

Theoretical Evaluation of Radii of Atoms in Molecules and their Dependence on Atomic Partial Charge

Amin Alibakhshi¹, Lars V. Schäfer^{1,*}

¹Center for Theoretical Chemistry, Ruhr University Bochum, 44780 Bochum, Germany

*Email: amin.alibakhshi@ruhr-uni-bochum.de , lars.schaefer@ruhr-uni-bochum.de

Abstract

Quantification of radii of atoms in molecules is of fundamental importance in understanding a wide range of physical and chemical phenomena. In the present study, we develop methods for evaluation of radii of atoms in molecules and assess their robustness by studying the agreement of van der Waals and solvent excluded surfaces constructed by them with reference iso-density surfaces. By studying a large dataset of 1235 molecules, we show that estimation of atomic radii via effective and free atomic volumes can accurately take the dependence of atomic radii on the chemical environment into account. A linear dependence of atomic radii on partial charge is found for radii estimated via effective volumes and partial charges computed based on iterative Hirshfeld and MBIS partitioning methods.

Introduction

Theoretical evaluation of atomic radii and their dependence on atomic partial charges is a fundamentally important subject that bridges various disciplines in physics and chemistry. Atomic radii play a dominant role in shaping many properties of compounds, including but not limited to molecular bonding, reactivity, and stability [1-4]. Crucially important quantities strongly influenced by atomic size are the inter- and intra-atomic interactions. They range from the classical example of Lennard-Jones interactions suggesting van der Waals (vdW) attraction and Pauli repulsion interactions to be proportional to 6th and 12th power of atomic radii [5], respectively, to a more recent theoretical demonstration of a linear relationship between atomic polarizability and the seventh power of atomic radii [6]. Similarly, the successful advancements in the theoretical treatment of solvation effects in molecular systems via continuum solvation models strongly rely on defining a solute cavity occupied by the solute molecule constructed via tailored atomic radii [7-10].

Inception of atomic radii dates back to the early 20th century and Bragg's pioneering work, who proposed that atomic distances within crystals can be interpreted as sums of individual atomic radii [11]. Although this groundbreaking approach, which was later further culminated by others [12-15], allows for the precise measurement of interatomic spacings and provides enhanced understanding of crystal structures, it is limited by the availability of suitable crystal structures, the assumption of static atomic radii that are unaffected by the chemical environment, and underestimation of atomic radii.

For quantifying radii of atoms in molecules, several studies have shown the necessity of refining atomic radii in a way to account for the variability due to diverse chemical environments [16-18]. This is, for example, reflected in the state-of-the-art continuum solvation models which exploit heuristic modifications on atomic radii, e.g. via scaling Bondi radii [19], empirical fine-tuning based on attaching functional groups in the conductor-like polarizable continuum model (CPCM) [20], or atomic partial charge in solvation model density (SMD) solvation models [21]. Thus, a method is needed that dynamically adapts to different molecular environments. Intuitively, one

can argue that radii of atoms in molecules are most strongly influenced by atomic partial charges, regarding that positively charged ions are smaller due to the stronger attraction of the electron cloud by a less shielded nucleus, and vice-versa for the negatively charged ions. The dependence of atomic radii on atomic partial charges has been intensely researched. The relationship between atomic partial charges and atomic radii fosters detailed studies of molecular interactions, such as hydrogen bonding, ionic and non-covalent interactions. However, the precise dependence between atomic radii and partial charges has not yet been demonstrated, to the best of our knowledge. One main reason behind this might be linked to the challenges in both quantifying radii of atoms in molecules and the ambiguities in defining atomic partial charges. As atomic partial charges are not experimental observables and no quantum mechanical operator exists for them, their definition and estimation introduce some degree of arbitrariness, as is for example reflected in the existence of many population analysis methods and a broad range of partial charges estimated by them.

The main aim of the present study is to tackle these challenges. Accordingly, we investigate the theoretical evaluation of radii of atoms in molecules based on two different approaches well-grounded in quantum chemical methodologies. The first approach relies on a theory initially conceived by Bader and co-workers, which defines atomic and molecular surfaces as iso-density surfaces beyond which the electron density is lower than a certain threshold (cut-off) [22]. Although this approach is mainly applicable to evaluate the total surface area of molecules, it can also be utilized to get a rough estimate of atomic radii via defining it as the closest distance of the nucleus to the iso-density surface [23]. One main challenge of this approach has long been the ambiguity in the precise value of the cut-off density. Different suggestions such as a cut-off density of 0.002 a.u. initially suggested by Bader [22] or 0.001 a.u. suggested by Boyd [24] yield different results [25]. In a recent study, we tackled this issue by employing an experimental evaluation of molecular surface areas based on thermodynamic phase change data [25]. By studying a dataset of 104 small molecules, we showed that a cut-off density of 0.0016 a.u. yields a close to perfect agreement (mean unsigned error of 1.6% and correlation coefficient of 0.995) between the experimental estimations and iso-density surfaces obtained from high-level quantum chemical calculations. We also showed that our suggested cut-off density yields atomic radii for noble gasses in excellent agreement with widely accepted experimental estimations. As a result, the iso-density surfaces constructed based on the cut-off density of 0.0016 a.u. is suggested as a reliable

representation of atomic and molecular surfaces. Accordingly, we consider this suggested cut-off density also in the present study to construct iso-density surfaces and refer to them as the reference molecular surfaces.

The second approach follows a relationship proposed by Tkatchenko and Scheffler with the extensively applied Tkatchenko-Scheffler method for estimating dispersion interactions [26]. This method employs the ratio of the effective volume occupied by an atom in a molecule to the volume of the same atom in the free state to evaluate the radius of the atom in the molecule (R_A) via

$$R_A = R_{A,free} \left(\frac{V_A^{eff}}{V_A^{free}} \right)^{1/3}, \quad (1)$$

$$\frac{V_A^{eff}}{V_A^{free}} = \frac{\int r^3 w_A(r) n(r) d^3(r)}{\int r^3 n_A^{free}(r) d^3(r)}, \quad (2)$$

where w_A is the atomic weight for partitioning the molecular space into atomic sub-spaces, and $R_{A,free}$ and n_A^{free} are the radius and the electron density of atom A in the free (isolated) state.

Estimation of atomic radii based on this approach inherently takes into account variations due to atomic partial charge via the partitioning weights in the effective volume estimations. Also, compared to the heuristic estimation of atomic radii based on the closest distance to the iso-density surface (see above), this method provides a more straightforward approach for direct estimation of spherically averaged atomic radii. Nevertheless, also for this method there are a number of challenges due to the ambiguity inherent to selecting partitioning methods for the calculation of effective volumes, as well as the appropriate values for free atomic radii. For partitioning the molecular space into atomic sub-spaces, Tkatchenko and Scheffler considered Hirshfeld partitioning in their work [26]. However, by selecting other methods such as iterative Hirshfeld (I-Hirshfeld) [27] and minimal basis iterative Stockholder (MBIS) [28] partitioning approaches, diverse estimates of effective volumes can be obtained. Similarly, for radii of free atoms, diverse parameterizations exist in the literature [8].

Regardless of the method used for approximating the radii of atoms in molecules, construction of molecular surfaces from the estimated radii also introduces some challenges. For that, the most

extensively applied surface types are vdW and solvent excluded surfaces (SES), which both consider atoms in perfect spherical shape. The SES surfaces are constructed in the same way as vdW surfaces with the only difference that the intersection area between adjacent atoms is smoothed via examining if a probe solvent atom can be packed in those areas [29]. This introduces the arbitrariness of selecting the probe radius used for this purpose, as the other ambiguity to the problem.

In the present study, employing our recently developed method that provides a thermodynamically consistent estimation of molecular surfaces [25], we aim to shed light on the above-mentioned issues. Developing methods for accurate estimation of radii of atom in molecules and their relation to atomic partial charges are the other main aims of the present study. For that, our general approach follows estimating radii of atoms in molecules based on different recipes and employ them to construct vdW and SES molecular surfaces. The robustness of the methods employed for evaluating atomic radii and molecular surfaces is then investigated via studying the agreement of the constructed molecular surfaces with electron iso-density surfaces contoured at 0.0016 a.u., which are considered in the present study as reference molecular surfaces, as discussed above.

Computational details

We benchmark our methods against a dataset of 1235 small molecules containing H, C, N, O, and S elements, covering the most essential elements in organic chemistry. A full list of the studied molecules is provided in the supplemental materials. For each compound, we first optimized the geometry at the PBE/def2TZVP density functional theory (DFT) level. The optimized structures were then used for higher level electronic structure computations with the DSD-PBEP86/def2QZVPD double-hybrid DFT method, which was shown in our recent study to accurately reproduce molecular iso-density surfaces computed at the CCSD(T) coupled-cluster level [25].

Radii of atoms in molecules were evaluated based on two different scenarios. As the most straightforward approach, we estimated atomic radii as the closest distance of each nucleus from the iso-density surface contoured at 0.0016 a.u. In the second approach, we estimated radii of atoms in molecules via effective volumes based on equation (1). To estimate effective volumes,

we considered Hirshfeld, I-Hirshfeld, and MBIS methods to partition the molecular space into atomic sub-spaces. The radii of free atoms required in equation (1) were optimized such that the constructed surfaces yield best agreement with reference molecular surfaces. For that, we considered separate optimizations of radii of free atoms for vdW and SES molecular surfaces. For calculating total vdW surface areas, we devised a simple but efficient algorithm that works based on uniformly distributing a pre-defined number of points over the surface of spheres centered at atomic positions and with a radius corresponding to the atomic radius being optimized. For each spherical shell, we then removed the uniformly distributed points on that shell which were found inside the spheres belonging to neighboring atoms. The vdW surface of each atom ($S_{vdW,A}$) was then approximated via

$$S_{vdW,A} = 4\pi r_A^2 \frac{n_{non-overlap}}{n_{tot}}, \quad (3)$$

where n_{tot} is the total number of initially distributed points over each spherical shell, and $n_{non-overlap}$ is the number of those points that were not inside other atomic shells. To estimate vdW surfaces, we considered 5000 points for each atom.

For estimating total SES surfaces, we employed the MolSurfComp MATLAB code developed by Quan and Stamm [30]. For each one of the vdW or SES surface types, optimization of atomic radii was carried out using an in-house MATLAB code based on multi-start optimization. For each molecule, to have a broad overview of the dependence of atomic radii on different partial charge models, we considered atomic partial charges based on ChelpG [31], CM5 [32], Hirshfeld [33], I-Hirshfeld [27], Merz-Singh-Kollman (MK)[34], RESP [35], EEM [36], ADCH [37], Gasteiger [38], and MBIS [28] methods, computed at DSD-PBEP86/def2QZVPD level of theory. All electronic structure computations were carried out with Orca 5.0.3. [39] and post-processing analysis of wavefunction files to calculate effective volumes and partial atomic charges was carried out using Multiwfn software [40].

The agreement between the theoretically estimated molecular surfaces in comparison to the experimental values is reported in terms of mean unsigned percentage error (MUPE) and Pearson correlation coefficient (R),

$$\text{MUPE} = \frac{1}{N} \sum \left(\left| \frac{y_{i,1} - y_{i,2}}{y_{i,1}} \right| \right) \times 100 \quad (4)$$

$$R = \frac{\sum(y_{i,1} - \bar{y}_1)(y_{i,2} - \bar{y}_2)}{\sqrt{\sum(y_{i,1} - \bar{y}_1)^2 \sum(y_{i,2} - \bar{y}_2)^2}} \quad (5)$$

where $y_{i,1}$ and $y_{i,2}$ are iso-density and TE surfaces of molecule i , respectively, and \bar{y}_1 and \bar{y}_2 are the mean values.

Results and Discussion

Using atomic radii estimated via distance of atomic positions to iso-density surfaces, the constructed vdW surfaces yielded MUPE and correlation coefficients of 4.26% and 0.84, respectively, compared to the reference molecular surfaces. For SES surfaces, we considered solvent probe radii ranging from 0.10 Å to 1.50 Å with 0.05 Å interval. SES surfaces calculated using a probe solvent radius of 0.40 Å showed the best agreement with reference molecular surfaces (Figure 1), with MUPE and Pearson correlation coefficient of 1.31% and 0.9993, respectively. For each molecule, the details of calculated SES and vdW surfaces based on the described recipe is provided in the supplemental materials.

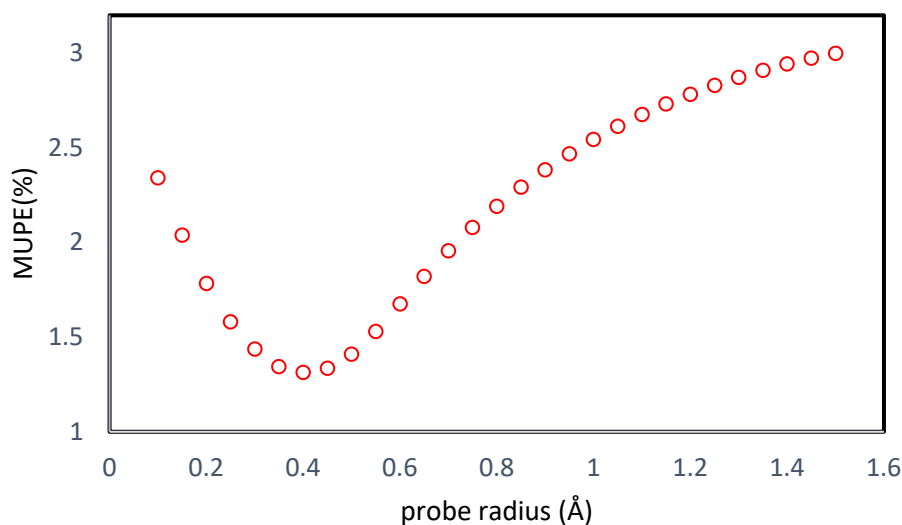


Figure 1- Agreement of SES surfaces with reference surfaces for different solvent probe radii

One potential application of this analysis might be to overcome the ambiguity of selecting surface type and solvent probe radii, two key challenges in continuum solvation models. For example, while in continuum solvation models implemented in the Gaussian software package, SES surfaces have been the default in version 03, vdW surfaces are considered as default in later releases of the same software. On the contrary, vdW surfaces were considered as default in continuum solvation models implemented in the Orca software up to version 4.0, but were then replaced with SES surfaces in more recent versions. In both cases, a solvent probe radius of 1.2 Å is considered as default for constructing SES surfaces. Here, our inference about surface type and probe radius is grounded in the direct comparison to iso-density surfaces that are consistent with thermodynamic phase change data. Accordingly, we suggest them as a reliable way to construct molecular surfaces from atomic radii and consider the probe radius of 0.4 Å for constructing SES surfaces within parameterizing radii of free atoms via effective volumes as well. The better agreement of SES surfaces with iso-density surfaces is consistent with the smooth variation of the electron density between atoms in molecules, as schematically illustrated in Figure 2.

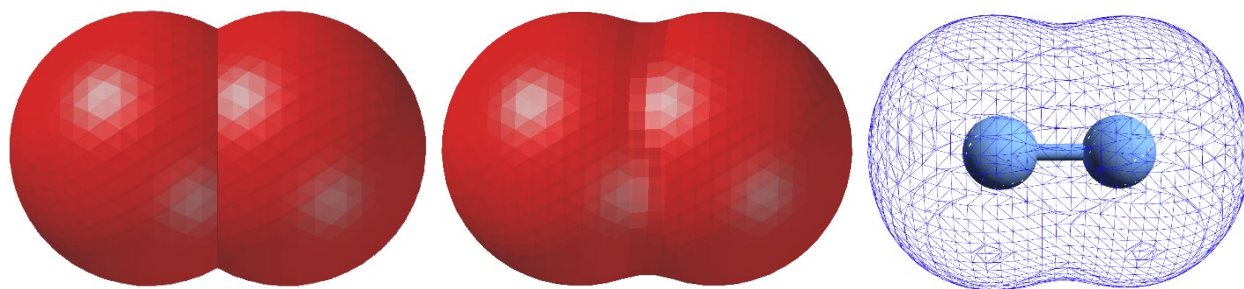


Figure 2- Comparison of vdW (left), SES (middle), and iso-density surfaces (right) in the F₂ molecule illustrates the agreement of SES surfaces with reference iso-density surfaces. The iso-density surface image (right) was generated with the Avogadro software using electron density data. The vdW and SES images (left and middle) were generated with the MolSurfComp Matlab code.

In comparison to radii estimated via closest distance to iso-density surfaces, we found the surfaces generated via effective and free atomic volumes to be in much better agreement with the reference molecular surfaces. According to the results reported in Table 1, the best agreement was obtained via atomic radii estimated based on Hirshfeld partitioning of atomic spaces and SES surface types, resulting in MUPE and correlation coefficient of 0.55% and 0.9997, respectively. Nevertheless, for other partitioning methods where radii of free atoms are optimized against SES surfaces, an almost similar accuracy and excellent agreement with reference molecular surfaces is found. A

comparison of the molecular surfaces approximated via effective and free volumes with reference molecular surfaces is shown in Figure 3.

Table 1- Accuracy of estimated molecular surfaces calculated via effective and free atomic volumes in reproducing iso-density surfaces and optimized values of radii of free atoms

	Estimated radius of free atom (Å)					MUPE (%)	R
	H	C	N	O	S		
Hirshfeld (SES)	1.42	1.94	1.70	1.62	2.06	0.55	0.9997
I-Hirshfeld (SES)	1.47	1.87	1.65	1.49	2.06	0.75	0.9996
MBIS (SES)	1.75	1.86	1.68	1.58	2.04	0.56	0.9997
Hirshfeld (vdW)	1.35	2.00	1.71	1.61	2.07	1.37	0.9969
I-Hirshfeld (vdW)	1.37	1.93	1.65	1.52	2.08	1.75	0.9982
MBIS (vdW)	1.53	1.99	1.68	1.58	2.05	1.96	0.9954
Homonuclear diatomic molecule	1.34		1.69	1.60	2.02		
COSMO-RS	1.30	2.00	1.83	1.72			

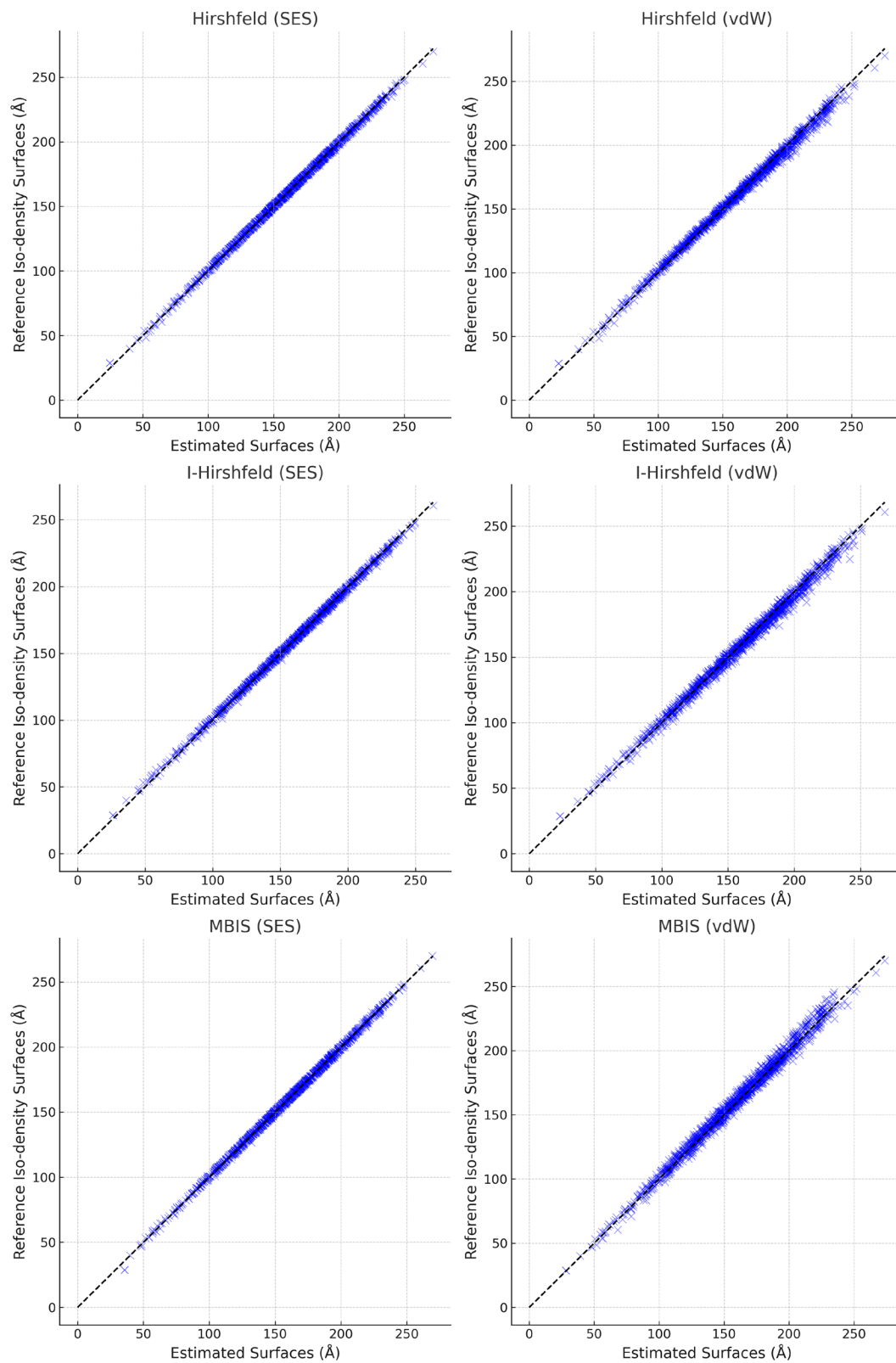


Figure 3- Comparison of estimated surfaces based on different partitioning methods with reference iso-density surfaces

The results show that the SES surfaces constructed with optimized radii of free atoms are in remarkably better agreement with the reference molecular surfaces compared to vdW surfaces, as expected. Nevertheless, the optimized radii of free atoms show a slight dependence to the applied partitioning method. We aimed to investigate this further and to shed light on the robustness of the applied partitioning schemes. By considering the parameterized radii of free atoms to be equivalent to the radii of atoms in molecules for zero atomic partial charge, one can exploit their agreement with the radii found by optimization using a homonuclear diatomic molecule to evaluate the performance of the different partitioning schemes. For homonuclear diatomic molecules, the condition of zero partial charge perfectly holds and undesired effects due to bond polarity or anisotropy of electron density can be neglected. Additionally, atomic radii for homonuclear diatomic molecules are obtained directly and without relying on any partitioning method and via fine tuning the atomic radii in a way that the resulting SES surfaces are in the best agreement with reference iso-density surfaces. Consequently, such a comparison allows a direct and straightforward investigation of the robustness of the partitioning methods applied. Accordingly, we recalculated the atomic radii of H, N, O, and S atoms via total iso-density surfaces of H₂, N₂, O₂, and S₂ molecules, respectively, where bond lengths were taken from the NIST database [41] and best matching radii were found by line-search. We also made a comparison with optimized radii employed in the COSMO-RS model, which is commonly considered as one of the most accurate implicit solvation models [42]. According to the results reported in table 2, atomic radii optimized using the whole dataset via Hirshfeld partitioning and vdW surface type, shows the best agreement with estimation of respective radii both via homonuclear diatomic molecules and COSMO-RS parameterization. For parameterization of radii of free atoms when SES surfaces are considered, Hirshfeld partitioning again yields the best agreement with Homonuclear diatomic molecule and COSMO-RS parameterizations. Accordingly, the Hirshfeld partitioning method performs well in estimating effective volumes.

Table 2- Comparison of mean average deviation (Å) of optimized free atomic radii with the estimation based on COSMO-RS parameterization and diatomic homonuclear molecules

	Homonuclear diatomic molecule	COSMO-RS
Hirshfeld (SES)	0.037	0.102
I-Hirshfeld (SES)	0.080	0.177
MBIS (SES)	0.115	0.220

Hirshfeld (vdW)	0.022	0.070
I-Hirshfeld (vdW)	0.052	0.130
MBIS (vdW)	0.062	0.132

Estimating the radii of atoms in molecules also allowed investigating their dependence to atomic partial charges for a broad range of population analysis methods. To do so, for each element we calculated the correlation coefficient of partial charges and radii estimated with different recipes. For the case of estimated radii via effective and free volumes, we here only consider with radii of free atoms optimized against vdW surfaces due to their slightly higher accuracy as explained earlier. According to the results reported in Table 3, there is a remarkable linear correlation between atomic radii and partial charges computed based on I-Hirshfeld and MBIS partitioning schemes. The reason is that in both these methods, atomic weights and thus effective volumes, are derived according to partial charges obtained in each iteration of partitioning. For I-Hirshfeld methods, the dependence to partial charge is implemented via linear scaling with respect to upper and lower charge state [27], while MBIS method takes it into account via direct inclusion of shell populations[28]. According to these results and considering the high accuracy of molecular surfaces estimated via these two methods, we suggest them for applications where dependence of radii to atomic partial charges are required. One interesting application might be development of next generation force fields, where well-established methods are available for estimating environment dependent partial charges. This allows defining charge dependent atomic radii which can then be transformed to charge dependent atomic polarizability or vdW parameters, as discussed in the introduction.

Table 3- Correlation coefficient between atomic radii and partial charges

Radii estimation method	Charge model	H	C	N	O	S
Distance to iso-density	ADCH	-0.1044	-0.0191	-0.1027	0.0541	-0.0044
Distance to iso-density	ChelpG	-0.8009	0.0048	-0.2358	0.0312	-0.0469
Distance to iso-density	CM5	-0.6598	-0.1811	-0.2922	0.1020	-0.0061
Distance to iso-density	EEM	-0.7857	-0.1655	-0.3736	-0.4542	-0.0448
Distance to iso-density	Gasteiger	-0.7863	-0.2927	-0.5679	-0.3090	-0.1813
Distance to iso-density	Hirshfeld	-0.5085	-0.2262	-0.3053	0.3291	0.0001
Distance to iso-density	I-Hirshfeld	-0.7752	-0.0978	-0.2441	0.3904	0.1187
Distance to iso-density	MBIS	-0.7990	-0.1551	-0.2594	0.2942	-0.0020
Distance to iso-density	MK	-0.7643	0.0470	-0.1666	0.0892	-0.0439
Distance to iso-density	RESP	-0.7724	0.0322	-0.1688	0.0990	-0.0460
Eq.(1)-Hirshfeld	ADCH	-0.0260	0.0481	0.0151	-0.0887	-0.6656

Eq.(1)-Hirshfeld	ChelpG	-0.0252	0.3103	0.1843	-0.0681	-0.6705
Eq.(1)-Hirshfeld	CM5	-0.2485	0.5128	0.3688	-0.1671	-0.7580
Eq.(1)-Hirshfeld	EEM	-0.0583	0.5855	0.3901	0.3193	-0.6798
Eq.(1)-Hirshfeld	Gasteiger	0.0595	0.2747	0.3197	0.1876	-0.1733
Eq.(1)-Hirshfeld	Hirshfeld	-0.3610	0.2910	-0.2217	-0.7344	-0.8024
Eq.(1)-Hirshfeld	I-Hirshfeld	-0.0962	0.5154	0.2770	-0.1454	-0.5906
Eq.(1)-Hirshfeld	MBIS	-0.0239	0.4966	0.3031	-0.1158	-0.6766
Eq.(1)-Hirshfeld	MK	-0.0182	0.3557	0.1770	-0.1069	-0.6737
Eq.(1)-Hirshfeld	RESP	-0.0170	0.3394	0.1731	-0.1120	-0.6742
Eq.(1)-I-Hirshfeld	ADCH	-0.0973	-0.1105	-0.3952	-0.1581	-0.8726
Eq.(1)-I-Hirshfeld	ChelpG	-0.7961	-0.8717	-0.9032	-0.6064	-0.9737
Eq.(1)-I-Hirshfeld	CM5	-0.6880	-0.8646	-0.8298	-0.5215	-0.9637
Eq.(1)-I-Hirshfeld	EEM	-0.8030	-0.8607	-0.8875	-0.3813	-0.9398
Eq.(1)-I-Hirshfeld	Gasteiger	-0.6775	-0.6997	-0.7948	-0.1733	-0.3915
Eq.(1)-I-Hirshfeld	Hirshfeld	-0.5392	-0.8209	-0.7588	-0.4020	-0.9540
Eq.(1)-I-Hirshfeld	I-Hirshfeld	-0.9634	-0.9898	-0.9885	-0.9132	-0.9981
Eq.(1)-I-Hirshfeld	MBIS	-0.9187	-0.9774	-0.9425	-0.8576	-0.9871
Eq.(1)-I-Hirshfeld	MK	-0.7580	-0.8668	-0.8936	-0.5889	-0.9719
Eq.(1)-I-Hirshfeld	RESP	-0.7544	-0.8645	-0.8939	-0.5904	-0.9716
Eq.(1)-MBIS	ADCH	-0.0633	-0.0249	-0.4890	-0.2337	-0.8860
Eq.(1)-MBIS	ChelpG	-0.8796	-0.8845	-0.9405	-0.8214	-0.9880
Eq.(1)-MBIS	CM5	-0.4133	-0.2043	-0.9259	-0.7581	-0.9585
Eq.(1)-MBIS	EEM	-0.4798	-0.1919	-0.9072	-0.4822	-0.9470
Eq.(1)-MBIS	Gasteiger	-0.4467	-0.1545	-0.7932	-0.1373	-0.4414
Eq.(1)-MBIS	Hirshfeld	-0.3092	-0.1854	-0.7773	-0.4799	-0.9460
Eq.(1)-MBIS	I-Hirshfeld	-0.9526	-0.9745	-0.9585	-0.8913	-0.9117
Eq.(1)-MBIS	MBIS	-0.9837	-0.9905	-0.9781	-0.9393	-0.9844
Eq.(1)-MBIS	MK	-0.8375	-0.8718	-0.9325	-0.8134	-0.9857
Eq.(1)-MBIS	RESP	-0.8367	-0.8720	-0.9352	-0.8184	-0.9858

For a schematic illustration, dependence of radii of different elements estimated via effective and free volumes based on I-Hirshfeld partitioning method to atomic partial charges calculated based on different population analysis methods is depicted in figure 4.

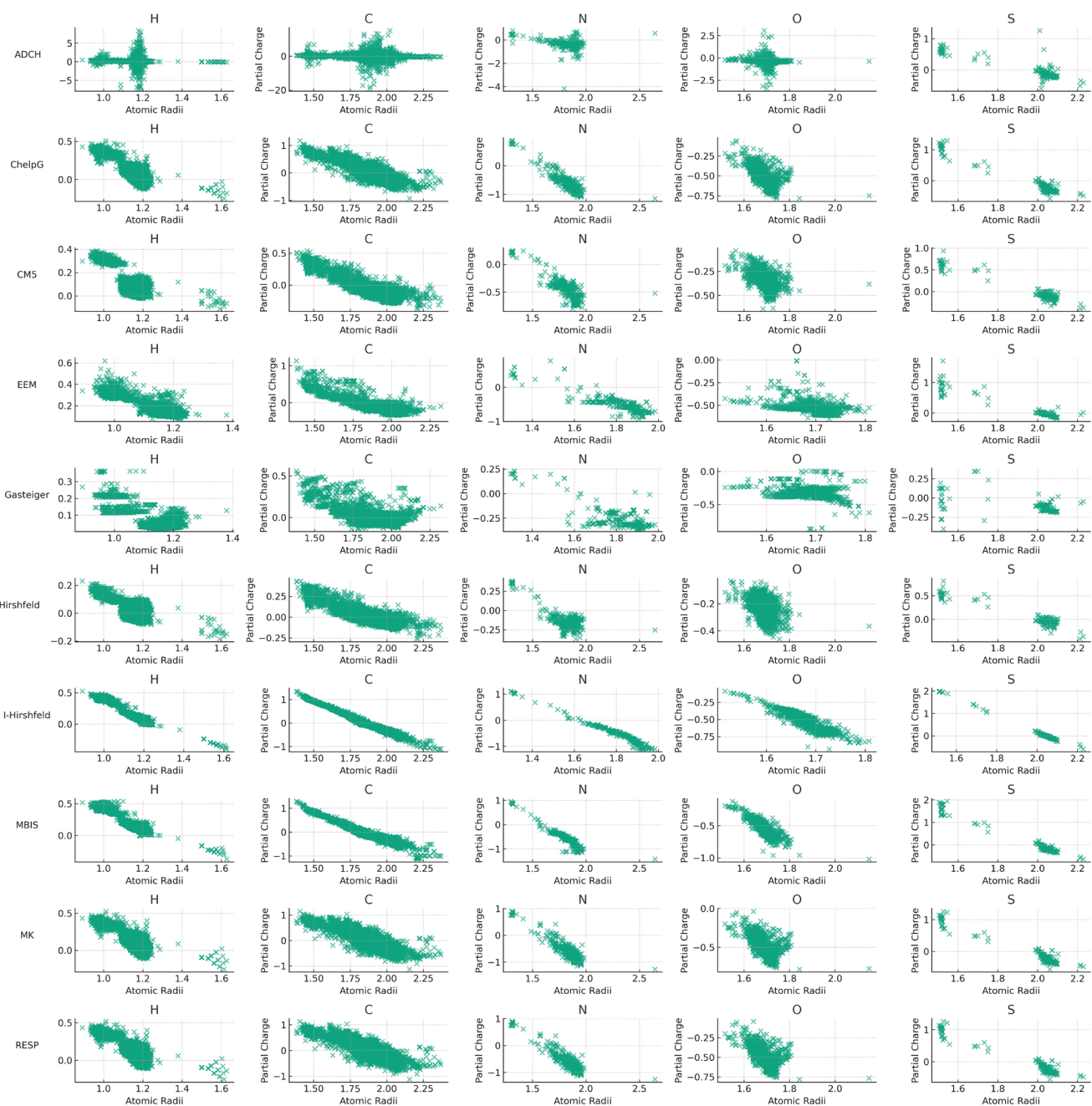


Figure 4- Schematic illustration of the relationship between atomic radii and partial charges for selected elements (H, C, N, O, S) for various charge models

Conclusion

In the present study, we show that reference molecular iso-density surfaces of a large dataset of 1235 molecules can be accurately reproduced by SES surfaces obtained with a 0.4 Å probe radius, where atomic radii are estimated via effective and free volumes. For that, our result showed a better performance for Hirshfeld partitioning scheme. This method also allows parameterizing radii of

free atoms which we found to be in good agreement with the values parameterized by the COSMO-RS solvation model. To study the dependence of atomic radii on atomic partial charges, we illustrated a drastic dependence to the selected charge model and also to the details of estimating radii of atoms in molecules. Our results showed a clear linear dependence between atomic radii and partial charges when they are calculated based on MBIS or iterative Hirshfeld partitioning schemes, while other partitioning or population analysis methods failed to capture this dependence. These results are expected to open new horizons towards better understanding molecular interactions and applications in the development of next generation force fields and continuum solvation models.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) under Germany's Excellence Strategy - EXC 2033 - 390677874 - RESOLV. We gratefully acknowledge the funding of this project by computing time provided by the Paderborn Center for Parallel Computing (PC2).

References:

1. Smith, M., et al., *Consequences of Overfitting the van der Waals Radii of Ions*. Journal of Chemical Theory and Computation, 2023. **19**(7): p. 2064-2074.
2. Spronk, S.A., et al., *A quantum chemical interaction energy dataset for accurately modeling protein-ligand interactions*. Scientific Data, 2023. **10**(1): p. 619.
3. Lu, T. and F. Chen, *Quantitative analysis of molecular surface based on improved Marching Tetrahedra algorithm*. Journal of Molecular Graphics and Modelling, 2012. **38**: p. 314-323.
4. Murray, J.S. and P. Politzer, *The electrostatic potential: an overview*. Wiley Interdisciplinary Reviews: Computational Molecular Science, 2011. **1**(2): p. 153-163.
5. Leach, A.R., *Molecular modelling: principles and applications*. 2001: Pearson education.
6. Fedorov, D.V., et al., *Quantum-mechanical relation between atomic dipole polarizability and the van der Waals radius*. Physical Review Letters, 2018. **121**(18): p. 183401.
7. Alibakhshi, A. and B. Hartke, *Implicitly perturbed Hamiltonian as a class of versatile and general-purpose molecular representations for machine learning*. Nature Communications, 2022. **13**(1): p. 1245.
8. Alibakhshi, A., *Thermodynamically effective molecular surfaces for more efficient study of condensed-phase thermodynamics*. 2021.
9. Alibakhshi, A. and B. Hartke, *Improved prediction of solvation free energies by machine-learning polarizable continuum solvation model*. Nature Communications, 2021. **12**(1): p. 1-7.
10. Ye, R., et al., *Variable van der waals radii derived from a hybrid gaussian charge distribution model for continuum-solvent electrostatic calculations*. Zeitschrift für Physikalische Chemie, 2016. **230**(5-7): p. 681-701.

11. Bragg, W.L., *XVIII. The arrangement of atoms in crystals*. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1920. **40**(236): p. 169-189.
12. Pauling, L., *The nature of the chemical bond: and the structure of molecules and crystals; an introduction to modern structural chemistry*. 1940.
13. Kitaigorodskii, A., *Organic Crystal Chemistry*. Izd. Akad. Nauk SSSR, Moscow, 1955: p. 15.
14. Bondi, A.v., *van der Waals volumes and radii*. The Journal of physical chemistry, 1964. **68**(3): p. 441-451.
15. Bondi, A., *Van der Waals volumes and radii of metals in covalent compounds*. The Journal of Physical Chemistry, 1966. **70**(9): p. 3006-3007.
16. del Valle, F.O. and M. Aguilar, *Solute-solvent interactions: Part 5. An extended polarizable continuum model including electrostatic and dispersion terms and electronic correlation in the solute*. Journal of Molecular Structure: THEOCHEM, 1993. **280**(1): p. 25-47.
17. del Valle, F.O. and M. Aguilar, *Correlation and solvation effects. IV. A systematic analysis of the influence of cavity size and shape on solvation properties in the polarizable continuum model with electron correlation*. Journal of Computational Chemistry, 1992. **13**(2): p. 115-134.
18. Miertuš, S., J. Bartoš, and M. Trebaticka, *Dependance of atomic radii and volumes on the electron distribution in solute molecule and on solute-solvent interaction*. Journal of molecular liquids, 1987. **33**(2-3): p. 139-156.
19. Garcia-Ratés, M. and F. Neese, *Effect of the Solute Cavity on the Solvation Energy and its Derivatives within the Framework of the Gaussian Charge Scheme*. Journal of Computational Chemistry, 2020. **41**(9): p. 922-939.
20. Barone, V. and M. Cossi, *Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model*. The Journal of Physical Chemistry A, 1998. **102**(11): p. 1995-2001.
21. Marenich, A.V., C.J. Cramer, and D.G. Truhlar, *Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions*. The Journal of Physical Chemistry B, 2009. **113**(18): p. 6378-6396.
22. Bader, R., W.H. Henneker, and P.E. Cade, *Molecular charge distributions and chemical binding*. The Journal of Chemical Physics, 1967. **46**(9): p. 3341-3363.
23. Rahm, M., R. Hoffmann, and N. Ashcroft, *Atomic and ionic radii of elements 1–96*. Chemistry–A European Journal, 2016. **22**(41): p. 14625-14632.
24. Boyd, R., *Relative sizes of atoms.[Density contour, scaling]*. J. Plasma Phys.:(United Kingdom), 1977. **10**(12).
25. Alibakhshi, A. and L. V. Schäfer, *Reconciling Experiment with Quantum Chemical Calculations: Electron Iso-Density Surfaces Represent Atomic and Molecular Surfaces*. 2024.
26. Tkatchenko, A. and M. Scheffler, *Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data*. Physical review letters, 2009. **102**(7): p. 073005.
27. Bultinck, P., et al., *Critical analysis and extension of the Hirshfeld atoms in molecules*. The Journal of chemical physics, 2007. **126**(14).
28. Verstraelen, T., et al., *Minimal basis iterative stockholder: atoms in molecules for force-field development*. Journal of Chemical Theory and Computation, 2016. **12**(8): p. 3894-3912.
29. Pascual-ahuir, J.-L., E. Silla, and I. Tunon, *GEPOL: An improved description of molecular surfaces. III. A new algorithm for the computation of a solvent-excluding surface*. Journal of Computational Chemistry, 1994. **15**(10): p. 1127-1138.
30. Quan, C. and B. Stamm, *Mathematical analysis and calculation of molecular surfaces*. Journal of Computational Physics, 2016. **322**: p. 760-782.

31. Breneman, C.M. and K.B. Wiberg, *Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis*. Journal of Computational Chemistry, 1990. **11**(3): p. 361-373.
32. Marenich, A.V., et al., *Charge model 5: An extension of Hirshfeld population analysis for the accurate description of molecular interactions in gaseous and condensed phases*. Journal of chemical theory and computation, 2012. **8**(2): p. 527-541.
33. Hirshfeld, F.L., *Bonded-atom fragments for describing molecular charge densities*. Theoretica chimica acta, 1977. **44**: p. 129-138.
34. Besler, B.H., K.M. Merz Jr, and P.A. Kollman, *Atomic charges derived from semiempirical methods*. Journal of computational chemistry, 1990. **11**(4): p. 431-439.
35. Bayly, C.I., et al., *A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model*. The Journal of Physical Chemistry, 1993. **97**(40): p. 10269-10280.
36. Mortier, W.J., K. Van Genechten, and J. Gasteiger, *Electronegativity equalization: application and parametrization*. Journal of the American Chemical Society, 1985. **107**(4): p. 829-835.
37. Lu, T. and F. Chen, *Atomic dipole moment corrected Hirshfeld population method*. Journal of Theoretical and Computational Chemistry, 2012. **11**(01): p. 163-183.
38. Gasteiger, J. and M. Marsili, *A new model for calculating atomic charges in molecules*. Tetrahedron letters, 1978. **19**(34): p. 3181-3184.
39. Neese, F., *The ORCA program system*. Wiley Interdisciplinary Reviews: Computational Molecular Science, 2012. **2**(1): p. 73-78.
40. Lu, T. and F. Chen, *Multiwfn: A multifunctional wavefunction analyzer*. Journal of computational chemistry, 2012. **33**(5): p. 580-592.
41. Database, N.I.C.S., *NIST Standard Reference Database Number 3, National Institute of Standards and Technology, Gaithersburg MD, 20899*.
42. Klamt, A., et al., *Refinement and parametrization of COSMO-RS*. The Journal of Physical Chemistry A, 1998. **102**(26): p. 5074-5085.