

RESEARCH

Open Access



Determination of bonding radii from small-molecule crystal structures

Eglė Šidlauskaitė^{1,2}, Andrius Merkys^{1*} , Antanas Vaitkus¹ , Algirdas Grybauskas¹ and Saulius Gražulis^{1,3}

Abstract

X-ray crystallography rarely captures chemical bonding between atoms of a structure in question. Most of the time distance-based heuristics are applied to establish the pairs of bonded atoms. One class of such heuristics depends on a set of bonding radii that estimate the idealised size of each chemical element in a bond. This publication describes an unsupervised workflow for deriving a bonding radii set from crystal structure data in the Crystallography Open Database.

Scientific contribution

An atomic bonding radii set is derived from the Crystallography Open Database using a novel, unsupervised and fully automated workflow. The method does not use any prior knowledge about the connectivity in input crystal structures, relying solely on their Voronoi tessellations.

Keywords Bonding radii, Connectivity, Crystallography

Introduction

Chemical connectivity in crystal structures is of interest to many derivative studies. Connectivity is an unavoidable prerequisite for, to name a few, the identification of chemical substances [1, 2], analysis of molecular geometry [3] and crystal properties [4]. X-ray crystallography – a well-established technique for the determination of 3D atom coordinates in crystal structures – is in practice rarely used to detect chemical connectivity, which, although possible using charge density analysis, is very demanding for data quality and the expertise of the

researcher [5]. Thus heuristics are often applied to derive the connectivity, usually based either on per-element contribution called a *bonding radius* or maximal bonding distance defined per each pair of elements [6]. The former is an often used bond determination heuristic which states that the distance between two chemically bonded atoms has to be smaller than or equal to the sum of their bonding radii [7–11]. Thus two atoms of chemical elements *i* and *j* separated by distance $d_{i,j}$ are treated as bonded if the following inequality holds:

$$d_{i,j} \leq r_i + r_j + \epsilon \quad (1)$$

Here r_i is the bonding radius of chemical element *i* and ϵ is a sensitivity constant, usually ranging between 0.35 Å and 0.45 Å [12]. Such radii are not constant intrinsic properties of atoms as their sizes are highly dependent on the environment [13]. Nevertheless bonding radii are approximated well enough and used in the distance-based connectivity criterion.

Although there are a number of published bonding radii sets [7, 13–16], a univocal method for bonding

*Correspondence:

Andrius Merkys
andrius.merkys@gmc.vu.lt

¹ Sector of Crystallography and Chemical Informatics, Institute of Biotechnology, Life Sciences Center, Vilnius University, Saulėtekio al. 7, LT-10257 Vilnius, Lithuania

² Department of Genetics, Cell Biology and Anatomy, University of Nebraska Medical Center, 42nd and Emile, Omaha, NE 68198, USA

³ Faculty of Mathematics and Informatics, Vilnius University, Naugarduko g. 24, LT-03225 Vilnius, Lithuania



radii derivation has not been established. A widely used bonding radii set published as a part of the Cambridge Structural Database (CSD) [17] has received critique for manual adjustments of radii aimed to fit specific molecules as well as default values being used in place of unknown radii [13]. In many other studies [7, 13] bonding radii sets are derived from the CSD crystal structures that are annotated with expert-curated connectivity information. Connectivity as derived by experienced chemists is usually of high quality, however, manual marking of connectivity is a slow and expensive process. Moreover, the CSD operates under a proprietary license, restricting the spread of the derivative datasets and preventing the replicability of study results. In addition, values in bonding radii sets can differ quite significantly, in some cases exhibiting differences of over 1 Å, as evident from Fig. 1. The need for an automated bonding radii derivation approach is addressed in recent studies published by Nikolaenko and coworkers [18, 19]. However, the main limitation of these studies is that the derived datasets involve only a few select elements thus reducing their applicability. For a more detailed review of currently available bonding radii sets see [20].

Bonding radii sets are usually interaction-type-specific. In this study we aim to determine radii encompassing all of the intramolecular interactions (covalent, ionic, metallic bonding, etc.) because such radii could be used to determine bonding in crystallographic structures. In the study published by Alvarez [21] an interatomic distance distribution is used to determine the typical distance between atoms that interact mainly by the van der Waals forces. In a typical interatomic distance distribution

intramolecular and intermolecular interactions are separated by the *van der Waals gap* due to a significant difference in the interaction strength. Identification of the van der Waals gap should be sufficient to determine the longest possible interatomic distance for each pair of elements that would still allow an intramolecular bond, at the same time rejecting longer intermolecular interactions. Throughout our study we maintain the view of interatomic distance distributions as postulated by Alvarez, on the other hand acknowledging departures from it due to the vast variability of chemical environments.

In this study we have developed an automated and unsupervised workflow for the derivation of bonding radii sets from any small-molecule crystallographic data, expressed in CIF files using the IUCr coreCIF dictionary data items. The set derived from the data in the Crystallography Open Database (COD) [22] is validated and compared to other similar bonding radii sets.

Methods

Input data

We used all 519 021 crystal structure records from the COD revision 295629 as input for our study. In the COD, crystal structures are stored as the Crystallographic Information File (CIF) format [23] files and are therefore directly suitable for the workflow described here. While the coreCIF dictionary defines data names for bond lengths (GEOM_BOND category), this information mostly concerns bonded interactions and is not guaranteed to be complete or free from mistakes [3], thus we analysed the atomic positions only.

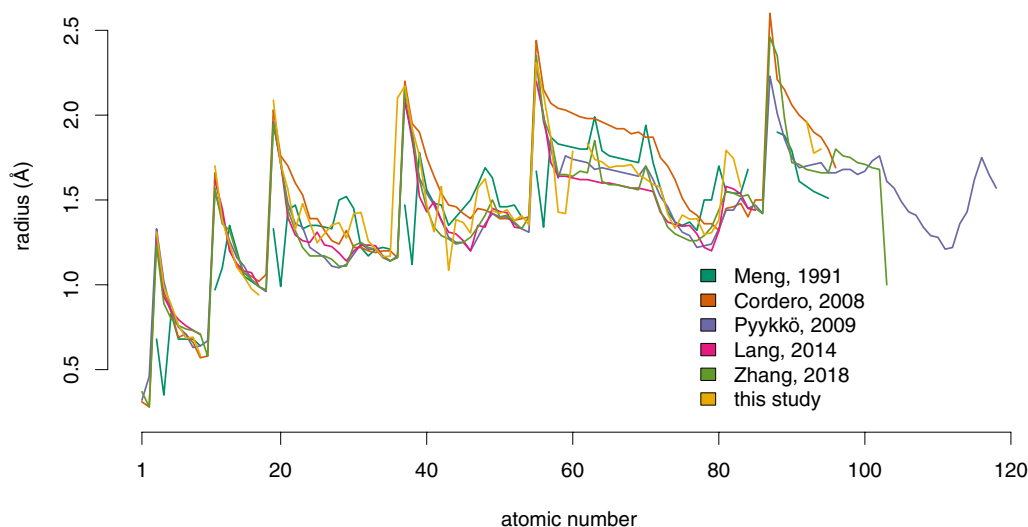


Fig. 1 Bonding radii sets

Structure selection criteria

We imposed several criteria to select only complete, experimentally measured structures and reject various biases and outliers. We strove to use only the structures having 3D coordinates for all of their sites, including the solvent molecules. This criterion stems from the need to correctly identify direct intramolecular and intermolecular interactions. Due to the difficulties in addressing disordered structures, we have decided to exclude them altogether. We employed several heuristics to exclude theoretical structures and the ones having one or more atoms which were not backed by experimental evidence. Outlier rejection also removes structures solved under unusual conditions such as high pressure.

We have excluded the following kinds of structures from our calculations:

- Disordered structures. We exclude structures that are either explicitly marked as disordered in their CIF files or that are detected as disordered by the `cif_mark_disorder` program from the *cod-tools* v3.10.0 package [24]. This is done because disordered structures are difficult to accommodate in the present analysis. Such structures provide either alternative sites for a set of atoms (positional disorder), alternative chemical elements occupying the same site (compositional disorder), or a combination of both. There is no clear solution for handling such structures as input for geometric analysis. One possible method would be to generate all possible combinations of disordered fragments and exclude the repetitive measurements from the stable part of the structure. However, alternative atom positions may affect the whole Voronoi partition rather than just the immediate environment of the disordered fragment. Moreover, Voronoi partition may depend on disorder in neighbouring crystal cells. Supercell modelling approach might offset the former problem for the price of increased computation complexity. Due to this complexity we decided to exclude disordered structures altogether.
- Structures with partially occupied atom sites. Partially occupied sites appear in structures with vacancy defects, crystallographically disordered structures as well as improperly modelled structures. All these cases are difficult to accommodate in the current statistical analysis, thus the corresponding structures are excluded.
- Structures with different declared and calculated formulae. The difference in heavy atom numbers between formulae usually means that the described structure is incomplete.
- Structures having attached hydrogen atoms. When the 3D coordinates of hydrogen atoms cannot be determined, their presence near a specific heavy atom is sometimes instead indicated as the number of attached hydrogen atoms (also known as virtual hydrogen atoms) using the `_atom_site_attached_hydrogens` CIF data item. Since in this case the 3D coordinates of the hydrogen sites are unknown, we cannot precisely identify contacts involving them. While it would be possible to ignore only the interactions of the atoms having attached hydrogen atoms, we have decided to exclude such structures altogether since they comprise less than 1% of the COD.
- Structures having dummy sites. Dummy sites in CIF files do not pertain to actual atom positions, but they may signal that certain atoms are missing from the structure. Therefore it is safer to exclude such structures. There are 420 such entries in the COD in total.
- Duplicate entries. The COD contains a couple thousand of duplicated entries, marked as such by its maintainers. These entries are excluded to avoid overrepresentation of certain observations.
- Structures with calculated heavy atom sites. We exclude structures with explicitly marked calculated sites to retain only experimentally located atoms.
- Structures solved in superspace groups. Superspace group structures are difficult to accommodate in our current analysis due to interatomic distances in them being dependent on modulation waves spanning many crystal cells.
- Theoretical structures. We base our insights on experimentally determined structures, thus we filter out COD entries that were marked as theoretical by the curators of the database.
- Structures with unmodelled solvent. Solvent is rarely interesting in structural analysis, moreover, disordered or mobile molecules may be difficult to represent in CIF files. Non-modelled solvent molecules are sometimes reported using the `_platon_squeeze_void_count_electrons` or `_smtbx_masks_void_count_electrons` CIF data items, thus we have excluded all structures containing these data items (27 189 in total). However, these data items are often absent, thus we also rely on other filters (formulae mismatch, unrealistic densities, etc.) to identify such structures.
- Unusual crystal densities. We have excluded crystals with density lower than $0.17 \frac{\text{g}}{\text{cm}^3}$ (density of extremely low density structures as reported in [25]) or higher than $22.59 \frac{\text{g}}{\text{cm}^3}$ (density of osmium as reported in [26]), both declared and calculated from formula and cell parameters. To ensure consistency between

the declared crystal density and the calculated one, we exclude structures with the ratios of said densities outside the range of [0.75, 1.25].

- Structures with large absolute density differences between original and relaxed form. To exclude incomplete or otherwise unrealistic crystal structures we have removed from consideration the ones which significantly diverge when relaxed using the M3GNet [27] graph deep learning interatomic potential (Justinas Šlepavičius, personal communication). It should be noted, however, that the relaxation appropriately preserves naturally porous structures such as metal-organic frameworks, as their framework structure is strong enough to be kept in place. Since the relaxation process can considerably alter both the cell parameters and atom coordinate values without actually changing the structure (this usually happens when there is a change in the unit cell basis), we based our exclusion criterion on the difference of crystal structure densities. Thus we exclude structures undergoing absolute density change of $0.5 \frac{\text{g}}{\text{cm}^3}$ or more.
- Structures with high R factor values. We require the crystallographic R factor to be known and be ≤ 0.1 , per the IUCr validation guidelines [28]. Requirement for R factor to be known is ignored for neutron diffraction structures in order to retain as many such structures as possible.
- Structures with high weighted R factor values. We require the weighted crystallographic R factor (wR) to be known and be ≤ 0.25 , per the IUCr validation guidelines [29]. Requirement for wR factor to be known is ignored for neutron diffraction structures in order to retain as many such structures as possible.
- Structures measured under extreme temperature. Extreme temperatures (higher than 320 K) may affect the crystal geometry.
- Structures measured under extreme pressure. Extreme pressures (lower than 80 kPa or higher than 120 kPa) may affect crystal geometry.

Most of the filters are constructed using the output produced by the `COD::CIF::Data::CIF2COD` and `COD::CIF::Data::CODFlags` modules from *cod-tools*. All filters were applied sequentially on the input data. In total, 249 032 entries were used for further analysis.

Structure processing workflow

The COD contains crystal structures described by their asymmetric units plus their space group. This description is enough to reconstruct all atom positions in the unit cell. We reconstruct the symmetrically equivalent

positions with the `cif_fill_cell` script from *cod-tools* [30]. The same tool constructs a $3 \times 3 \times 3$ supercell to represent all interactions where atoms from the central unit cell participate.

Interaction measurements

We have harnessed Voronoi tessellation to detect the pairs of possibly interacting atoms. To perform the tessellation, we have chosen an algorithm implemented in *voronota* v1.18.1877 [31]. For each pair of directly contacting atoms *voronota* calculates both the contact distance and the contact area which is later used as a weight of a contact. In *voronota*, which is mainly used to locate and describe contacts in protein structures, a probe is used to cut off distant pairwise interactions. We have disabled the probe feature in our calculations by setting it to the longest diagonal of the supercell, as we are interested in all pairwise interactions regardless of their distance. To use *voronota* in our calculations, we have wrapped it into a tool named `cif_contacts` of the *crystal-contacts* package [32], revision 56.

Due to the abundance and relatively low precision of their site 3D coordinates determined by X-ray diffraction, hydrogen atoms require special treatment. While hydrogen coordinates in neutron diffraction structures can usually be trusted, explicit hydrogen sites in X-ray structures usually are imprecise and/or derived using geometric constraints. Thus taking their contacts into consideration is unlikely to result in meaningful radii values. Nevertheless, however imprecise, hydrogen sites act as natural “barriers” separating heavier atoms with distant interactions. Therefore we use all hydrogen atom sites in the input for Voronoi tessellation, but for non-neutron diffraction structures we afterwards discard all the observed contacts involving hydrogen atoms. Neutron diffraction structures, on the other hand, are processed differently: in them, hydrogen atoms are treated just as any other atoms and their contacts are preserved. This is done to obtain a bonding radius for hydrogen. After filtering the COD structures as described in [Structure selection criteria](#), 489 neutron diffraction structures remain in the filtered dataset.

After finding and calculating all the contacts from the $3 \times 3 \times 3$ supercell with `cif_contacts`, we retain only those contacts that involve at least one atom from the asymmetric unit. This is done by `cif_bonds_angles` from the *atomclasses* package [3]. By removing interactions between symmetric equivalents we prevent overrepresentation of contacts.

From the total pool of calculated interatomic distances we exclude unusually short and unusually long observations. Distances involving hydrogen (from neutron diffraction structures only) are limited to the inclusive range

of [0.5, 6] Å, the other distances being limited to the inclusive range of [1, 6] Å. Cutoff on the lower end is imposed to filter out steric clashes, possibly arising due to unmarked and undetected disorder. Cutoff on the upper end is supposed to reduce the influence of very distant interactions, for example, in MOFs or other porous crystals. The upper limit is derived based on the longest possible covalent bond permitted by the covalent radii set of Cordero et al. [13] (5.2 Å for Fr–Fr).

Determination of the bond peak

Having processed all pairs of directly interacting atoms in the input data, we weigh each observation by the area of its contact [33] to reduce the influence of long distances, group them by chemical elements and for each pair of elements (we call it a *class*) fit a density model and identify the *bond peak*, which we deem to contain the longest intramolecular interactions. Each of these steps is described in more detail in the following paragraphs.

To facilitate the analysis of the density of a class, we fit a Gaussian mixture model on the sample composed of the class distances (Fig. 2). Fitting is done using the expectation maximisation algorithm with a modification to accept weights. In order to reduce the influence of long nonbonded interactions we use the following weighting scheme for interaction observed between atoms i and j :

$$w_{ij} = 1 - (1 - \frac{S_{ij}}{S_i})(1 - \frac{S_{ij}}{S_j}) \quad (2)$$

Here S_i is the total sum of all the contact areas for atom i , and S_{ij} is the contact area between atoms i and j . To find the best fitting Gaussian mixture model for each class, we use the expectation maximisation method to determine parameters for mixtures with the number of normal components ranging from 1 to 10. Classes larger than 200 000 observations are reduced by randomly picking 200 000 observations from them. On the other end, we require a minimal 1:30 ratio of model parameters to observations. This means that for a mixture of N Gaussian components to be considered, a sample needs at least $30 \times (3N - 1)$ observations. Component centres for the initial steps are taken from sample quantiles. From the fitted mixtures, all components accounting for less than 10 observations are removed to avoid fitting to the possible outliers. From these ten mixtures we select the one with an optimal Bayesian information criterion (BIC [34]) value. Both the fitting and the BIC computation is performed using *MixtureFitting* v0.7.0 [35] package written in *R*.

We have tailored the method for bond peak selection based on the chemical elements participating in the class:

- For classes involving at least one alkali or alkaline earth metal, the first mixture component is deemed to be the bond peak, as these elements very rarely form higher order bonds.

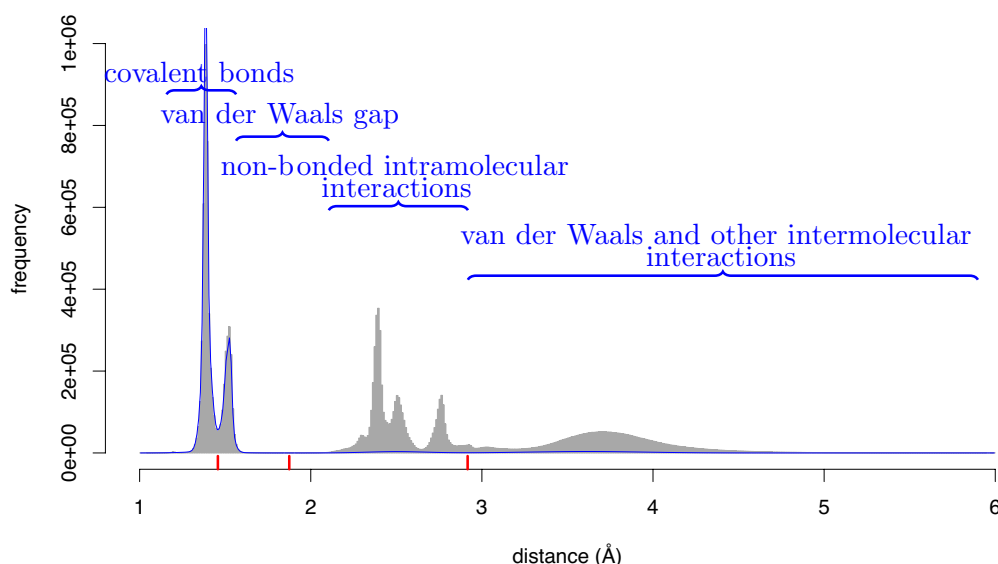


Fig. 2 Histogram of C–C interatomic distances in the COD. Brackets mark approximate distance regions populated by observations from certain interactions. Blue curve shows the fitted mixture model density. Red ticks mark the antimodes as detected by our approach. Note the proportion differences between components matching covalent bonds and noncovalent interactions. The mismatch between the histogram and model density in range [2, 6] Å is due to the observation weighting as given in Eq. 2

- For interactions between halogens and non-organogenic elements (not C, N, O, P, S, Se), the first mixture component is deemed to be the bond peak.
- For interactions between oxygen and U, Np and Pu, the last mixture component is deemed to be the bond peak.
- For the rest, the component preceding the lowest-density antimode (supposedly corresponding to the van der Waals gap) is deemed to be the bond peak.

Different treatment of classes based on the types of elements participating in them is necessary to accommodate the variability of chemical interactions.

To select the lowest-density antimode, we inspect the density zones between each pair of neighbouring component centres. As a midpoint between the component centres might fall on a slope of the density function, we employ the Nelder–Mead method [36] to locate the actual minimum. If the Nelder–Mead method locates the density minimum outside the range limited by neighbouring component centres, it is held that these components are inseparable.

More rules dictated by the common chemical sense could be further applied to select the correct bond peaks. In one instance, the F–F class, we have decided to completely remove the class from consideration due to its negative influence on the overall bonding radii derivation (described in section [Bonding radii](#)). Proper F–F bonds are too rare in the COD to form a detectable bond peak (Fig. 3), thus a peak formed by non-bonded interaction distances is chosen as the bond peak; no van der Waals

gap is visible. While such aberrations are possible under our methodology, we expect them to be dominated by correctly determined bond peaks. Unfortunately, F–F class is the most populated among all F-involving classes, thus the overly long bonding distance is assigned the largest weight. In order to offset the influence of F–F class we have decided to remove this class altogether.

Bonding radii

We have derived the bonding radii by approximately solving an overdetermined equation system via the least squares method:

$$\begin{cases} W_{Ac,Ac} \cdot (r_{Ac} + r_{Ac}) = W_{Ac,Ac} \cdot d_{Ac,Ac} \\ W_{Ac,Ag} \cdot (r_{Ac} + r_{Ag}) = W_{Ac,Ag} \cdot d_{Ac,Ag} \\ W_{Ac,Al} \cdot (r_{Ac} + r_{Al}) = W_{Ac,Al} \cdot d_{Ac,Al} \\ \dots \\ W_{Zr,Zr} \cdot (r_{Zr} + r_{Zr}) = W_{Zr,Zr} \cdot d_{Zr,Zr} \end{cases} \quad (3)$$

Where r_i is the bonding radii for chemical element i , $d_{i,j}$ is the bond peak centre in class of chemical elements i and j (including $i = j$), and $W_{i,j}$ is the weight of the equation equal to the number of observations in the class. Weights are used to increase the influence of larger classes in the least squares calculation, as their models should be more accurate. *R* package *Rlinsolve* v0.3.2 [37] was used to approximately solve the system of equations.

Since both the mixture model fitting and the antimode selection could be affected by peculiarities of certain classes, we have applied an iterative outlier removal procedure to arrive to self-consistent radii set. During each iteration, Eq. 3 is approximately solved and the

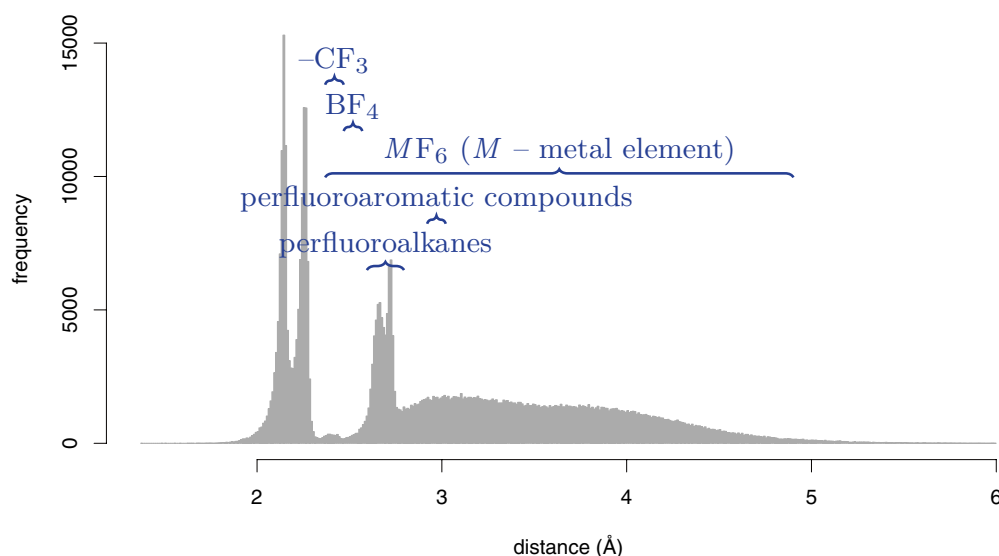


Fig. 3 Histogram of F–F interatomic distances in the COD. Brackets mark approximate distance regions populated by observations from certain chemical compounds

largest class having the difference between its bond peak and radii sum greater than $0.35 \text{ \AA}$ is selected. Then the selected class is excluded from Eq. 3 and the next iteration commences. This is repeated until no classes with $d_{i,j} - (r_i + r_j) > 0.35 \text{ \AA}$ are found. It is important to note that the difference in this criterion is not absolute, that is, only classes with bond peak position larger than radii sum are excluded. This is done in order to eliminate classes where the bond peak centres might be detected incorrectly due to the relative abundance of long intermolecular distances.

We have also attempted to use the iterative reweighting approach for the least squares calculation (IRLS) [38], which iteratively decreases the weights of equations producing the largest errors in least squares calculation. However, IRLS did not perform better than our iterative outlier removal approach, which is specifically tailored to alleviate the shortcomings of our methods.

Results

Overview

In total, 3107 contact classes were observed, consisting of 97 unique chemical elements (Fig. 5). Mixture models were derived for 954 classes. This number excludes classes with too few observations to reliably fit even a one-component Gaussian mixture model. Correspondingly, coverage of the periodic table decreased to 84 chemical elements. Although BIC was applied as a measure to prefer simpler mixture models, there is a tendency of larger classes to be approximated with mixtures

containing greater number of components, as seen in Fig. 4.

The resulting radii set (Table 1) is close to the one derived by Pyykkö and Atsumi [14], especially in treatment of organogenic, alkali and alkaline earth metals where the absolute differences for most elements do not exceed $0.075 \text{ \AA}$. With the exception of considerably different radii for Kr, Pr, Tl and Pb, the largest deviations between these two radii sets are observed in the d-block elements. A comparison with published bonding radii sets is provided in Fig. 1

Validation

Comparison with a chemical perception pipeline

The derived bonding radii set was evaluated for its suitability for interatomic bonding determination in crystallographic data. For this purpose, we analysed the bonding presented in a collection of chemically annotated crystallographic structures in SDF format that were previously derived from the COD data using a chemical perception pipeline [2]. The chemical perception pipeline establishes the bonding based on the bonding radii set compiled by Pyykkö and Atsumi paired with the following *ad hoc* rules as implemented in the `cif-perceive-chemistry` program version 0.4.0 [39], therefore it can be viewed as a completely independent approach.

- Removal of incorrectly assigned intraring bonds. Due to short interatomic distances in flat 4-membered rings, an extra covalent bond is sometimes

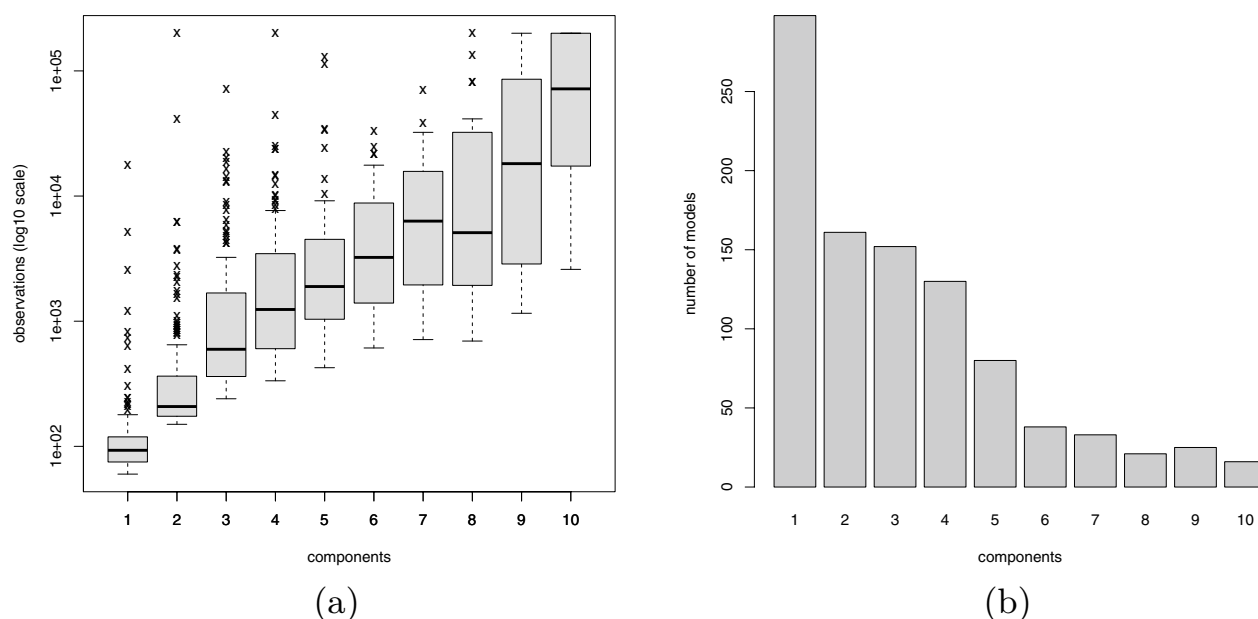


Fig. 4 Statistics of the selected models. **a** Shows the dependence of the number of mixture components in models on the number of observations, vertical axis is in log₁₀ scale. **b** Shows number of model components versus the number of models

Table 1 Bonding radii set derived using the described approach

Atom	Radius (Å)	Z
H	0.25	1
He	–	2
Li	1.31	3
Be	0.99	4
B	0.88	5
C	0.76	6
N	0.68	7
O	0.69	8
F	0.58	9
Ne	–	10
Na	1.70	11
Mg	1.38	12
Al	1.24	13
Si	1.10	14
P	1.05	15
S	0.98	16
Cl	0.94	17
Ar	–	18
K	2.09	19
Ca	1.71	20
Sc	1.56	21
Ti	1.32	22
V	1.48	23
Cr	1.36	24
Mn	1.25	25
Fe	1.30	26
Co	1.35	27
Ni	1.37	28
Cu	1.28	29
Zn	1.42	30
Ga	1.43	31
Ge	1.26	32
As	1.21	33
Se	1.17	34
Br	1.17	35
Kr	2.10	36
Rb	2.17	37
Sr	1.92	38
Y	1.76	39
Zr	1.47	40
Nb	1.31	41
Mo	1.58	42
Tc	1.09	43
Ru	1.38	44
Rh	1.37	45
Pd	1.31	46
Ag	1.57	47
Cd	1.63	48
In	1.44	49

Table 1 (continued)

Atom	Radius (Å)	Z
Sn	1.42	50
Sb	1.44	51
Te	1.38	52
I	1.41	53
Xe	1.35	54
Cs	2.31	55
Ba	2.12	56
La	1.87	57
Ce	1.43	58
Pr	1.42	59
Nd	1.79	60
Pm	–	61
Sm	1.83	62
Eu	1.74	63
Gd	1.72	64
Tb	1.69	65
Dy	1.70	66
Ho	1.70	67
Er	1.71	68
Tm	1.65	69
Yb	1.62	70
Lu	1.59	71
Hf	1.57	72
Ta	1.42	73
W	1.33	74
Re	1.41	75
Os	1.39	76
Ir	1.39	77
Pt	1.30	78
Au	1.31	79
Hg	1.38	80
Tl	1.79	81
Pb	1.75	82
Bi	1.58	83
Po	–	84
At	–	85
Rn	–	86
Fr	–	87
Ra	–	88
Ac	–	879
Th	1.73	90
Pa	–	91
U	1.96	92
Np	1.78	93
Pu	1.80	94

placed between the diagonal atoms. Since non-flat rings may contain legitimate diagonal bonds (e.g., bicyclobutane, tetrahedrane derivatives), such bonds

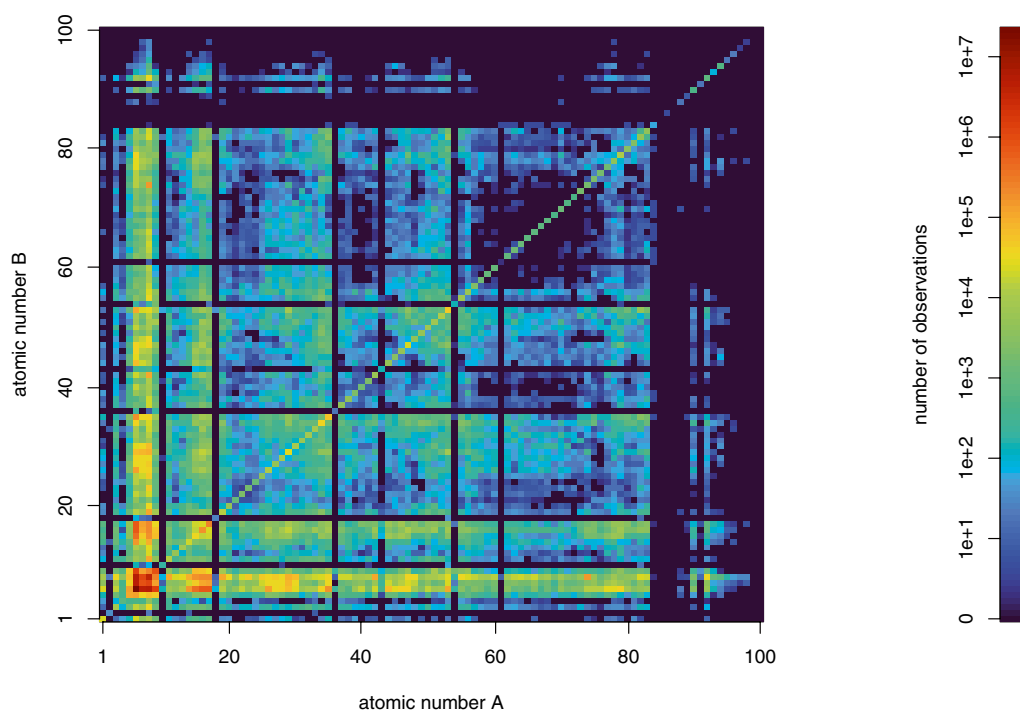


Fig. 5 Prevalence of pairwise distances between specific elements observed in the COD

are removed, but only if the absolute dihedral angle between the atoms of the ring is smaller than 20° (see Fig. 6).

- Removal of incorrectly assigned metal-ligand bonds. Short interatomic distances can also result in spurious bonding between a metal centre and atoms in
- its second coordination sphere, i.e., atoms covalently bound to directly coordinated ligands. Such bonds are therefore removed.
- Addition of bonds to aromatic π -ligands. Steric effects and bulky substituents may increase metal-ligand distances thus leading to underestimation of

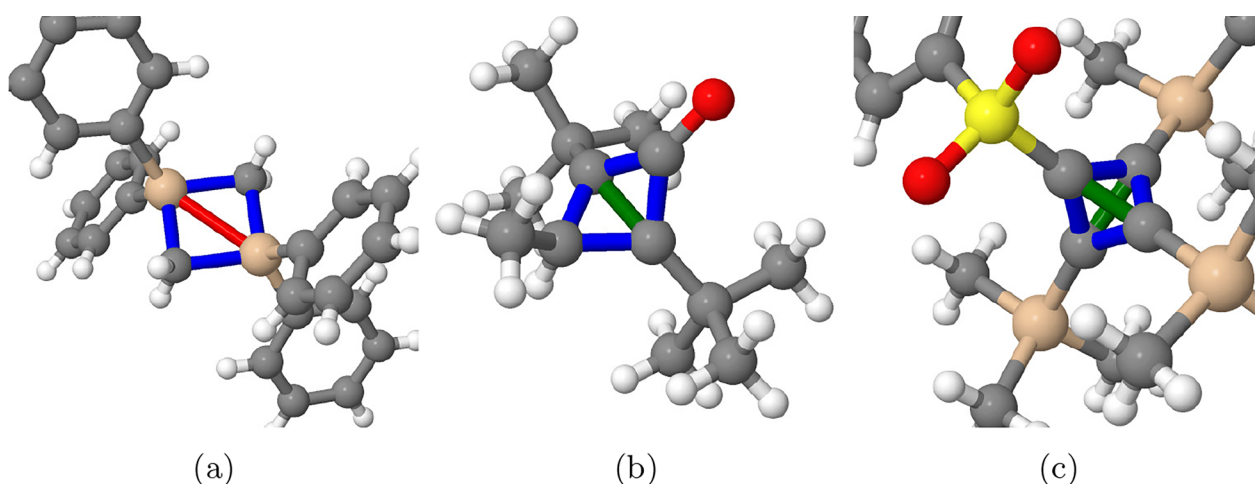


Fig. 6 Examples of intraring bonds in 4-membered rings that would be detected in COD entries [4077811](#) (a), [4100665](#) (b) and [4106489](#) (c) when judged on the atom radii sum criterion alone. Ring connections are coloured blue, legitimate intraring connections are coloured green, and false-positive intraring bonds are coloured red. While the diagonal intraring C–C bonds in bicyclobutane (b) and tetrahedrane (c) derivatives are legitimate, the Si–Si bond in (a) is a false-positive. Given the angle sums of 360° (b), 336° (b) and 237° (c) and the $[340, 380]^\circ$ ring flatness restriction, only the incorrect Si–Si bond would be removed during the bond pruning stage

ligand hapticity (e.g., a cyclopentadienyl ligand may be marked as connected only by 3 atoms instead of the expected 5). Therefore, if an atom ring is flat and more than half of the ring atoms have connections to the same metal atom, the ring is treated as being fully coordinated by that metal atom and metal-ligand bonds are assigned to all ring atoms.

Only the structures which passed all of the data quality checks imposed by the pipeline (data consistency, chemical validity, etc.) are used in the analysis to ensure that the presented connectivity reflects proper intramolecular bonds.

We subjected the coordinates-only part of chemically annotated structures to connectivity detection using Eq. 1 ($\epsilon = 0.45$ Å) looking for two types of bonding differences: missing bonds (input structure lacks an explicit bond between atoms) and overlong bonds (input structure contains an explicit bond which is deemed to be too long by our new criteria). The connectivity detection also reproduces the flat 4-membered ring diagonal bond removal employed by the chemical perception pipeline. This is done by identifying pairs of triangles sharing an edge and having sums of their angles within $[340, 380]^\circ$. Every such pair is treated as constituting a flat 4-membered ring and their shared bond is removed.

In over 365,000 input structures 80,487 mismatches were detected. Around 92% of them are due to the new bonds assigned by our radii set, with the rest being missing (i.e. unreproduced) bonds. Classes with the most differences of both kinds are shown in Table 2, with classes involving carbon and other organogenic elements provided for reference. In both the newly assigned and unreproduced bonds, certain bond classes led to improved bonding accuracy while others introduced errors.

The majority of the newly detected bonds are between the same metal atoms, highlighting the fact that our radii for these elements are longer than the ones derived by Pyykkö and Atsumi. For Mo and V elements, the elongation of bonding radii is influenced by metal-OH contacts in polyoxometalates shifting the van der Waals gap to the longer side. In turn, longer Mo and V bonding radii treat as bonded the indirect contacts between metal atoms in MO_6 or, more generally, MX_6 coordination clusters ($M \in \{\text{Mo}, \text{V}\}$, X – any element). Clearly, these contacts are recognised as bonds incorrectly, which showcases a limitation of distance-based connectivity heuristic, as distance alone is not enough to determine connectivity in such cases. Nevertheless, *Jmol* 3D viewer [40] recognises the same indirect Mo–Mo and V–V contacts as bonds. In practical applications, such indirect contacts could possibly be filtered out by employing the Voronoi tessellation. On the other hand, the newly assigned Cu–O

Table 2 Results of an attempt to reproduce connectivities derived by a chemical perception pipeline [2] using our bonding radii set

Class	New bonds	Missing bonds	Common bonds
Mo–Mo	4289	0	1196
V–V	2854	0	6
Ag–Ag	2713	0	2199
Cu–Cu	2669	0	2114
Cu–O	2256	0	22517
...			
C–C	51	0	7534316
C–O	22	0	956486
C–S	0	0	168082
C–P	0	0	201895
C–N	0	2	1463763
...			
N–Pr	0	389	58
O–Pr	3	494	937
Ce–O	0	605	2174
Ce–N	20	655	706
C–Ce	0	1045	54

Only 5 classes with the largest number of differences are shown for each kind of difference. In addition, five classes involving carbon and other organogenic elements are provided for reference. Column “New bonds” lists the number of newly assigned bonds by our radii, while “Missing bonds” lists unreproduced bonds. Column “Common bonds” lists the number of reproduced bonds

bonds mostly represent proper coordination interactions between Cu and axial water or ketone ligands that were not detected by the chemical perception pipeline.

Most of the unreproduced bonds can be attributed to the significantly shorter bonding radii for Ce and Pr. This is caused by a small component present in both Ce–O and O–Pr classes, which primarily consist of interactions involving CeO_6 and PrN_2O_4 . While such radii effectively identify smaller coordination spheres, they fail to capture connectivity in complexes with the coordination number of 7 or higher.

Outside the d-block the largest increases in radii lengths with regards to Pyykkö and Atsumi are Kr, Tl and Pb. While increases in Kr and Tl radii do not affect the validation outcomes much, number of new O–Pb bonds exceed 1000 observations. Most of them appear to be proper coordination sphere contacts between Pb and oxygen in ketones or carboxyl groups that were not detected by the chemical perception pipeline.

MaterialsCoord benchmark

We have attempted validation by plugging in our radii set to *MaterialsCoord* benchmark [12] (Git commit d9158f3 [41]) which evaluates the performance of several bonding heuristics as implemented in the

pymatgen v2022.11.7 package [42] for the task of predicting coordination numbers in inorganic structures. *MaterialsCoord* has the `JmolNN` class which implements the criterion described in Eq. 1 with $\epsilon = 0.45$ Å. Due to the similarity of the `JmolNN` class to our method, we inherited `JmolNN` class to allow for easier plugging in of other radii sets, naming the derived class `CovalentRadiiTableNeighbours`. *MaterialsCoord* uses a set of 116 diverse experimentally determined structures, notably excluding organic compounds.

The comparison results are plotted in Fig. 7. Although *MaterialsCoord* does not encompass organic structures, we can see that the bonding radii set presented in this paper fares better than the earlier radii sets. Our radii set fared consistently better than *Jmol's* *autoBond* algorithm, the only radii sum based method described in [12]. It should be noted that *MaterialsCoord* does not perform diagonal bond removal from 4-membered rings, but that places our method on a par with the other compared methods.

Bonding distances

We have evaluated the usability of the determined bond peaks (described in section [Determination of the bond peak](#)) in element-pair-specific distance-based connectivity detection. To do so, we have repeated the comparison with the chemical perception pipeline (as described in section [Comparison with a chemical perception pipeline](#)) by replacing radii sum in Eq. 1 with bond peak distance and using $\epsilon = 0.45$ Å. For classes where bond peak distances were unavailable due to the lack of observations, the original radii sum approach was retained using radii from Table 1.

In over 365 000 input structures 109,178 mismatches were detected, a substantial increase from 80,487 which were observed using radii sum criterion in section [Comparison with a chemical perception pipeline](#). Classes with the most new and missing bonds are shown in Table 3, with classes involving carbon and other organogenic elements provided for reference.

Table 3 Results of an attempt to reproduce connectivities derived by a chemical perception pipeline [2] using derived bonding distances

Class	New bonds	Missing bonds	Common bonds
C–Zn	11365	0	2189
C–K	5665	0	1511
Cu–Cu	3679	0	2114
Ag–Ag	2731	0	2199
C–Na	2730	0	315
...			
C–C	53	0	7534316
C–O	15	0	956486
C–S	0	0	168082
C–P	0	0	201895
C–N	0	0	1463765
...			
N–W	1	1187	1802
N–V	3	1239	1471
N–Ti	8	1699	3187
Mo–N	2	2014	2937
N–U	0	2643	1968

Only 5 classes with the largest number of differences are shown for each kind of difference. In addition, five classes involving carbon and other organogenic elements are provided for reference. Column "New bonds" lists the number of newly assigned bonds by our bonding distances, while "Missing bonds" lists unreproduced bonds. Column "Common bonds" lists the number of reproduced bonds

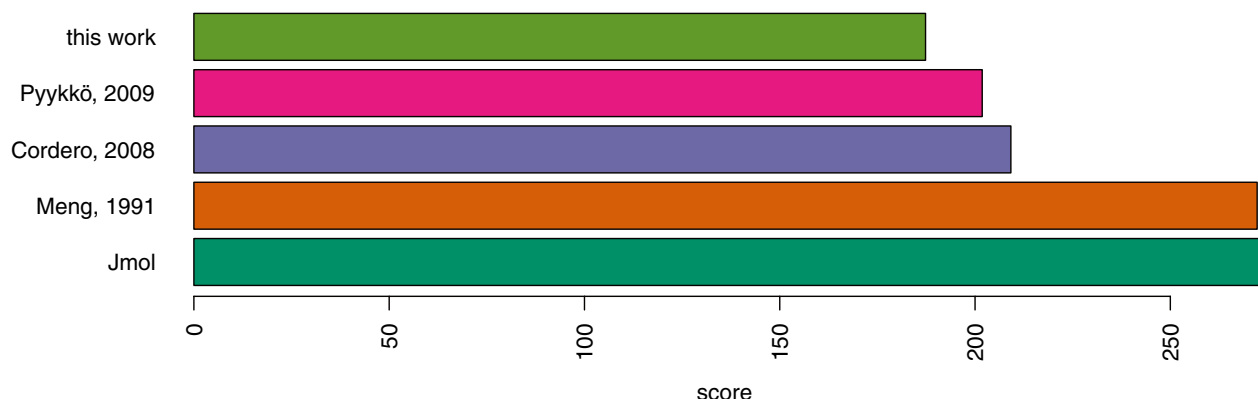


Fig. 7 Bonding radii sets validated using *MaterialsCoord* benchmark. Shown are the total scores (lower is better) for the whole set of structures presented in the benchmark. A lower *MaterialsCoord* score means a radii set yields coordination numbers closer to the expected values, with 0 being the ideal match score

Out of the classes with the most of newly detected bonds, only the new bonds in C–K class are proper bonds not detected by the chemical perception pipeline. Newly detected C–K bonds occur mostly in half-sandwich compounds between potassium atoms bound to benzene rings, thus longer C–K bonding distance might be useful in detecting such compounds. New bonds in other classes, C–Zn, Cu–Cu, Ag–Ag and C–Na, correspond to clearly nonbonded interactions between an atom and its second coordination sphere neighbours. For the class of C–Zn, a single component model is selected as the best, which results in the bond peak being placed at 2.49 Å.

The most populous classes of unreproduced bonds contain nitrogens in coordination interactions. In these cases the selected bonding distances are too short and miss proper coordination bonds.

Thus while the bonding distance-based connectivity criterion allows much more control over which interactions should be regarded as bonds, it is more difficult to derive in an unsupervised manner. Lack of data, problems associated with outliers or model selection strongly affect the derived bonding distances, whereas bonding radii are much less sensitive to problematic classes due to them being averaged out by the less problematic ones. Therefore, with the current methodology, we recommend using bonding radii criterion over the one based on the bonding distances. Nevertheless we provide both the bonding radii and distance sets [43].

Radii set browser

To better assess and compare the different algorithms and the derived sets, we have developed an interactive website [44]. The website facilitates browsing of calculated radii sets, comparing between different sets and reviewing the details of radii derivation. Radii are presented in the form of a periodic table, and a drop-down menu presents the option of switching between our radii sets and those published by other authors [7, 13, 14].

Clicking on a chemical element in the table opens a list of diagrams showing all the classes where the selected element participates in. Each diagram shows a class histogram, fitted Gaussian mixture model, detected van der Waals gap and the sum of radii of the participating elements. We have intensively used this website during the development of our methods.

Connectivity validator

We have created a website [45] exposing our connectivity validator which we have described in section [Comparison with a chemical perception pipeline](#). This website accepts uploads in structural data file (SDF) and chemical markup language (CML) formats, and highlights the

differences between connectivity described in the input file and the one determined using our radii set.

Discussion

Biases

The derived radii set is greatly skewed towards organic environments due to the abundance of organic and metalorganic compounds in the COD. Weights of organogenic elements dominate the other elements in approximate least squares matrix solution, some of the rarer elements have contacts with organogenic elements only. Class sizes are shown in Fig. 5.

Classes are not free from outlier observations, warranting stricter structure selection criteria, which could result in exclusion of valid observations. In particular, it has been observed that relaxation-based criterion does not identify structures having insignificant amounts of atoms with significant displacements from their equilibrium positions. Moving such atoms to their equilibrium positions does not affect the unit cell, therefore no changes in crystal density are observed. A more sensitive approach would be to identify atoms with high potential energies.

Applicability

Our method of approximating the distribution of interatomic distances with a Gaussian mixture and then selecting an antimode as a limit between bonded and non-bonded interactions has several limitations. All of them are related to the inability of our method to generalise over the variety of chemical interactions.

Many interaction classes do not contain pronounced gaps in their densities. In some cases even antimodes are not present at all (as in unimodal mixtures, mixtures with components having very low proportions, or substantially overlapping components), or are not pronounced. Cu–O class has a pair of antimodes, but both of them are still populated. Interestingly, first of the pronounced antimodes lands in between components of equatorial and axial interactions in complexes affected by the Jahn-Teller effect, both of which we would like to detect as bonds. Cu–Cu distance distribution is highly skewed towards longer distances and does not contain a gap; even the detected antimodes are not evident. A special case of this problem is the unimodal distribution where it is impossible to say whether the mode contains bonded or non-bonded interactions. Currently we treat this mode as containing bonded interactions, but this is clearly incorrect for element pairs which do not form bonds or of which bonded interactions are not observed in the COD.

Our method is sensitive to the absence and overrepresentation of certain chemical environments. For example, F–F class (Fig. 3) is dominated by nonbonded F–F interactions, therefore identification of the van der Waals gap

there is unreliable While this is a correct representation of F–F interactions, it needs *ad hoc* handling.

We have attempted to group classes by the shape of their distributions in an attempt to identify common traits and use them to apply different methods for separation of bonded and non-bonded interactions. We have used both two-sample Kolmogorov–Smirnov test and our own method of mixture model difference integral:

$$\Delta(A, B) = \int_0^6 |p_A(x) - p_B(x)| dx$$

Where A and B are classes and $p_X(x)$ is a mixture model density function for class X . Both methods resulted in similar distance matrices, however, neither of them highlighted sensible clusters. Most likely neither of the methods is sensitive enough to detect the similarities evident to the human eye without being penalised for differences, usually in location and scale. We have attempted aligning the distributions and models using normalisation and matching of the components surrounding the perceived van der Waals gap, but none of them yielded better outcomes.

Usage

Radii for some of the elements in our set are missing due to the lack of observations in the COD (inert gases, most of the elements with $Z > 83$). Some sources suggest replacing missing radii from the other radii sets or using fixed reasonable values [46]. We recommend taking the missing radii values from Pyykkö and Atsumi set.

Some of the other elements have unusually short or long radii, as evident from Fig. 1. Again, replacing them with values from other radii sets is a possibility. Another solution would be to average over all radii, or fitting a smoothing spline to each of the periods. We suggest a conservative approach of averaging over the outlying radii values, however care should be taken to maintain the periodic trends.

Alternative connectivity heuristics

While the described radii sum heuristic is very widely used in practice, there are many ongoing efforts to improve it or present alternatives to them, summarised and evaluated in, for example, Pan et al. [12], which measures the performance of various heuristics in the task of predicting coordination numbers in crystal structures. Some of the methods already incorporate Voronoi tessellation [42, 47, 48].

Conclusions

In this publication we present a method for unsupervised derivation of sets of bonding radii from crystallographic small molecule datasets. The devised method allows for fully automated bonding radii derivation, making the

method suitable for periodic analysis of large crystallographic resources. This property is especially useful for deriving bonding radii from crystallographic databases, such as the COD.

In addition, we present a bonding radii set derived from the COD. The presented set follows the trends also visible in other widely used bonding radii sets. We have evaluated our bonding radii set by applying it on structures with known connectivities to find the apriori and derived connectivities to be in a good agreement.

It is clear that the bonding radii sum based methods will not supersede more precise methods which pay more respect to chemical environments. However, they have their place in the computational chemist's toolbox as an indispensable approach for the derivation of connectivity when only the coordinates of atoms are known.

Acknowledgements

The authors thank Justinas Šlepavičius for allowing to use the results of his machine learning-based validation of the COD, and Kliment Olechnovič for his valuable insights into the subject matter.

Author contributions

E. Š., A. M. and A. V. have conceived and performed the research with the help of the rest of the authors. All authors contributed to the preparation of the manuscript. All authors read and approved the final manuscript.

Funding

This research has received funding from Vilnius University Science Support Fund under grant agreement No. MSF-JM-15/2021 and the Research Council of Lithuania under grant agreement No. MIP-23-87.

Data availability

The software and the data used to derive and validate the bonding radii sets described in this article are available in a Subversion repository at svn://www.crystallography.net/contacts-in-COD. This article is based on revision 2854 of the repository. All the software (scripts, libraries and workflow definitions) developed exclusively for this study are licensed under the BSD-3-Clause license, whereas all the data exclusive to this project are licensed under the CC0-1.0 license. The repository contains embedded external software and data which are subject to their own copyright details. To make the licensing details more explicit, we have adopted REUSE v3.3 recommendations by introducing REUSE.toml files throughout the repository tree.

Declarations

Competing interests

The authors declare no competing interests.

Received: 6 January 2026 Accepted: 24 May 2026

Published online: 10 June 2026

References

1. Merkys A, Vaitkus A, Grybauskas A, Kononov A, Quirós M, Gražulis S (2023) Graph isomorphism-based algorithm for cross-checking chemical and crystallographic descriptions. *J Cheminform* 15(1):25. <https://doi.org/10.1186/s13321-023-00692-1>
2. Vaitkus A, Merkys A, Sander T, Quirós M, Thiessen PA, Bolton EE, Gražulis S (2023) A workflow for deriving chemical entities from crystallographic data and its application to the Crystallography Open Database. *J Cheminform* 15(1):123. <https://doi.org/10.1186/s13321-023-00780-2>

3. Merkys A (2018) Extraction and Usage of Crystallographic Knowledge for Refinement and Validation of Molecular Models. PhD thesis, Vilnius University. https://www.lvb.lt/permalink/370LABT_NETWORK/1m6ui06/alma9910036312108451
4. Mounet N, Gibertini M, Schwaller P, Campi D, Merkys A, Marrazzo A, Sohler T, Castelli IE, Cepellotti A, Pizzi G, Marzari N (2018) Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds. *Nat Nanotechnol* 13:246–252. <https://doi.org/10.1038/s41565-017-0035-5>
5. Koritsanszky TS, Coppens P (2001) Chemical applications of X-ray charge-density analysis. *Chem Rev* 101(6):1583–1628. <https://doi.org/10.1021/cr990112c>
6. Bruno IJ, Shields GP, Taylor R (2011) Deducing chemical structure from crystallographically determined atomic coordinates. *Acta Crystallogr B* 67(4):333–349. <https://doi.org/10.1107/S0108768111024608>
7. Meng EC, Lewis RA (1991) Determination of molecular topology and atomic hybridization states from heavy atom coordinates. *J Comput Chem* 12(7):891–898. <https://doi.org/10.1002/jcc.540120716>
8. Baber JC, Hodgkin EE (1992) Automatic assignment of chemical connectivity to organic molecules in the Cambridge Structural Database. *J Chem Inf Model* 32(5):401–406. <https://doi.org/10.1021/ci00009a001>
9. Hendlich M, Rippmann F, Barnickel G (1997) BAL: automatic assignment of bond and atom types for protein ligands in the Brookhaven Protein Databank. *J Chem Inf Comput Sci* 37(4):774–778. <https://doi.org/10.1021/ci9603487>
10. Labute P (2005) On the perception of molecules from 3D atomic coordinates. *J Chem Inf Model* 45(2):215–221. <https://doi.org/10.1021/ci049915d>
11. Zhao Y, Cheng T, Wang R (2007) Automatic perception of organic molecules based on essential structural information. *J Chem Inf Model* 47(4):1379–1385. <https://doi.org/10.1021/ci700028w>
12. Pan H, Ganose AM, Horton M, Aykol M, Persson KA, Zimmermann NER, Jain A (2021) Benchmarking coordination number prediction algorithms on inorganic crystal structures. *Inorg Chem* 60(3):1590–1603. <https://doi.org/10.1021/acs.inorgchem.0c02996>
13. Cordero B, Gómez V, Platero-Prats AE, Revés M, Echeverría J, Cremades E, Barragán F, Alvarez S (2008) Covalent radii revisited. *Dalton Trans* 21:2832–2838. <https://doi.org/10.1039/B801115J>
14. Pykkö P, Atsumi M (2009) Molecular double-bond covalent radii for elements Li–E112. *Chemistry* 15(46):12770–12779. <https://doi.org/10.1002/chem.200901472>
15. Lang PF, Smith BC (2014) Electronegativity effects and single covalent bond lengths of molecules in the gas phase. *Dalton Trans* 43(21):8016–8025. <https://doi.org/10.1039/c4dt00807c>
16. Zhang Y (2018) Covalent radii and new applications. *Int J Cheminform Chem Eng* 7(1):42–51. <https://doi.org/10.4018/ijcce.2018010103>
17. Groom CR, Bruno IJ, Lightfoot MP, Ward SC (2016) The Cambridge Structural Database. *Acta Crystallographica Section B Struct Sci Crystal Eng Mater* 72:171–179. <https://doi.org/10.1107/S2052520616003954>
18. Nikolaenko TY, Chuiko VS, Bulavin LA (2020) The covalent radii derived from the first-principle data. *Mol Phys* 118(21–22):1742937. <https://doi.org/10.1080/00268976.2020.1742937>
19. Nikolaenko TY, Bulavin LA (2021) The machine-learned radii of atoms. *Comput Theor Chem* 1204:113389. <https://doi.org/10.1016/j.comptc.2021.113389>
20. Šidlauskaitė E (2023) Determination of atomic radii from small-molecule crystal structures. https://www.lvb.lt/permalink/370LABT_NETWORK/1m6ui06/alma9912936686808451
21. Alvarez S (2013) A cartography of the van der Waals territories. *Dalton Trans* 42(24):8617–8636. <https://doi.org/10.1039/C3DT50599E>
22. Gražulis S, Daškevič A, Merkys A, Chateigner D, Lutterottand L, Quirós M, Serebryanaya NR, Moeck P, Downs RT, Le Bail A (2012) Crystallography open database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. *Nucleic Acids Res* 40:420–427. <https://doi.org/10.1093/nar/gkr900>
23. Hall SR, Allen FH, Brown ID (1991) The crystallographic information file (CIF): a new standard archive file for crystallography. *Acta Crystallogr A* 47(6):655–685. <https://doi.org/10.1107/S010876739101067X>
24. Vaitkus A, Merkys A, Gražulis S (2026) cod-tools. <https://wiki.crystallography.net/cod-tools>
25. El-Kaderi HM, Hunt JR, Mendoza-Cortés JL, Côté AP, Taylor RE, O’Keeffe M, Yaghi OM (2007) Designed synthesis of 3D covalent organic frameworks. *Science* 316(5822):268–272. <https://doi.org/10.1126/science.1139915>
26. Arblaster JW (2014) Osmium density challenge. *Anal Bioanal Chem* 406(17):4009–4011. <https://doi.org/10.1007/s00216-014-7839-z>
27. Chen C, Ong SP (2022) A universal graph deep learning interatomic potential for the periodic table. *Nat Comput Sci* 2(11):718–728. <https://doi.org/10.1038/s43588-022-00349-3>
28. IUCr: Data validation criteria, version 2000.06.09. Accessed: 2026-04-23 (2000). https://journals.iucr.org/services/cif/checking/RFACG_01.html
29. IUCr: Data validation criteria, version 2000.06.09. Accessed: 2026-04-23 (2000). https://journals.iucr.org/services/cif/checking/RFACR_01.html
30. Gražulis S, Merkys A, Vaitkus A, Okulič-Kazarinas M (2015) Computing stoichiometric molecular composition from crystal structures. *J Appl Crystallogr* 48(1):85–91. <https://doi.org/10.1107/S1600576714025904>
31. Olechnovič K, Venclovas Č (2014) Voronota: a fast and reliable tool for computing the vertices of the Voronoi diagram of atomic balls. *J Comput Chem* 35:672–681. <https://doi.org/10.1002/jcc.23538>
32. Merkys A. crystal-contacts. [svn://saulius-grazulis.lt/crystal-contacts](https://github.com/saulius-grazulis/crystal-contacts)
33. Olechnovič K, Kulberkytė E, Venclovas Č (2012) CAD-score: a new contact area difference-based function for evaluation of protein structural models. *Proteins: Struct, Funct, Bioinf* 81(1):149–162. <https://doi.org/10.1002/prot.24172>
34. Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* 6:461–464. <https://doi.org/10.1214/aos/1176344136>
35. Merkys A (2026) MixtureFitting: R package for fitting mixture distributions to data using various approaches. <https://github.com/merkys/MixtureFitting>
36. Nelder JA, Mead R (1965) A simplex method for function minimization. *Comput J* 7:308–313. <https://doi.org/10.1093/comjnl/7.4.308>
37. You K (2021) Rlinsolve: Iterative Solvers for (Sparse) Linear System of Equations. R package version 0.3.2. <https://CRAN.R-project.org/package=Rlinsolve>
38. Wolke R, Schwetlick H (1988) Iteratively reweighted least squares: algorithms, convergence analysis, and numerical comparisons. *SIAM J Sci Stat Comput* 9(5):907–921. <https://doi.org/10.1137/0909062>
39. Vaitkus A. Cif-perceive-chemistry, Version 0.4.0. <https://doi.org/10.5281/zenodo.7510339>
40. Hanson RM (2010) Jmol – a paradigm shift in crystallographic visualization. *J Appl Crystallogr* 43(5):1250–1260. <https://doi.org/10.1107/S0021889810030256>
41. Ganose AM, Aykol M, Dagdelen J, Pan H, Zimmermann NER (2022) MaterialsCoord: infrastructure to benchmark crystal structure coordination algorithms. <https://github.com/hackingmaterials/materials-coord>
42. Ong SP, Richards WD, Jain A, Hautier G, Kocher M, Cholia S, Gunter D, Chevrier VL, Persson KA, Ceder G (2013) Python Materials Genomics (pymatgen): a robust, open-source python library for materials analysis. *Comput Mater Sci* 68:314–319. <https://doi.org/10.1016/j.commatsci.2012.10.028>
43. Merkys A. Sets of Derived Bonding Radii and Distances. <https://radii.crystallography.net/radii-sets/this-study/>
44. Šidlauskaitė E. Bonding Radii Set. <https://radii.crystallography.net/brows er>
45. Merkys A. Connectivity Validator. <https://radii.crystallography.net/checker>
46. The Cambridge Crystallographic Data Centre: Element Data and Radii <https://web.archive.org/web/20130115112446/http://www.ccdc.cam.ac.uk/products/csd/radii/>
47. O’Keeffe M (1979) A proposed rigorous definition of coordination number. *Acta Crystallogr A* 35(5):772–775. <https://doi.org/10.1107/S0567739479001765>
48. Zimmermann NER, Horton MK, Jain A, Haranczyk M (2017) Assessing local structure motifs using order parameters for motif recognition, interstitial identification, and diffusion path characterization. *Front Mater* 4:34. <https://doi.org/10.3389/fmats.2017.00034>

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.