių sumai, kovalentiniai spinduliai išvedami sudarius persąlygotą (angl. overdetermined) lygčių sistemą ir išsprendžiant ją naudojant mažiausių kvadratų metodą.

Tyrimo metu buvo sukurtas automatizuotas įrankis kovalentinių spindulių nustatymui. Kovalentiniai spinduliai buvo apskaičiuoti naudojant duomenis iš atviros prieigos duomenų bazės Crystallography Open Database. Apskaičiuoti kovalentiniai spinduliai buvo įvertinti lyginant juos su kovalentiniais spinduliais pateikiamais publikuotose kovalentinių spindulių lentelėse.

Raktiniai žodžiai Kovalentinių spindulių nustatymas, Van der Waals plyšys, Normaliųjų skirstinių mišinio modelis, Simplekso metodas, Crystallography Open Database

Contents

1 Introduction

X-ray crystallography is a widely used method of crystal structure determination. It is used to determine the exact 3D positions of the atoms that constitute the crystal structure. However, this method cannot capture chemical bonding [GMVOK15].

Covalent radius is an usual radius of an atom in a covalent bond. In practice, 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 element-specific covalent radii. This distance is called covalent bond length.

Covalent radii tables are created by statistically analysing lengths of observed chemical bonds and are often used to determine chemical bonds between atoms. There is no univocal method for covalent radii derivation, therefore there are multiple covalent radii tables available and there is a relatively big difference between results provided in them (as shown in Figure 1).

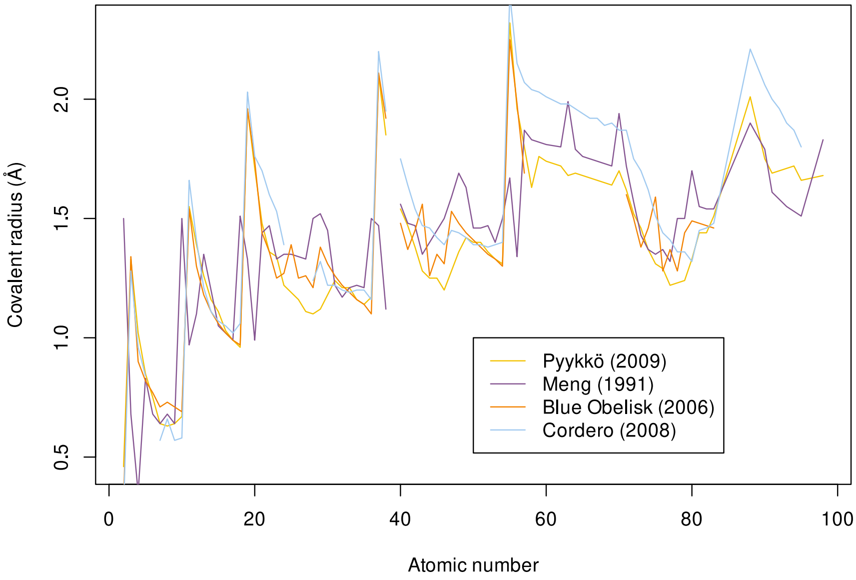

Figure 1: Comparison of covalent radii in selected covalent radii tables.

Additionally, covalent radii tables are often derived from data from databases that are not open-access (for example Cambridge Structural Database – CSD [GBLW16]), hence restricting usage of their data and dissemination of derived results. This issue motivates research for a method for determination of covalent radii in order to recreate covalent radii tables based on data from an open-access database (for example Crystallography Open Database – COD [GDM$^+$12]).

1.1 Objectives

- Create methodology for covalent radii derivation and analyse different approaches that can be used to reduce the impact of the random distribution in the data to the covalent radii.

- Develop an automated tool which would allow to derive covalent radii by statistically analysing data. Use the tool to derive covalent radii from data from an open-access database – COD.

- Evaluate calculated covalent radii based on comparison with published covalent radii tables. Inspect the discrepancies by analysing the structure of crystals in which the grossest outliers in terms of atom distance were observed.

2 Literature overview

There are a few selected covalent radii tables that will be used for comparison in this research. The oldest one of them is published by Meng and Lewis in 1991 [ML91]. The article reports covalent radii from CSD [ABC+79], which were generated from covalent bond lengths retrieved using X-ray crystallography and neutron diffraction (for bonds involving hydrogen atoms) [AKW87]. In Meng and Lewis (1991) it is also stated that tolerance value of 0.4 Å is used to help determine connectivity in low-resolution structures. This shows that a relatively large tolerance is applicable in terms of covalent radii.

However, the main criticism for CSD covalent radii is very well described in Cordero et al. (2008) article [CGPP+08] – over the years these covalent radii values were manually amended to conform to connectivity of specific structures. Additionally, elements for which radii could not be derived due to lack of data were assigned a default radius of 1.50 Å (hence the large tolerance value). The method used to generate Cordero et al. (2008) covalent radii tables starts by defining covalent radii for organic elements – nitrogen, carbon and oxygen – from experimental bond distances. By applying certain limitations, N–N, C–C, C–N and C–O atom distance classes were considered and covalent radii were derived from average bond length in each of them. Covalent radii of remaining elements were derived with respect to nitrogen, carbon and oxygen by analysing average bond length in atom distance classes where one of these elements constitutes the bond.

A different approach is taken in research for single-bond covalent radii published by Pyykkö and Atsumi in 2008 [PA08]. In this case, all elements are treated as equal and covalent radius of each element is derived independently using least-squares fit (method applied in this research as well – see Section 3.6).

The fourth covalent radii table that will be used in this research is published by Guha and coauthors in 2006 and is available as part of the Blue Obelisk Data Repository [GHH+06]. It is known that this table is heavily based on Pyykkö and Atsumi (2008) publication, however it is not clear what methods were applied to make alterations to the data.

3 Methods

3.1 Approach of this research

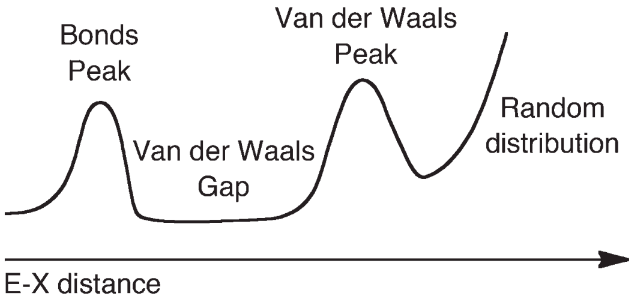

Figure 2: A qualitative illustration of a typical atom distance distribution from article [Alv13]. E and X – elements constituting the atom distance class, bonds peak represents covalent bond distances, Van der Waals gap – interval with no atom distances, Van der Waals peak represents distances between atoms affected by Van der Waals forces, random distribution – exponential component which includes other atom distances and noise observed in the data.

Four main sections can be identified in the distribution of atom distances of most atom distance classes (based on Alvarez (2013) [Alv13] and shown in Figure 2). The exponential component, also called random distribution, includes distances between atoms that are induced by other forces or noise in the data. This research postulates that exponential component can be reduced using Voronoi cells. During determination of distances between atoms, crystals are sectioned into Voronoi cells with atoms in the cell centre [OV14] and only distances between atoms in neighbouring cells are included in final atom distance classes. This way only the nearest neighbours of each atom are considered.

After reducing the impact of the exponential component, the distribution of distances between atoms can be approximated by a Gaussian mixture model. In such case the first component would represent covalent bonds, the second component would represent atoms that are affected by the influence of the Van der Waals forces and remaining components would represent other types of interactions as well as noise (as shown in Figure 3). Covalent bonds may as well form more than one component, but this issue will be addressed further on.

As in Van der Waals gap distances between atoms are not observed [Alv13], this interval corresponds to the lowest density region between components. Once

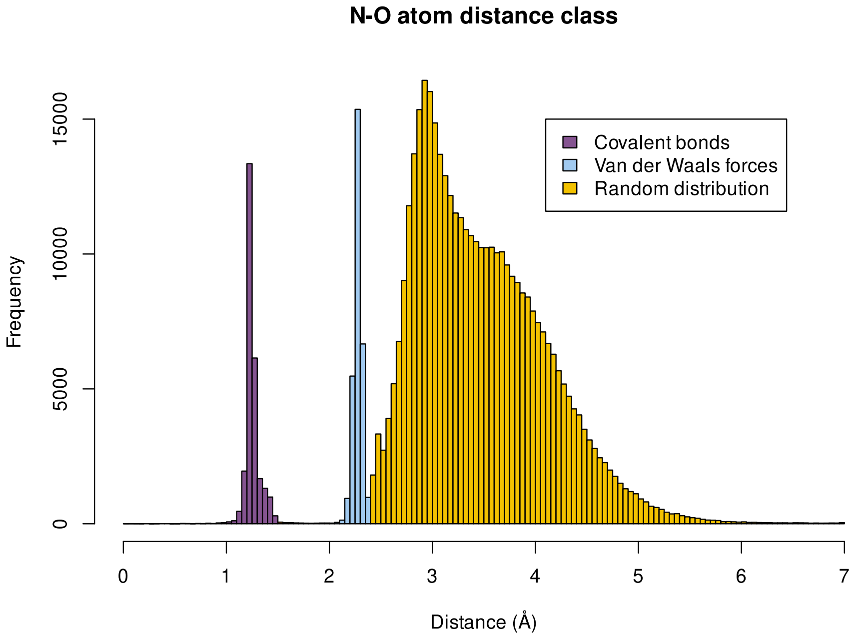

Figure 3: Distribution of distances between atoms in N–O atom distance class. Van der Waals gap is clearly visible in the histogram, in the interval from 1.5 to 2 Å.

Gaussian mixture model is applied to the distribution of distances between atoms, midpoint of the lowest density region between correctly selected components would represent the length of the covalent bond plus roughly a half of the Van der Waals gap width. As previously mentioned in Section 2, a tolerance value is often used by other authors in published covalent radii tables, therefore the addition of half of the Van der Waals gap width does not have a significant effect on determination of a covalent bond. Thus further in this research, covalent radii with the addition of a part of the Van der Waals gap will be referred to as calculated covalent radii.

The challenge arises while analysing distance distributions of elements which do not have a clear Van der Waals gap, for example alkaline elements, copper, silver, etc. Additionally, this approach is complicated by random distribution that remains in the data even after Voronoi cells are used. Noise can appear in the dataset due to computational mistakes. For example, in Co–Co atom distance class there are instances where observed bond length is less than 0.1 Å which are caused by incorrectly merging symmetrical copies of an atom which is located in the special crystallographic position. Another reason for noise observed in the data is atoms being affected by forces other than covalent or Van der Waals. In order to deal with aforementioned biases to covalent radii calculations five different methods for determination of suitable mixture model were analysed.

3.2 Data

Data for the research was generated by analysing 129320 crystal structures form COD entries in range 4000000 – 4999999 (revision 199925). Tool `cif_fillcell` from *cod-tools* [1] (revision 8653) [MVB+16] was used to restore crystal structures from CIF data files to 3 × 3 × 3 supercells. Tool `cif_contacts` from *crystal-contacts*[2] package (revision 42) was used to determine distances between atoms in crystals by separating crystals into Voronoi cells with atoms in the cell centre [OV14] and calculating distances between atoms in neighbouring cells. All data is stored in *molecules-in-COD* repository[3] which will be opened for public access later.

Data resulting from atom distance measurements is separated into classes, each of which represents distances observed between a specific pair of chemical elements. Each entry in the atom distance class consists of length of the distance between the atoms, atom indicators and identifier of the COD entry in which the distance was observed. Atom indicators are constructed by specifying name of the atom in the original structure, symmetry operator which was used to restore the atom and translation in the crystal cell, as used in [GMVOK15].

Published tables of Van der Waals radii are used in the research to evaluate the methods and create more informative histograms. Van der Waals distance was calculated by finding the sum of Van der Waals radii of each atom constituting the atom distance class. The values of Van der Waals radii were taken from `COD::AtomProperties` module which describes chemical properties of the atoms and is a part of *cod-tools* (revision 8653) package. The module uses Van der Waals radii published by Alvarez in 2013 [Alv13].

Calculated covalent radii need to be compared to covalent radii from published covalent radii tables in order to determine the accuracy of used methods. For this purpose, data from four published covalent radii tables was used, published by Meng

---

[1] Located at `svn://www.crystallography.net/cod-tools/trunk`, revision 8653

[2] Located at `svn://saulius-grazulis.lt/crystal-contacts/trunk`, revision 42

[3] Located at `svn://www.crystallography.net/molecules-in-COD/branches/contacts/classes`, revision 2724

and Lewis (1991) [ML91], Guha et al. (2006) [GHH+06] (later referred to as the Blue Obelisk), Cordero et al. (2008) [CGPP+08] and Pyykkö and Atsumi (2009) [PA09].

3.3 Fitting of the mixture model

During the research it was decided to use Gaussian mixture models [Alv13]. Program `fit-model` [Mer18] was used in order to fit a suitable mixture model to the data. The program uses functions from the R package *MixtureFitting v0.3.4*[4] to determine parameters of selected type of mixture model by utilizing expectation-maximisation algorithm [Mer18]. This tool tries to apply 10 mixture models to the provided data. Each model differs in number of components in mixture (from 1 to 10). Program also determines which of the models fits data the best based on Bayesian Information Criterion (BIC) [Sch78]. Models are calculated for each atom distance class. Output of the generated model contains tab-separated fields with the following data (all distances in Å):

1. Number of instances in dataset;
2. Type of applied mixture model ("gmm" – Gaussian mixture model);
3. Number of components in the model;
4. Logarithm of likelihood of the model given data;
5. BIC;
6. Evaluation of the model ("best" indicates the best fitting model, "NONE" otherwise);
7. Proportions of components;
8. Centres of components;
9. Standard deviations of components.

---

[4] Located at `https://github.com/merkys/MixtureFitting`

3.4 Selecting suitable mixture model

To correctly determine the location of Van der Waals gap and to improve the precision of the calculated covalent radii, a suitable mixture model needs to be selected. The main objective in this case is to select the mixture model which would correctly apply components on intervals of data that portray covalent bonds and Van der Waals interactions in order to find the lowest density region between them. To determine the best approach, five methods have been used for further calculations:

Method 1 Mixture model which has only two components is selected even if it is not the best model by BIC. Lowest density region between the two components is identified (see Section 3.5). Atom distance classes for which two component mixture model is not generated are omitted from further calculations.

Method 2 Mixture model which is evaluated as the best fitting based on BIC is selected and smallest by proportion components are removed until only two components remain in the model. Lowest density region has to be identified (see Section 3.5) between the two remaining components. Atom distance classes for which one component Gaussian mixture model is selected as best are omitted from further calculations.

Method 3 Mixture model which is evaluated as the best fitting based on BIC is selected, and components with proportions less than 0.04 are removed. Lowest density region between each two neighbouring components is calculated (see Section 3.5) and the point with the lowest density is selected. Atom distance classes for which one component Gaussian mixture model is selected as best are omitted from further calculations.

Method 4 Mixture model which is evaluated as the best fitting based on BIC is selected. Lowest density region is identified (see Section 3.5) between the first two components (i.e. two components with centre points closest to zero). Atom distance classes for which one component Gaussian mixture model is selected as best are omitted from further calculations.

Method 5 Mixture model which is evaluated as the best fitting based on BIC is selected, and components with proportions less than 0.04 are removed. Lowest density region is identified (see Section 3.5) between the first two components (i.e. two components with centre points closest to zero). Atom distance classes that after small component removal have less than two components remaining are omitted from further calculations.

To evaluate methods used in the research, all calculations were repeated after clearing the data of distances between atoms which are less than 0.1 Å or exceed the Van der Waals distance of that class. This action removes a big part of the random distribution and allows Van der Waals gap to be identified easier. Therefore, comparing results from the original data set and cleared data set allows to evaluate how much impact does the noise have on the effectiveness of the method in question. Additionally, if method is unable to deliver suitable results with a cleared dataset, it can be determined to be unusable.

Additionally, all methods described were repeated after removing atom distance classes that have less than 30 instances in the dataset. This removes approximately 20% of the atom distance classes. These calculations were done to evaluate the impact of possible mistakes in scarcely populated classes. However, removal of these classes did not have a significant impact, therefore all results in this research are derived from the full data set.

3.5 Calculating covalent bond length

To determine the exact location of lowest density region simplex method is used. The simplex method is an iterative process that finds the most favourable of all feasible solutions. It consecutively checks if the marginal points of the set are optimal and if they are not – moves to neighbouring marginal points which have a lower value. Simplex function is part of R package *MixtureFitting v0.3.4*.

Before simplex method is invoked, the function of mixture model density is altered – interval from 0 to the centre of the first component is assigned the maximum value of the first component and interval from the centre of the second component to infinity is assigned the maximum value of the second component. This alteration provides certainty that the lowest density point will be determined between selected components.

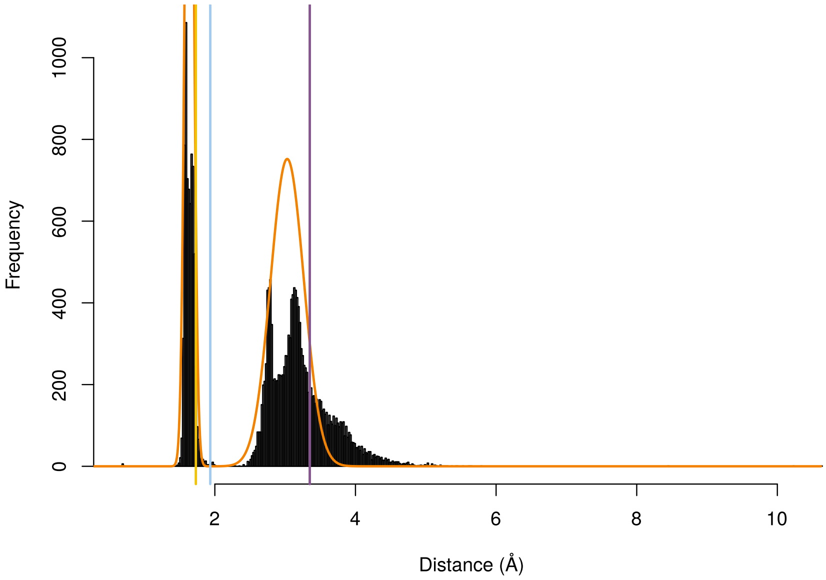

Figure 4: Distribution of distances between atoms in N–P atom distance class. Orange – Gaussian mixture model, yellow – published covalent bond length [ML91], purple – Van der Waals interaction distance [Alv13], blue – calculated covalent bond length.

Results of the calculations from all classes are collected in a result file. Additionally, a histogram for each atom distance class is plotted (as seen in Figure 4) which includes generated mixture model, covalent bond distance (calculated from a published covalent radii table [ML91]), Van der Waals distance (calculated from a published Van der Waals radii table [Alv13]) and calculated covalent bond length.

3.6 Calculating covalent radii

As described previously, covalent bond length is equal to the sum of covalent radii of bonded atoms. An overdetermined system of equations from all atom distance classes is then created which, when solved, yields values of individual radii. To derive covalent radii from calculated covalent bond lengths R package *Rlinsolve v0.3.1* is used. This package provides basic stationary iterative solvers for systems of linear equations.

The package provides multiple solvers, hence a number of them were tested. Results were compared to two of the existing covalent radii tables Meng and Lewis (1991) [ML91] and Blue Obelisk (2006) [GHH$^+$06] by finding the difference between covalent radii of each atom and calculating average difference for each method. Sum of average difference from both tables was used as a measure for comparison. After comparison it was determined that Symmetric Successive Over-Relaxation (SSOR) method delivered the most accurate results (as shown in Table 1), hence it was used in further calculations.

Table 1: Average difference (Å) between existing covalent radii tables and radii calculated using different solvers from *Rlinsolve* package.

Method	**ssor**	**gs**	**bicg**	**bicgstab**	**cgs**	**cheby**	**gmres**	**qmr**
Method 1	0.6922	0.6922	1.4971	0.7278	0.7634	0.7635	0.7636	1.7091
Method 2	0.6522	0.6522	1.4259	0.6346	0.6169	0.6169	0.6168	1.5155
Method 3	0.6984	0.6984	1.5189	0.7060	0.7136	0.7136	0.7135	1.6337
Method 4	0.4712	0.4721	1.3746	0.5100	0.5480	0.5480	0.5479	1.5194
Method 5	0.5125	0.5125	1.5675	0.5452	0.5779	0.5779	0.5778	1.6503

Calculated covalent radii are stored in YAML format tables. Additionally, each set of calculated covalent radii is visualised in a graph which includes a comparison with existing covalent radii tables.

3.7 Connectivity testing

The validity of calculated radii can be tested by comparing connectivity between atoms in molecules based on published versus calculated radii. SMILES (simplified molecular-input line-entry system) notation is a suitable format for this comparison. For chemical elements, for which covalent radius could not be determined during this research (for example, hydrogen), values from Meng (1991) table are used.

To generate SMILES strings, personal branch[5] of *cod-tools* is used where SMILES writer was included in program `cif_molecule` during research for term paper [Š20]. In the term paper research it was concluded that there were 9 molecules for which generated SMILES string matched the manually-curated SMILES string identically. Therefore, the same 9 molecules will be used for radii comparison. Detailed explanation of process of generating SMILES strings is provided in the term paper.

Program `cif_molecule` adds a tolerance value of 0.35 Å to Meng (1991) covalent radii. When performing calculations with derived covalent radii, additional tolerance value is not added as calculated covalent radii already include a part of Van der Waals

---

[5] Located at `svn://www.crystallography.net/cod-tools/branches/feature/egle-SMILES`, revision 8680

gap as a tolerance value. After generating the SMILES strings they are converted to canonical SMILES using *Open Babel* software package (version 2.3.2.) [OBJ+11] in order to simplify the comparison process.

4 Results

As previously mentioned in Section 3.4, to evaluate impact of random distribution and errors in data, two types of data sets are used in calculations:

Dataset 1 full data set.

Dataset 2 data set with instances between 0.1 Å and Van der Waals distance.

4.1 Two component model (Method 1)

Results received by using Dataset 1 differ from published covalent radii tables quite significantly (as shown in Figure 5). The most alerting discrepancies are related to organic atoms (for example carbon, oxygen, nitrogen) that differ from published covalent radii tables in more than 1 Å.

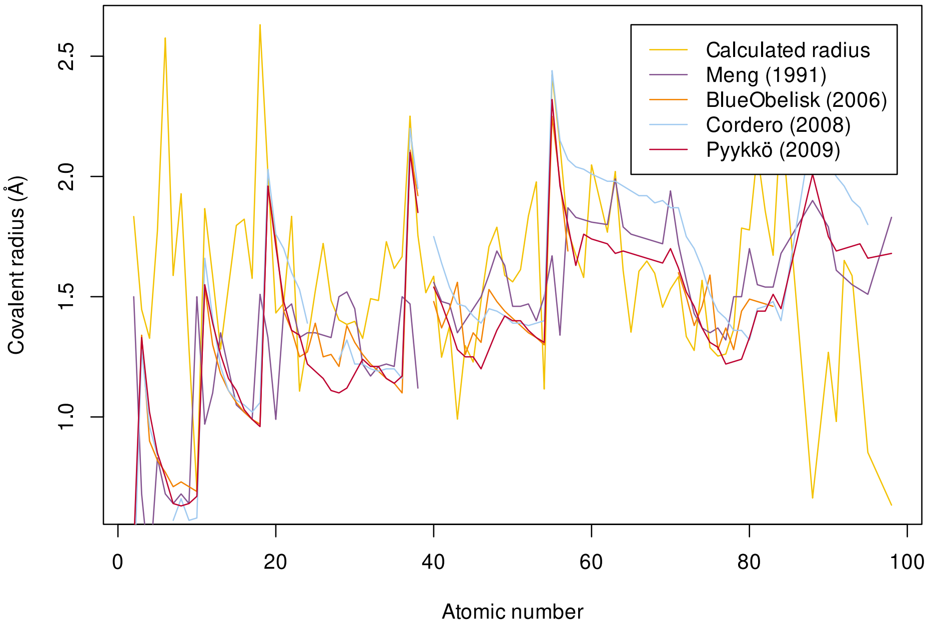

Figure 5: Calculated covalent radii (Method 1) comparison.

Significant decrease in number of covalent radii that differ from covalent radii in published tables in more than 0.5 Å is observed once manually cleared Dataset 2 is used (as shown in Table 2). It shows that this method is highly impacted by random distribution. Full list of atoms that differ from published covalent radii in more than 0.5 Å using Dataset 1 can be found in Appendix 1.

Table 2: Number of atoms for which calculated radii differ from published covalent radii tables in more than 0.5 Å.

Table	**Dataset 1**	**Dataset 2**
Meng (1991)	27	9
Blue Obelisk (2006)	15	3
Cordero (2008)	20	5
Pyykkö (2009)	20	4

These issues occur as two component model in most cases is not the best fitting – only in 586 out of 1608 classes 2 component model was determined to be best fitting as per BIC. Therefore, while fitting the model, some components are merged into the first component and random distribution is treated as the second component.

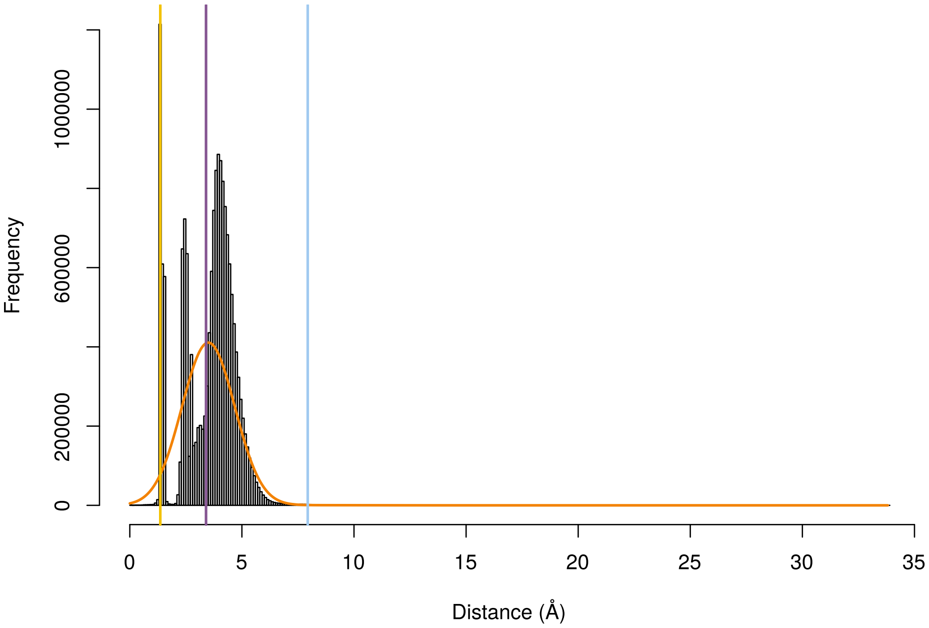

Figure 6: C–C atom distance class. Orange – Gaussian mixture model, yellow – published covalent bond length [ML91], purple – Van der Waals interaction distance [Alv13], blue – calculated covalent bond length.

Example of large discrepancy is carbon atom. Calculated covalent radius of carbon is 2.58 Å. One of the classes that include carbon and has extremely large calculated bond length is C–C atom distance class. For C–C atom distance class two component model was not selected as best. Calculated covalent bond length of this atom distance class is 7.93 Å. During the review of the class histogram (as shown in Figure 6) it was observed that only one component is clearly visible and calculated bond length is longer than its centre. After reviewing properties of the applied mixture model it was determined that a small peak of random distribution (proportion of the component is 0.002, location of the component centre is at 7.936 Å) has been located and the second component, therefore covalent radius was calculated incorrectly.

Similar situation is observed in all cases where calculated covalent radii significantly exceed covalent radii in published tables. It can be concluded that this method is not able to reduce the impact random distribution has on covalent radii determination.

4.2 Largest components in best fitting model (Method 2)

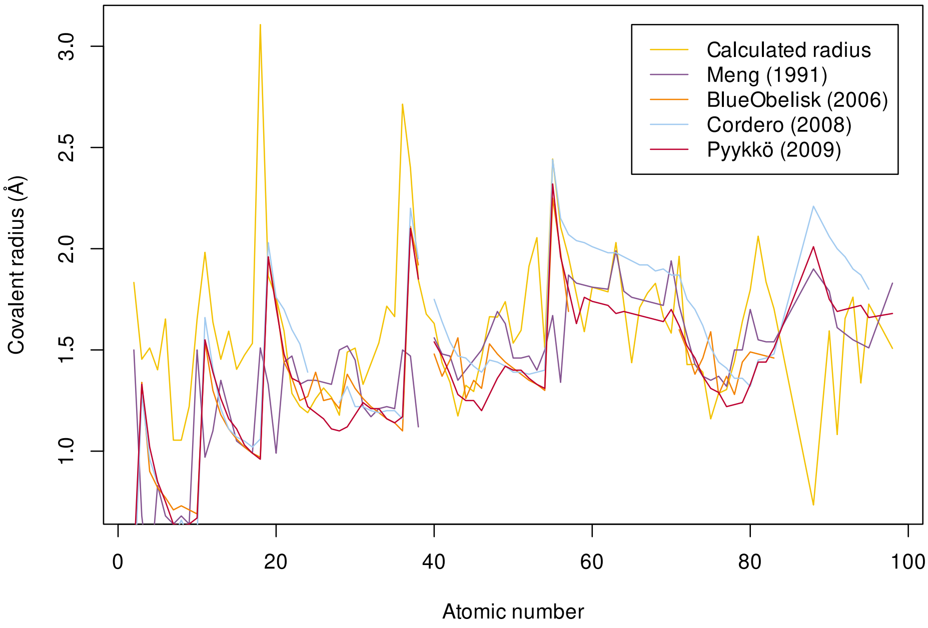

Figure 7: Calculated covalent radii (Method 2) comparison.

From the results of Method 1 it is clear that generated two component mixture models in often cases are unable to fit data precisely. Therefore, in further methods mixture models that were determined to fit data best as per BIC are used. However, in case best fitting mixture model has more than two components, correct components that represent covalent bonds and Van der Waals interactions need to be selected to be able to correctly identify the Van der Waals gap. As described in Section 3.2, Voronoi cells were used when retrieving the data for this research.

This approach should reduce the amount of random distribution observed in the data sets. Hence, in this method only the two biggest components were kept in the mixture model.

Results (as shown in Figure 7) show slight improvement from Method 1, however big difference in radii can still be observed. More detailed comparison can be seen in Table 3 – number of atom classes that differ from published covalent radii tables was reduced by approximately 20% for all authors, except for Blue Obelisk. However, number of mistakes is still inadequate. Additionally, comparing the results to the results of Dataset 2 shows that impact of random distribution is still significant. Full list of atoms that differ from published covalent radii in more than 0.5 Å using Dataset 1 can be found in Appendix 2.

Table 3: Number of atoms for which calculated radii differ from published covalent radii tables in more than 0.5 Å.

Table	**Dataset 1**	**Dataset 2**
Meng (1991)	20	8
Blue Obelisk (2006)	14	4
Cordero (2008)	16	6
Pyykkö (2009)	15	4

Using this method calculated covalent radius of carbon is equal to 1.69 Å which is smaller than using Method 1, however still significantly exceeds radii from published tables. In this case, atom distance class with largest covalent bond distance is C-Se where covalent bond distance equals to 4.92 Å. In the C-Se atom distance class histogram (as shown in Figure 8) it is clearly visible that both largest components are part of the random distribution, therefore covalent bond length is identified incorrectly. It can be concluded that even after Voronoi cells are used to clean up the data, peaks of random distribution can still be significantly large, therefore this method is unable to reduce the impact of random distribution.

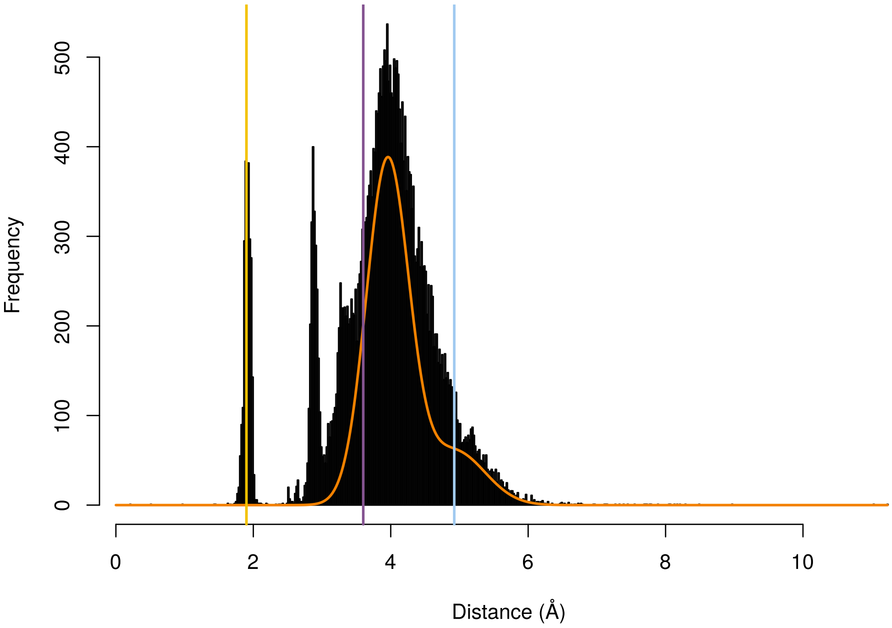

Figure 8: C–Se atom distance class. Orange – Gaussian mixture model, yellow – published covalent bond length [ML91], purple – Van der Waals interaction distance [Alv13], blue – calculated covalent bond length.

4.3 Lowest density region in best fitting model (Method 3)

As discussed in Section 3.1, in theory Van der Waals gap is an interval of distances between atoms in which no instances are observed. This supports the idea that if lowest density point was calculated between each two neighbouring components, the area where Van der Waals gap is located should include lowest result. However, after the investigation of atom distance classes it was determined that in most classes there are instances of atom distances that are very close to zero (<0.2 Å) or are very large (>10 Å). These instances are part of random distribution which is caused by computational errors. Furthermore, after best fitting model is applied to the dataset, these instances can be identified as extremely small proportioned (<0.04) components, located far from actual atom distances. Hence, as shown in Table 4 lowest density region is identified incorrectly.

Table 4: Number of atoms for which calculated radii differ from published covalent radii tables in more than 0.5 Å before removal of small components.

Table	Dataset 1	Dataset 2
Meng (1991)	41	7
Blue Obelisk (2006)	21	4
Cordero (2008)	35	8
Pyykkö (2009)	32	6

Therefore, an attempt was made to improve this method by removing components from the best fitting model that have proportions less than 0.04 before calculations. Removal of small proportioned components has definitely improved covalent radii calculation results, although in often cases they are still not satisfactory compared to published covalent radii tables (as shown in Figure 9).

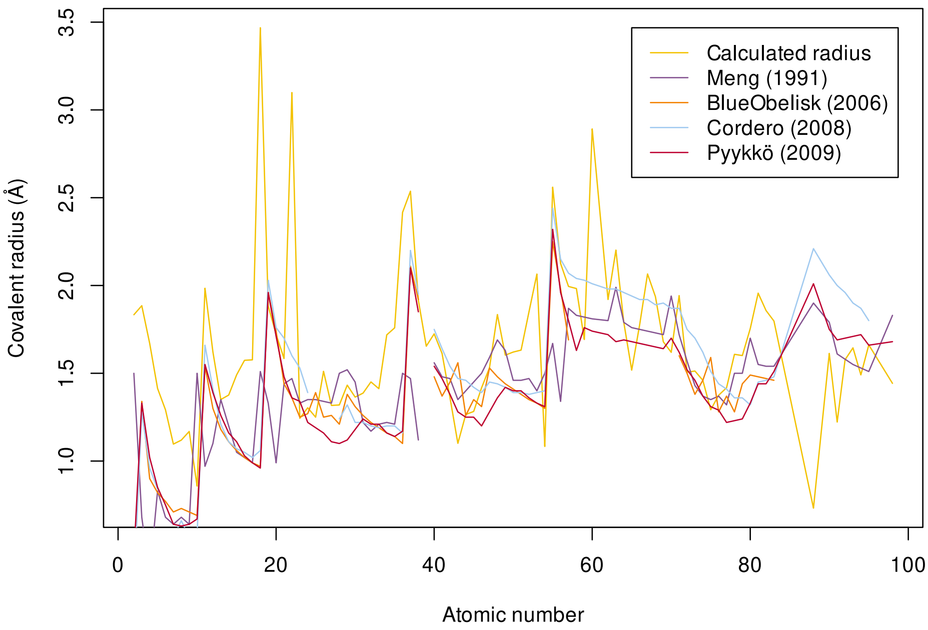

Figure 9: Calculated covalent radii (Method 3) comparison.

After removal of small components, number of atoms for which calculated covalent radii exceed published covalent radii was reduced almost by half (as per Table 5), however usage of Dataset 2 still delivers significantly better results. Therefore, this method is highly impacted by random distribution observed in the data. This is partially caused by removal of small components not being the ideal method for outlier removal. For example, there are two atom distance classes (Nd–Ti and Li–Ti) for which calculated bond length exceeds 26 Å. In both of these classes the component which causes huge deviation has an approximate proportion of 0.05. However, increasing the limit of proportion for removal of components has a negative effect on overall results. The value of calculated covalent radius for carbon using this method is 1.29 Å which is still more than 0.5 Å higher than in Meng and Lewis (1991), Blue Obelisk (2006) and Pyykkö and Atsumi (2009) covalent radii tables. Full list of atoms that differ from published covalent radii in more than 0.5 Å using Dataset 1 can be found in Appendix 3.

Table 5: Number of atoms for which calculated radii differ from published covalent radii tables in more than 0.5 Å after removal of small components.

Table	**Dataset 1**	**Dataset 2**
Meng (1991)	23	9
Blue Obelisk (2006)	14	3
Cordero (2008)	17	9
Pyykkö (2009)	18	5

4.4 First two components in best fitting model (Method 4)

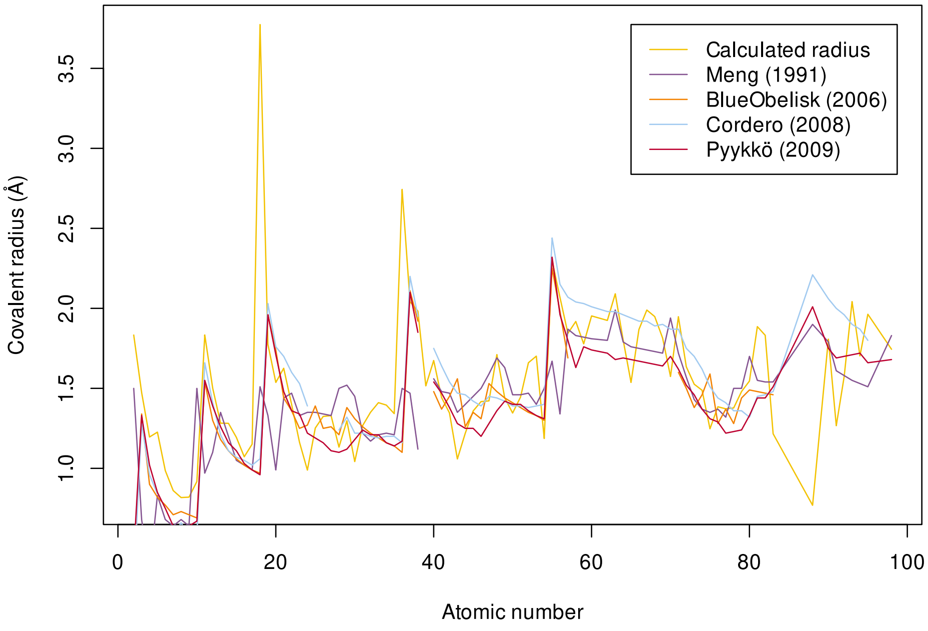

Figure 10: Calculated covalent radii (Method 4) comparison.

Approach used in Method 4 has delivered much better results. As shown in Figure 10 in most cases derived covalent radii results are pretty similar to covalent radii in published tables. In Table 6 it is evident that number of elements for which calculated covalent radius strongly differs from published tables is quite low. In almost all tables largest differences occur with helium, argon, krypton (noble gasses) and radium (alkaline earth metal, very rare in COD). Full list of atoms that differ from published covalent radii in more than 0.5 Å using Dataset 1 can be found in Appendix 4.

Table 6: Number of atoms for which calculated radii differ from published covalent radii tables in more than 0.5 Å.

Table	**Dataset 1**	**Dataset 2**
Meng (1991)	12	9
Blue Obelisk (2006)	4	3
Cordero (2008)	5	8
Pyykkö (2009)	4	6

However, from Table 6 it can be noticed that in some cases removal of random distribution (using Dataset 2) increases the amount of errors. It hints that in this method random distribution still has an impact on the derived covalent radii, however these mistakes compensate one another in the process and final results appear correct. For example there are six atom distance classes in which calculated atom bond length is smaller than 0.1 Å (as shown in Figure 7), which clearly signifies errors.

Table 7: Atom distance classes with extremely small calculated atom covalent distance. C1 stands for first component, C2 stands for second component in the model.

Class	**Bond length**	**C1 location**	**C1 proportion**	**C2 location**	**C2 proportion**
Bi–Ca	0.00077 Å	0.00013 Å	0.02469	3.31948 Å	0.75740
Br–I	0.00082 Å	0.00018 Å	0.01506	1.04689 Å	0.00803
Bi–Zn	0.00104 Å	0.00033 Å	0.05172	0.39622 Å	0.62345
Se–Te	0.00144 Å	0.00033 Å	0.02928	1.22505 Å	0.02650
Mn–Zn	0.08471 Å	0.00941 Å	0.09091	1.18050 Å	0.59091

4.5 Small component removal (Method 5)

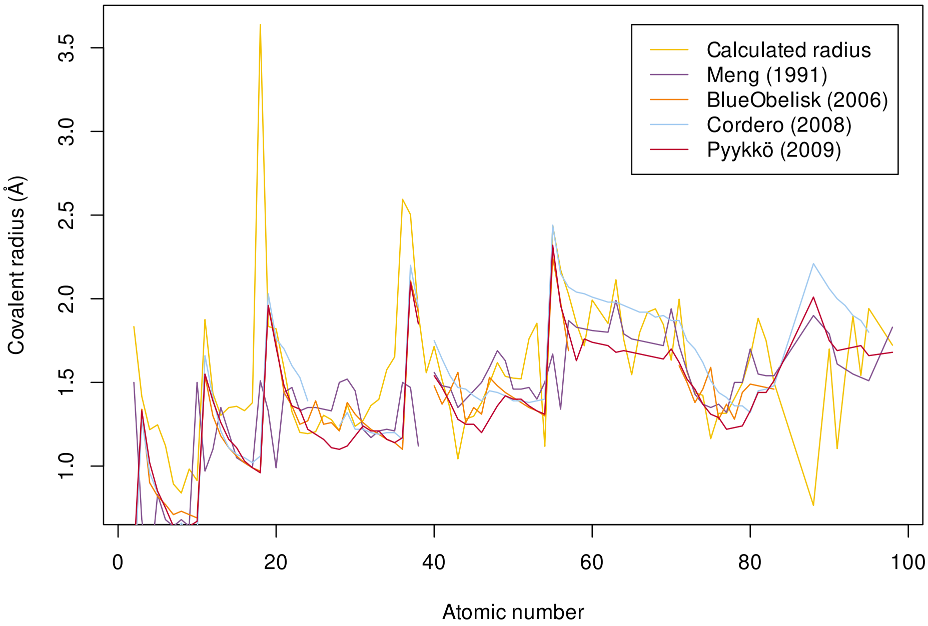

Figure 11: Calculated covalent radii (Method 5) comparison.

As previously shown in Table 7, proportions of components that cause very short covalent distances in some of the classes are small. Therefore, in this method components that have a proportion smaller or equal to 0.04 were removed. As seen in Figure 11 calculated covalent radii appear to be quite similar to ones derived using Method 4.

Table 8: Number of atoms for which calculated radii differ from published covalent radii tables in more than 0.5 Å.

Table	**Dataset 1**	**Dataset 2**
Meng (1991)	14	10
Blue Obelisk (2006)	6	4
Cordero (2008)	5	7
Pyykkö (2009)	7	4

Table 8 shows that compared to Method 4 a few more calculated radii differ from published tables in more than 0.5 Å. However, by comparing results to data derived using Dataset 2 it can be determined that the impact of random distribution to the results is very low. In order to confirm that derived results are acceptable, all cases where calculated radii differ in more than 0.5 Å need to be analysed. These cases are listed in Tables 9, 10, 11 and 12.

Table 9: Comparison with Meng (1991) in Å (14 classes).

Atom	**Atomic No.**	**Calculated**	**Meng (1991)**	**Diff**	**Type**
Li	3	1.42	0.68	0.7366	Alkali metals
Be	4	1.22	0.35	0.8676	Alkaline earth metals
Ne	10	0.91	1.50	-0.5865	Noble gasses
Na	11	1.88	0.97	0.9055	Alkali metals
Ar	18	3.64	1.51	2.1284	Noble gasses
K	19	1.83	1.33	0.5044	Alkali metals
Ca	20	1.82	0.99	0.8302	Alkaline earth metals
Kr	36	2.59	1.50	1.0945	Noble gasses
Rb	37	2.50	1.47	1.0335	Alkali metals
Sr	38	1.91	1.12	0.7914	Alkaline earth metals
Cs	55	2.44	1.67	0.7674	Alkali metals
Ba	56	2.17	1.34	0.8337	Alkaline earth metals
Ra	88	0.77	1.90	-1.1348	Alkaline earth metals
Pa	91	1.10	1.61	-0.5057	Actinoids

Table 10: Comparison with Blue Obelisk (2006) in Å (6 classes).

Atom	**Atomic No.**	**Calculated**	**Blue Obelisk (2006)**	**Diff**	**Type**
He	2	1.83	0.32	1.5133	Noble gasses
Ar	18	3.64	0.97	2.6684	Noble gasses
Br	35	1.65	1.14	0.5125	Reactive nonmetals
Kr	36	2.59	1.10	1.4945	Noble gasses
Tc	43	1.04	1.56	-0.5166	Transition metals
I	53	1.85	1.33	0.5237	Reactive nonmetals

Table 11: Comparison with Cordero (2008) in Å (5 classes).

Atom	**Atomic No.**	**Calculated**	**Cordero (2008)**	**Diff**	**Type**
He	2	1.83	0.28	1.5534	Noble gasses
Ar	18	3.64	1.06	2.5784	Noble gasses
Kr	36	2.59	1.16	1.4345	Noble gasses
Ra	88	0.77	2.21	-1.4448	Alkaline earth metals
Pa	91	1.10	2.00	-0.8957	Actinoids

Table 12: Comparison with Pyykkö (2009) in Å (7 classes).

Atom	**Atomic No.**	**Calculated**	**Pyykkö (2009)**	**Diff**	**Type**
He	2	1.83	0.46	1.3734	Noble gasses
Ar	18	3.64	0.96	2.6784	Noble gasses
Br	35	1.65	1.14	0.5125	Reactive nonmetals
Kr	36	2.59	1.17	1.4245	Noble gasses
I	53	1.85	1.33	0.5237	Reactive nonmetals
Ra	88	0.77	2.01	-1.2448	Alkaline earth metals
Pa	91	1.10	1.69	-0.5857	Actinoids

Comparison of all radii can be found in Appendix 5. Most of the cases where calculated covalent radii differ from published covalent radii tables in more than 0.5 Å can be explained:

- Elements that are categorised as noble gasses rarely create bonds as outer shell of valence electrons is full [Leab]. Therefore, covalent bond peak and Van der Waals gap cannot be identified in atom distance classes that involve these atoms.
- Alkaline earth metals have two electrons in the outer shell of the atom and are prone to form ionic bonds instead of covalent. However, beryllium is an exception, as it forms covalent bonds [Leaa].
- Remaining elements exceed the 0.5 Å difference limit compared to only a few on published covalent radii tables. The calculated values cannot be clearly evaluated as consensus for covalent radii of these elements does not exist in published tables.

4.6 Connectivity in SMILES

As described in Section 3.7, calculated covalent radii were tested by trying to determine connectivity between atoms and generate SMILES strings for 9 crystal structures in the COD. Results are provided in Table 13.

In 8 out of 9 cases SMILES strings generated using calculated radii match the strings generated using Meng (1991) radii. Calculated covalent radii in most cases exceed values provided in Meng (1991) covalent radii table. Hence, for structure with COD ID 7008430 when Meng (1991) covalent radii are used, there are two bromium atoms that create a separate molecule in the crystal. However, when calculated covalent radii are used all atoms in the crystal are bonded into a single molecule. After consulting the original publication it is clear that the result generated using Meng (1991) radii is the correct one. Failure of the calculated radii can be attributed to bromium covalent radius being too large.

Table 13: Comparison of SMILES strings generated using calculated radii and Meng (1991) radii.

COD ID	**Using Meng (1991) radii**	**Using calculated radii**
7007287	[OH2][Fe](Cl)(Cl)([OH2])([OH2])[OH2]	[OH2][Fe](Cl)(Cl)([OH2])([OH2])[OH2]
7007471	Cl[Pd]1(Cl)[NH]2CC[N]1(C)CC2	Cl[Pd]1(Cl)[NH]2CC[N]1(C)CC2
7007645	CCP([Te](Cl)Cl)(CC)CC	CCP([Te](Cl)Cl)(CC)CC
7007646	CCP([Te](Br)Br)(CC)CC	CCP([Te](Br)Br)(CC)CC
7007647	CCP([Te](I)I)(CC)CC	CCP([Te](I)I)(CC)CC
7008430	CC(P([Au](Br)(Br)Br)(C(C)C)C(C)C)C.BrBr	Br[Br][Br][Au](P(C(C)C)(C(C)C)C(C)C)(Br)Br
7008658	CCN1P(Cl)N(CC)P(N(P1Cl)CC)Cl	CCN1P(Cl)N(CC)P(N(P1Cl)CC)Cl
7008979	CC(O[Ta]1(OC(C)C)(OC(C)C)(OC(C)C)[O](C)[Ta]([O]1C)(OC(C)C)(OC(C)C)(OC(C)C)OC(C)C)C	CC(O[Ta]1(OC(C)C)(OC(C)C)(OC(C)C)[O](C)[Ta]([O]1C)(OC(C)C)(OC(C)C)(OC(C)C)OC(C)C)C
7009026	C[S]([Os](Cl)(Cl)(Cl)([NH3])[NH3])C	C[S]([Os](Cl)(Cl)(Cl)([NH3])[NH3])C

5 Conclusions

During the research all objectives have been achieved:

- Methodology for automated covalent radii derivation was created. Five different methods were analysed and removal of small proportioned components from best fitting mixture model (Method 5) was determined to be the most suitable as it manages to reduce the impact of the random distribution to derived covalent radii the most.
- An automated tool for covalent radii derivation was created[6] and it can be easily altered to test alternate versions of methods described in the research. Covalent radii were derived using data from an open-access database – COD (see Appendix 5).
- Calculated covalent radii were evaluated based on published covalent radii tables. Additionally, calculated covalent radii were tested by checking atom connectivity in molecules using SMILES nomenclature.

The determination of covalent radii could still be improved. One of these approaches should be tested by amending the created automated tool:

- Removal of classes that include chemical elements that rarely participate in covalent bonds (for example, noble gasses and alkaline earth metals) could improve derived covalent radii. These classes may affect derivation of covalent radii of other atoms as atom distances observed in these classes are usually larger and do not represent covalent bonds.
- Atom distance classes used in the research could be additionally refined by removing all instances that are larger than 10 Å. During the research it was noticed that Van der Waals bond peak should not exceed this limit in any of the analysed classes. Therefore, this restriction would allow to remove a big part of noise from the data, hence calculations could be more precise and still would not require usage of published Van der Waals radii tables.

---

[6] Located at `svn://saulius-grazulis.lt/Egle-Sidlauskaite/7semestras`, revision 294

References

[ABC$^+$79] F. H. Allen, M. D. Bellard, S.and Brice, B. A. Cartwright, A. Doubleday, H. Higgs, T. Hummelink, B. G. Hummelink-Peters, O. Kennard, W. D. S. Motherwell, J. R. Rodgers, and D. G. Watson. The Cambridge Crystallographic Data Centre: computer-based search, retrieval, analysis and display of information. *Acta Crystallographica Section B*, 35(10):2331–2339, 1979. `doi:10.1107/S0567740879009249`.

[AKW87] Frank H. Allen, Olga Kennard, and David G. Watson. Tables of Bond Lengths determined by X-Ray and Neutron Diffraction. Part I. Bond Lengths in Organic Compounds. *Journal of the Chemical Society, Perkin Transactions 2*, pages S1–S19, 1987. `doi:10.1039/P298700000S1`.

[Alv13] Santiago Alvarez. A cartography of the Van der Waals territories. *Dalton Trans.*, 42(24):8617–8636, 2013. URL: `http://dx.doi.org/10.1039/C3DT50599E`, `doi:10.1039/C3DT50599E`.

[CGPP$^+$08] Beatriz Cordero, Verónica Gómez, Ana E. Platero-Prats, Marc Revés, Jorge Echeverría, Eduard Cremades, Flavia Barragán, and Santiago Alvarez. Covalent radii revisited. *Dalton Transactions*, (21):2832–2838, 2008. `doi:10.1039/B801115J`.

[GBLW16] Colin R. Groom, Ian J Bruno, Matthew P. Lightfoot, and Suzanna C. Ward. The Cambridge Structural Database. *Acta Cryst*, B72:171–179, 2016. `doi:10.1107/S2052520616003954`.

[GDM$^+$12] Saulius Gražulis, Adriana Daškevič, Andrius Merkys, Daniel Chateignerand, Luca Lutterottiand, Miguel Quirós, Nadiezhda R. Serebryanaya, Peter Moeck, Robert T. Downs, and Armel Le Bail. Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. *Nucleic Acids Research*, 40:D420–D427, 2012. `doi:10.1093/nar/gkr900`.

[GHH+06] Rajarshi Guha, Michael T Howard, Geoffrey R Hutchison, Peter Murray-Rust, Henry Rzepa, Christoph Steinbeck, Jörg Wegner, and Egon L Willighagen. The Blue Obelisk-interoperability in chemical informatics. *Journal of Chemical Information and Modeling*, 46(3):991–998, 2006. `doi:10.1021/ci050400b`.

[GMVOK15] Saulius Gražulis, Andrius Merkys, Antanas Vaitkus, and Mykolas Okulič-Kazarinas. Computing stoichiometric molecular composition from crystal structures. *Journal of Applied Crystallography*, 48(1):85–91, 2015. `doi:10.1107/s1600576714025904`.

[Leaa] Lumen Learning. Introduction to Chemistry: The Alkaline Earth Metals. Accessed: 2021-05-15. URL: `https://courses.lumenlearning.com/introchem/chapter/the-alkaline-earth-metals/`.

[Leab] Lumen Learning. Introduction to Chemistry: The Noble Gases (Group 18). Accessed: 2021-05-15. URL: `https://courses.lumenlearning.com/introchem/chapter/the-noble-gases-group-18/`.

[Mer18] Andrius Merkys. *Extraction and Usage of Crystallographic Knowledge for Refinement and Validation of Molecular Models*. PhD thesis, Vilnius University, 2018.

[ML91] Elaine C. Meng and Richard A. Lewis. Determination of molecular topology and atomic hybridization states from heavy atom coordinates. *Journal of Computational Chemistry*, 12(7):891–898, sep 1991. `doi:10.1002/jcc.540120716`.

[MVB+16] Andrius Merkys, Antanas Vaitkus, Justas Butkus, Mykolas Okulič-Kazarinas, Visvaldas Kairys, and Saulius Gražulis. COD::CIF::Parser: an error-correcting CIF parser for the Perl language. *Journal of Applied Crystallography*, 49(1):292–301, 2016. `doi:10.1107/s1600576715022396`.

[OBJ+11] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison. Open Babel: An open chemical toolbox. *J. Cheminf.*, 33(3), 2011. `doi:10.1186/1758-2946-3-33`.

[OV14] Kliment Olechnovič and Česlovas Venclovas. A fast and reliable tool for computing the vertices of the voronoi diagram of atomic balls. *Journal of Computational Chemistry*, 35:672–681, 2014. `doi:10.1002/jcc.23538`.

[PA08] Pekka Pyykkö and Michiko Atsumi. Molecular Single–Bond Covalent Radii for Elements 1–118. *Chemistry—A European Journal*, 15(1):186–197, 2008. `doi:10.1002/chem.200800987`.

[PA09] Pekka Pyykkö and Michiko Atsumi. Molecular Double–Bond Covalent Radii for Elements Li–E112. *Chemistry—A European Journal*, 15(46):12770–12779, 2009. `doi:10.1002/chem.200901472`.

[Sch78] Gideon Schwarz. Estimating the Dimension of a Model. *Annals of Statistics*, 6:461–464, 1978. `doi:10.1214/aos/1176344136`.

[Š20] Eglė Šidlauskaitė. Atomų atstumų kristaluose tyrimas, 2020. Term paper.

Appendix 1

List of atoms for which calculated radii differ from published covalent radii in more than 0.5 Å using Method 1 with Dataset 1.

Table 14: Comparison with Meng (1991) in Å (27 classes).

Atom	**Calculated**	**Meng (1991)**	**Diff**
Li	1.45	0.68	0.7669
Be	1.33	0.35	0.9771
B	1.77	0.83	0.9433
C	2.58	0.68	1.8960
N	1.59	0.68	0.9082
O	1.93	0.68	1.2482
F	1.28	0.64	0.6450
Ne	0.69	1.50	-0.8052
Na	1.87	0.97	0.8965
P	1.80	1.05	0.7463
S	1.82	1.02	0.8024
Cl	1.58	0.99	0.5863
Ar	2.63	1.51	1.1208
K	2.01	1.33	0.6810
Se	1.73	1.22	0.5088
Rb	2.25	1.47	0.7808
Sr	1.76	1.12	0.6351
I	1.98	1.40	0.5775
Cs	2.42	1.67	0.7525
Ba	2.13	1.34	0.7882
Tl	2.10	1.55	0.5493
Po	2.25	1.68	0.5714
Ra	0.66	1.90	-1.2371
Th	1.27	1.79	-0.5211
Pa	0.98	1.61	-0.6289
Am	0.85	1.51	-0.6574
Cf	0.63	1.83	-1.1965

Table 15: Comparison with Blue Obelisk (2006) in Å (15 classes).

Atom	**Calculated**	**Blue Obelisk (2006)**	**Diff**
He	1.83	0.32	1.5133
B	1.77	0.82	0.9533
C	2.58	0.77	1.8060
N	1.59	0.75	0.8382
O	1.93	0.73	1.1982
F	1.28	0.71	0.5750
P	1.80	1.06	0.7363
S	1.82	1.02	0.8024
Cl	1.58	0.99	0.5863
Ar	2.63	0.97	1.6608
Se	1.73	1.16	0.5688
Kr	1.67	1.10	0.5667
Tc	0.99	1.56	-0.5691
I	1.98	1.33	0.6475
Tl	2.10	1.48	0.6193

Table 16: Comparison with Cordero (2008) in Å (20 classes).

Atom	**Calculated**	**Cordero (2008)**	**Diff**
He	1.83	0.28	1.5534
B	1.77	0.84	0.9333
N	1.59	0.71	0.8782
O	1.93	0.66	1.2682
F	1.28	0.57	0.7150
P	1.80	1.07	0.7263
S	1.82	1.05	0.7724
Cl	1.58	1.02	0.5563
Ar	2.63	1.06	1.5708
Se	1.73	1.20	0.5288
Kr	1.67	1.16	0.5067
I	1.98	1.39	0.5875
Tb	1.35	1.94	-0.5874
Tl	2.10	1.45	0.6493
Po	2.25	1.40	0.8514
Ra	0.66	2.21	-1.5471
Th	1.27	2.06	-0.7911
Pa	0.98	2.00	-1.0189
Pu	1.22	1.87	-0.6463
Am	0.85	1.80	-0.9474

Table 17: Comparison with Pyykkö (2009) in Å (20 classes).

Atom	**Calculated**	**Pyykkö (2009)**	**Diff**
He	1.83	0.46	1.3734
B	1.77	0.85	0.9233
C	2.58	0.75	1.8260
N	1.59	0.71	0.8782
O	1.93	0.63	1.2982
F	1.28	0.64	0.6450
P	1.80	1.11	0.6863
S	1.82	1.03	0.7924
Cl	1.58	0.99	0.5863
Ar	2.63	0.96	1.6708
Fe	1.72	1.16	0.5615
Se	1.73	1.16	0.5688
I	1.98	1.33	0.6475
Au	1.79	1.24	0.5455
Tl	2.10	1.44	0.6593
Po	2.25	1.45	0.8014
Ra	0.66	2.01	-1.3471
Pa	0.98	1.69	-0.7089
Am	0.85	1.66	-0.8074
Cf	0.63	1.68	-1.0465

Appendix 2

List of atoms for which calculated radii differ from published covalent radii in more than 0.5 Å using Method 2 with Dataset 1.

Table 18: Comparison with Meng (1991) in Å (20 classes).

Atom	**Calculated**	**Meng (1991)**	**Diff**
Li	1.45	0.68	0.7739
Be	1.51	0.35	1.1597
B	1.40	0.83	0.5715
C	1.65	0.68	0.9729
F	1.22	0.64	0.5812
Na	1.98	0.97	1.0121
Mg	1.64	1.10	0.5380
Cl	1.53	0.99	0.5426
Ar	3.11	1.51	1.5969
K	1.87	1.33	0.5355
Ca	1.76	0.99	0.7678
Kr	2.71	1.50	1.2140
Rb	2.40	1.47	0.9290
Sr	1.85	1.12	0.7259
I	2.05	1.40	0.6547
Cs	2.44	1.67	0.7730
Ba	2.11	1.34	0.7682
Tl	2.06	1.55	0.5119
Ra	0.73	1.90	-1.1658
Pa	1.08	1.61	-0.5280

Table 19: Comparison with Blue Obelisk (2006) in Å (14 classes).

Atom	**Calculated**	**Blue Obelisk (2006)**	**Diff**
He	1.83	0.32	1.5133
Be	1.51	0.90	0.6097
B	1.40	0.82	0.5815
C	1.65	0.77	0.8829
F	1.22	0.71	0.5112
Ne	1.65	0.69	0.9621
Cl	1.53	0.99	0.5426
Ar	3.11	0.97	2.1369
Se	1.72	1.16	0.5562
Br	1.66	1.14	0.5245
Kr	2.71	1.10	1.6140
Te	1.91	1.35	0.5635
I	2.05	1.33	0.7247
Tl	2.06	1.48	0.5819

Table 20: Comparison with Cordero (2008) in Å (16 classes).

Atom	**Calculated**	**Cordero (2008)**	**Diff**
He	1.83	0.28	1.5534
Be	1.51	0.96	0.5497
B	1.40	0.84	0.5615
F	1.22	0.57	0.6512
Ne	1.65	0.58	1.0721
Cl	1.53	1.02	0.5126
Ar	3.11	1.06	2.0469
Se	1.72	1.20	0.5162
Kr	2.71	1.16	1.5540
Te	1.91	1.38	0.5335
I	2.05	1.39	0.6647
Tb	1.44	1.94	-0.5032
Tl	2.06	1.45	0.6119
Ra	0.73	2.21	-1.4758
Pa	1.08	2.00	-0.9180
Pu	1.34	1.87	-0.5334

Table 21: Comparison with Pyykkö (2009) in Å (15 classes).

Atom	**Calculated**	**Pyykkö (2009)**	**Diff**
He	1.83	0.46	1.3734
B	1.40	0.85	0.5515
C	1.65	0.75	0.9029
F	1.22	0.64	0.5812
Ne	1.65	0.67	0.9821
Cl	1.53	0.99	0.5426
Ar	3.11	0.96	2.1469
Se	1.72	1.16	0.5562
Br	1.66	1.14	0.5245
Kr	2.71	1.17	1.5440
Te	1.91	1.36	0.5535
I	2.05	1.33	0.7247
Tl	2.06	1.44	0.6219
Ra	0.73	2.01	-1.2758
Pa	1.08	1.69	-0.6080

Appendix 3

List of atoms for which calculated radii differ from published covalent radii in more than 0.5 Å using Method 3 with Dataset 1.

Table 22: Comparison with Meng (1991) in Å (23 classes).

Atom	**Calculated**	**Meng (1991)**	**Diff**
Li	1.88	0.68	1.2049
Be	1.67	0.35	1.3203
B	1.41	0.83	0.5833
C	1.29	0.68	0.6119
F	1.17	0.64	0.5286
Ne	0.86	1.50	-0.6425
Na	1.98	0.97	1.0144
Mg	1.62	1.10	0.5220
S	1.57	1.02	0.5539
Cl	1.58	0.99	0.5865
Ar	3.47	1.51	1.9579
K	1.88	1.33	0.5531
Ca	1.72	0.99	0.7311
Ti	3.10	1.47	1.6289
Br	1.76	1.21	0.5487
Kr	2.42	1.50	0.9162
Rb	2.54	1.47	1.0671
Sr	1.92	1.12	0.7956
I	2.07	1.40	0.6651
Cs	2.56	1.67	0.8900
Ba	2.13	1.34	0.7852
Nd	2.89	1.81	1.0818
Ra	0.73	1.90	-1.1681

Table 23: Comparison with Blue Obelisk (2006) in Å (14 classes).

Atom	**Calculated**	**Blue Obelisk (2006)**	**Diff**
He	1.83	0.32	1.5133
Li	1.88	1.34	0.5449
Be	1.67	0.90	0.7703
B	1.41	0.82	0.5933
C	1.29	0.77	0.5219
S	1.57	1.02	0.5539
Cl	1.58	0.99	0.5865
Ar	3.47	0.97	2.4979
Ti	3.10	1.36	1.7389
Se	1.72	1.16	0.5587
Br	1.76	1.14	0.6187
Kr	2.42	1.10	1.3162
Te	1.85	1.35	0.5001
I	2.07	1.33	0.7351

Table 24: Comparison with Cordero (2008) in Å (17 classes).

Atom	**Calculated**	**Cordero (2008)**	**Diff**
He	1.83	0.28	1.5534
Li	1.88	1.28	0.6049
Be	1.67	0.96	0.7103
B	1.41	0.84	0.5733
F	1.17	0.57	0.5986
S	1.57	1.05	0.5239
Cl	1.58	1.02	0.5565
Ar	3.47	1.06	2.4079
Ti	3.10	1.60	1.4989
Se	1.72	1.20	0.5187
Br	1.76	1.20	0.5587
Kr	2.42	1.16	1.2562
I	2.07	1.39	0.6751
Nd	2.89	2.01	0.8818
Tl	1.96	1.45	0.5064
Ra	0.73	2.21	-1.4781
Pa	1.22	2.00	-0.7777

Table 25: Comparison with Pyykkö (2009) in Å (18 classes).

Atom	**Calculated**	**Pyykkö (2009)**	**Diff**
He	1.83	0.46	1.3734
Li	1.88	1.33	0.5549
Be	1.67	1.02	0.6503
B	1.41	0.85	0.5633
C	1.29	0.75	0.5419
F	1.17	0.64	0.5286
S	1.57	1.03	0.5439
Cl	1.58	0.99	0.5865
Ar	3.47	0.96	2.5079
Ti	3.10	1.36	1.7389
Se	1.72	1.16	0.5587
Br	1.76	1.14	0.6187
Kr	2.42	1.17	1.2462
I	2.07	1.33	0.7351
Nd	2.89	1.74	1.1518
Eu	2.20	1.68	0.5217
Tl	1.96	1.44	0.5164
Ra	0.73	2.01	-1.2781

Appendix 4

List of atoms for which calculated radii differ from published covalent radii in more than 0.5 Å using Method 4 with Dataset 1.

Table 26: Comparison with Meng (1991) in Å (12 classes).

Atom	**Calculated**	**Meng (1991)**	**Diff**
Li	1.47	0.68	0.7940
Be	1.20	0.35	0.8459
Ne	0.92	1.50	-0.5822
Na	1.83	0.97	0.8639
Ar	3.77	1.51	2.2640
Ca	1.54	0.99	0.5470
Kr	2.74	1.50	1.2435
Rb	2.04	1.47	0.5715
Sr	1.98	1.12	0.8629
Cs	2.30	1.67	0.6291
Ba	2.07	1.34	0.7277
Ra	0.77	1.90	-1.1308

Table 27: Comparison with Blue Obelisk (2006) in Å (4 classes).

Atom	**Calculated**	**Blue Obelisk (2006)**	**Diff**
He	1.83	0.32	1.5133
Ar	3.77	0.97	2.8040
Kr	2.74	1.10	1.6435
Tc	1.06	1.56	-0.5013

Table 28: Comparison with Cordero (2008) in Å (5 classes).

Atom	**Calculated**	**Cordero (2008)**	**Diff**
He	1.83	0.28	1.5534
Ar	3.77	1.06	2.7140
Kr	2.74	1.16	1.5835
Ra	0.77	2.21	-1.4408
Pa	1.27	2.00	-0.7333

Table 29: Comparison with Pyykkö (2009) in Å (4 classes).

Atom	**Calculated**	**Pyykkö (2009)**	**Diff**
He	1.83	0.46	1.3734
Ar	3.77	0.96	2.8140
Kr	2.74	1.17	1.5735
Ra	0.77	2.01	-1.2408

Appendix 5

Covalent radii table generated using Method 5 with Dataset 1 compared to published covalent radii tables.

Table 30: Calculated covalent radii table in comparison with published covalent radii tables in Å.

Atom	**Calculated**	**Pyykkö (2009)**	**Meng (1991)**	**Blue Obelisk (2006)**	**Cordero (2008)**
H	–	0.32	–	0.37	0.31
He	1.83	0.46	1.50	0.32	0.28
Li	1.42	1.33	0.68	1.34	1.28
Be	1.22	1.02	0.35	0.90	0.96
B	1.25	0.85	0.83	0.82	0.84
C	1.12	0.75	0.68	0.77	–
N	0.89	0.71	0.68	0.75	0.71
O	0.84	0.63	0.68	0.73	0.66
F	0.98	0.64	0.64	0.71	0.57
Ne	0.91	0.67	1.50	0.69	0.58
Na	1.88	1.55	0.97	1.54	1.66
Mg	1.43	1.39	1.10	1.30	1.41
Al	1.30	1.26	1.35	1.18	1.21
Si	1.35	1.16	1.20	1.11	1.11
P	1.36	1.11	1.05	1.06	1.07
S	1.33	1.03	1.02	1.02	1.05
Cl	1.38	0.99	0.99	0.99	1.02
Ar	3.64	0.96	1.51	0.97	1.06
K	1.83	1.96	1.33	1.96	2.03
Ca	1.82	1.71	0.99	1.74	1.76
Sc	1.59	1.48	1.44	1.44	1.70
Ti	1.33	1.36	1.47	1.36	1.60
V	1.20	1.34	1.33	1.25	1.53
Cr	1.19	1.22	1.35	1.27	1.39
Mn	1.21	1.19	1.35	1.39	–
Fe	1.30	1.16	1.34	1.25	–
Co	1.28	1.11	1.33	1.26	–
Ni	1.21	1.10	1.50	–	1.24
Cu	1.37	1.12	1.52	1.38	1.32
Zn	1.24	1.18	1.45	1.31	1.22
Ga	1.27	1.24	1.22	1.26	1.22
Ge	1.36	1.21	1.17	1.22	1.20
As	1.40	1.21	1.21	1.19	1.19
Se	1.58	1.16	1.22	1.16	1.20
Br	1.65	1.14	1.21	1.14	1.20

Atom	**Calculated**	**Pyykkö (2009)**	**Meng (1991)**	**Blue Obelisk (2006)**	**Cordero (2008)**
Kr	2.59	1.17	1.50	1.10	1.16
Rb	2.50	2.10	1.47	2.11	2.20
Sr	1.91	1.85	1.12	–	1.95
Y	1.56	1.63	1.78	1.62	1.90
Zr	1.71	1.54	1.56	1.48	1.75
Nb	1.51	1.47	1.48	1.37	1.64
Mo	1.35	1.38	1.47	1.45	1.54
Tc	1.04	1.28	1.35	1.56	1.47
Ru	1.28	1.25	1.40	1.26	1.46
Rh	1.30	1.25	1.45	1.35	1.42
Pd	1.38	1.20	1.50	1.31	1.39
Ag	1.49	1.28	1.59	1.53	1.45
Cd	1.62	1.36	1.69	1.48	1.44
In	1.54	1.42	1.63	1.44	1.42
Sn	1.53	1.40	1.46	1.41	1.39
Sb	1.52	1.40	1.46	1.38	1.39
Te	1.76	1.36	1.47	1.35	1.38
I	1.85	1.33	1.40	1.33	1.39
Xe	1.12	1.31	1.50	1.30	1.40
Cs	2.44	2.32	1.67	2.25	–
Ba	2.17	1.96	1.34	1.98	2.15
La	2.03	1.80	1.87	1.69	2.07
Ce	1.85	1.63	1.83	–	2.04
Pr	1.72	1.76	1.82	–	2.03
Nd	1.99	1.74	1.81	–	2.01
Pm	–	1.73	–	–	1.99
Sm	1.85	1.72	1.80	–	1.98
Eu	2.11	1.68	1.99	–	1.98
Gd	1.75	1.69	1.79	–	1.96
Tb	1.55	1.68	1.76	–	1.94
Dy	1.80	1.67	1.75	–	1.92
Ho	1.92	1.66	1.74	–	–
Er	1.94	1.65	1.73	–	1.89
Tm	1.85	1.64	1.72	–	1.90
Yb	1.63	1.70	1.94	–	1.87
Lu	2.00	1.62	1.72	1.60	1.87
Hf	1.56	1.52	–	1.50	1.75
Ta	1.44	1.46	–	1.38	1.70
W	1.42	1.37	1.37	1.46	1.62
Re	1.16	1.31	1.35	1.59	1.51
Os	1.32	1.29	1.37	1.28	1.44
Ir	1.32	1.22	–	1.37	1.41
Pt	1.40	1.23	1.50	1.28	1.36
Au	1.50	1.24	–	1.44	1.36
Hg	1.65	1.33	1.70	1.49	1.32

Atom	**Calculated**	**Pyykkö (2009)**	**Meng (1991)**	**Blue Obelisk (2006)**	**Cordero (2008)**
Tl	1.88	1.44	1.55	1.48	1.45
Pb	1.75	1.44	1.54	1.47	1.46
Po	–	1.45	1.68	–	1.40
At	–	1.47	–	–	1.50
Rn	–	1.42	–	1.45	1.50
Fr	–	2.23	–	–	2.60
Ra	0.77	2.01	–	–	2.21
Ac	–	1.86	–	–	2.15
Th	1.70	1.75	1.79	–	2.06
Pa	1.10	1.69	1.61	–	2.00
U	1.53	1.70	1.58	–	1.96
Np	1.90	1.71	1.55	–	–
Pu	1.54	1.72	1.53	–	1.87
Am	1.94	1.66	1.51	–	1.80
Cm	–	1.66	–	–	1.69
Bk	–	1.68	–	–	–
Cf	1.72	1.68	–	–	–
Es	–	1.65	–	–	–
Fm	–	1.67	–	–	–
Md	–	1.73	–	–	–
No	–	1.76	–	–	–
Lr	–	1.61	–	–	–
Rf	–	1.57	–	–	–
Db	–	1.49	–	–	–
Sg	–	1.43	–	–	–
Bh	–	1.41	–	–	–
Hs	–	1.34	–	–	–
Mt	–	1.29	–	–	–
Ds	–	1.28	–	–	–
Rg	–	1.21	–	–	–
Cn	–	1.22	–	–	–
Uut	–	1.36	–	–	–
Fl	–	1.43	–	–	–
Uup	–	1.62	–	–	–
Lv	–	1.75	–	–	–
Uus	–	1.65	–	–	–
Uuo	–	1.57	–	–	–