

University of Groningen

## Variations in Proteins Dielectric Constants

Amin, Muhamed; Küpper, Jochen

*Published in:*  
 Chemistryopen

*DOI:*  
[10.1002/open.202000108](https://doi.org/10.1002/open.202000108)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
 Publisher's PDF, also known as Version of record

*Publication date:*  
 2020

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*  
 Amin, M., & Küpper, J. (2020). Variations in Proteins Dielectric Constants. *Chemistryopen*, 9(6), 691-694.  
<https://doi.org/10.1002/open.202000108>

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

# Variations in Proteins Dielectric Constants

Muhamed Amin<sup>\*,[a, b, c]</sup> and Jochen Küpper<sup>\*,[a, d, e]</sup>

Using a new semi-empirical method for calculating molecular polarizabilities and the Clausius–Mossotti relation, we calculated the static dielectric constants of dry proteins for all structures in the protein data bank (PDB). The mean dielectric constant of more than 150,000 proteins is  $\epsilon_r = 3.23$  with a standard deviation of 0.04, which agrees well with previous measurement for dry proteins. The small standard deviation results from the strong correlation between the molecular polarizability and the volume of the proteins. We note that non-amino acid cofactors such as Chlorophyll may alter the dielectric environment significantly. Furthermore, our model shows anisotropies of the dielectric constant within the same molecule according to the constituents amino acids and cofactors. Finally, by changing the amino acid protonation states, we show that a change of pH does not have a significant effect on the dielectric constants of proteins.

The intermolecular electrostatic interactions in proteins are scaled by their dielectric constants, which vary according to the size and composition of the proteins. The accurate determination of the dielectric constant is essential to understand a variety of biochemical interactions such as electron and proton transfer,<sup>[1,2]</sup> voltage gating,<sup>[3,4]</sup> ion channel selectivity,<sup>[5]</sup> charge separation,<sup>[6]</sup> and protein-protein and protein-ligand interactions.<sup>[7]</sup> To a large extent, these interactions are governed by the electrostatic-potential surfaces of proteins.

Direct measurements of dielectric constants  $\epsilon_r$  of dry proteins span a range from 2.5 to 3.5. These values are determined by measuring the capacity of crystalline samples,<sup>[8,9]</sup> which agree with chemical shift perturbation measurements.<sup>[10]</sup> However, in addition to amino acids, proteins in practice contain solvent molecules as well as organic and inorganic cofactors. These affect their dielectric constants and in most cases the effective dielectric constant is significantly different from the measured values for the dry proteins. The effective dielectric constants are usually determined indirectly using the Poisson-Boltzmann equation to calculate the electrostatic interactions that reproduce measured  $pK_a$ 's of some amino acids. These measurements include the effect of solvent molecules on the dielectric constant.<sup>[10]</sup> The contribution of the solvent to the effective dielectric constant was studied theoretically based on Kirkwood-Fröhlich dielectric theory.<sup>[11]</sup>

In addition, computational studies based on continuum electrostatics and molecular dynamic simulations showed that different structural motifs within the same protein may yield significantly different values of  $\epsilon_r$  according to the polarity of their constituents molecules.<sup>[12–14]</sup>

The dielectric constant  $\epsilon_r$ , the average polarizability  $\alpha$ , and the volume  $V$  of a molecule are related by the Clausius–Mossotti relation:<sup>[15]</sup>

$$\frac{4\pi\alpha}{3V} = \frac{\epsilon_r - 1}{\epsilon_r + 2} \quad (1)$$

However, calculations of the molecular polarizabilities of macromolecules are challenging and computationally demanding. Previously, we proposed a model<sup>[16]</sup> for calculating the complete polarizability tensor of a protein through scaling of the tensor of a perfect conductor of the same shape based on a molecular basis set. The scaling factor was obtained from a regression model that correlated the polarizabilities of the molecule and a corresponding perfect-conductor of the constituents molecules of the proteins, i.e. the amino.

Here, we propose a new method for the calculation of the average (scalar) polarizabilities of proteins based on their amino acid compositions, which utilizes the fact that objects with the same volume  $V$  and dielectric constant  $\epsilon_r$  have the same average polarizabilities  $\alpha$  independent of shape, see also (1). The static dielectric constants are then calculated using the Clausius–Mossotti relation. This method is computationally highly efficient and facilitated the calculations of the average polarizabilities and dielectric constants of all proteins in the protein data bank (PDB).<sup>[17]</sup>

The average polarizability of a molecule can be calculated from the sum over hybridization configurations of the atoms in the molecule,<sup>[18]</sup>

[a] M. Amin, J. Küpper  
Center for Free-Electron Laser Science, Deutsches Elektronen-Synchrotron DESY, Notkestrasse 85, 22607 Hamburg, Germany  
E-mail: muhamed.amin@cfel.de  
jochen.kuepper@cfel.de

Homepage: <https://www.controlled-molecule-imaging.org>

[b] M. Amin  
Rijksuniversiteit Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Groningen, Netherlands

[c] M. Amin  
Department of Sciences, University College Groningen, University of Groningen, Groningen, Netherlands

[d] J. Küpper  
Department of Physics, Universität Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany

[e] J. Küpper  
Center for Ultrafast Imaging, Universität Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany

Supporting information for this article is available on the WWW under <https://doi.org/10.1002/open.202000108>

© 2020 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

$$\alpha = \frac{4}{N} \left( \sum_A \tau_A \right)^2 \quad (2)$$

with the number of electrons in the molecule  $N$  and the hybrid component  $\tau_A$  of each atom  $A$ , obtained by approximating the zeroth order wavefunction by an antisymmetrized product of molecular orbitals and spin functions. Average polarizabilities predicted by this method showed a very good agreement with experimental polarizabilities for more than 400 relatively small molecules with only  $\sim 2\%$  error.

Furthermore, since the atomic hybridizations of the atoms within the constituents amino acids do not change in proteins, (2) could be rearranged to obtain the average polarizability of a protein  $\alpha_p$  by summing over effective amino-acid hybrid components:

$$\alpha_p = \frac{4}{N_p} \left( \sum_{aa} \tau_{aa} \right)^2 \quad (3)$$

Here,  $N_p$  is the number of electrons in the protein and  $\tau_{aa}$  are the hybridization components of amino acid  $aa$ , which are obtained as

$$\tau_{aa} = \frac{\sqrt{N_{aa} \alpha_{aa}}}{2} \quad (4)$$

with the number of electrons  $N_{aa}$  in an amino acid  $aa$  and its average polarizability  $\alpha_{aa}$ . The latter could be obtained from quantum-chemical calculations and, therefore, the values of  $\tau$  not only include the summation of the atomic hybrid components within the amino acids, but also exchange correlation interactions at the level of quantum-chemistry employed.

Furthermore, for (2) to be applicable for very polar compounds,  $\tau_A$  has to be modified to include the effect of the atoms to which  $A$  is bonded. However,  $\tau_{aa}$  already includes this effect since it reproduces the exact polarizabilities calculated from first principles.

The values of  $\tau$  and  $\alpha_{aa}$  obtained with DFT are reported in Table 1 for the 6-31G+( $d,p$ ) and 6-311G++(3 $df,3pd$ ) basis set using B3LYP functional. The 6-31G+( $d,p$ ) basis sets allow us to compare the predicted average polarizability against the calculated ones for the Trp-cage mini protein, whereas the DFT calculations were not feasible for the larger basis sets. The average polarizability of the Trp-cage protein calculated by DFT is  $221 \text{ \AA}^3$ ; this calculation consumed more than 2000 CPU hours. The average polarizability of Trp cage calculated with our semi-empirical approach is  $215 \text{ \AA}^3$ , with an error against DFT of 2.7%; calculated in less than 200  $\mu\text{s}$ . Thus, this approach allows the calculations of the average polarizabilities and hence the dielectric constants of all the structures stored in the PDB. However, for these calculations we will use the amino acids polarizabilities obtained with the larger basis sets 6-311G++(3 $df,3pd$ ) to get more accurate predictions; for Trp cage this approach yields  $234 \text{ \AA}^3$ .

**Table 1.** Polarizabilities, volumes, and a Clausius-Mosotti term of the amino acids. The amino acids are sorted according to their molecular weight.  $\alpha'$ ,  $\alpha$  are the average polarizabilities calculated using the 6-31G+( $d,p$ ) and 6-311G++(3 $df,3pd$ ) basis sets, respectively. All polarizability values are reported in units of  $\text{\AA}^3 = \text{pm}^6$

	$\alpha'$ ( $\text{\AA}^3$ )	$\alpha$ ( $\text{\AA}^3$ )	$V$ ( $\text{\AA}^3$ )	$\frac{4\pi\alpha}{3V}$
G	6	6	63	0.41
A	7	8	81	0.42
S	8	9	92	0.39
P	10	11	109	0.41
V	11	12	119	0.41
T	10	11	109	0.41
C	10	11	98	0.47
I	13	14	141	0.41
L	11	13	139	0.40
N	10	11	112	0.41
D	11	12	102	0.49
Q	12	13	132	0.41
K	13	14	158	0.36
E	14	15	121	0.52
M	14	15	139	0.46
H	14	15	138	0.45
F	17	18	160	0.48
R	15	17	173	0.40
Y	18	19	168	0.48
W	22	23	193	0.50

To compare with our previous method, which allows the calculations of the full polarizability tensor, we calculated the polarizability tensor for perfect conductors of the same shape of the proteins by solving Laplace's equation with Dirichlet boundary conditions and using Monte Carlo path integral methods.<sup>[19]</sup> Then, all tensors are diagonalized to transform the proteins to the polarizability frame and the average of the diagonal elements are scaled by 0.26, which was the slope of the best-fit line that described the correlation between the amino acids and perfect conductors of their shapes.<sup>[16]</sup> The obtained polarizabilities from the summation of the square of the atomic hybridization components highly correlate with those obtained by scaling the polarizabilities of perfect conductors with  $R^2=0.8$  and a slope of 1.6, with the intercept set to zero. Thus, the later, method produced polarizabilities that are 60% higher, which we ascribe to effects of the uneven concentration of the individual amino acids in each protein. Overall, the method presented here provides a computationally highly efficient method for the calculation of the scalar polarizabilities. If the tensorial properties of the polarizability are needed, the current method could be used to generate the scaling factor that is applied to the tensor elements obtained in our previous method.<sup>[16]</sup>

In order to solve the Clausius-Mosotti equation, the volumes of the proteins are calculated as the summation of the volume of the constituents amino acids. The volume of the 20 amino acids are calculated using the Volume Assessor software by rolling a virtual sphere with a probe radius of 1 pm on the surface of the amino acids.<sup>[20]</sup> The calculated volumes are reported in Table 1.

The average static dielectric constant  $\epsilon$ , for more than 150,000 protein structures stored in the PDB database based on their amino acid decomposition is 3.23 with a standard

deviation of 0.04, see Figure 1a. According to the Clausius–Mossotti relation, the ratio between the average polarizability and the volume,  $\alpha/V$ , is the factor that determines the value of  $\epsilon_r$ . Thus, due to the strong correlation between the average polarizability and the molecular volume with  $R^2=1$ , Figure 1b, the standard deviation of  $\epsilon_r$  is very small. According to the regression model shown in Figure 1b, the polarizability  $\alpha$  of proteins could be calculated according to the straight line equation  $\alpha = 0.1 \cdot V + 32$  with negligible residuals. Both the volume and the average polarizabilities exhibit a skewed normal distribution, shown in Figure 1c, d.

The maximum dielectric constant of 3.7 is observed for N-terminal human brand 3 peptide with PDB ID 2BTA,<sup>[21]</sup> which has an average polarizability of  $212.7 \text{ \AA}^3$  and a volume of  $1879 \text{ \AA}^3$ . The large polarizability of this peptide is attributed to the ASP and GLU amino acids, which represent 50% of the constituent amino acids and have high  $\alpha/V$  ratios. The minimum  $\epsilon_r$  of 2.8 is observed for peptide-membrane PDB ID 6HNG,<sup>[22]</sup> which is formed by only eight leucine and six lysine amino acids. The lysine amino acid generally has a small  $\alpha/V$  ratio, because it is positively charged, i.e., it has less electrons than neutral or negatively charged amino acids which are also stronger bound.

Within the same protein the value of  $\epsilon_r$  may change according to the composition of the different parts. For example, in norrin, a Wnt signaling activator, PDB ID 5BPU,<sup>[23]</sup> the chains A, B, D, E, and F have  $\epsilon_r = 3.20$ , while chains H and I have  $\epsilon_r = 4.26$  as they are only formed by GLU amino acids. Thus,  $\epsilon_r$  distributions can be inhomogeneous within a protein, which agrees with previous studies based on MD simulations and continuum electrostatics simulations.<sup>[12–14]</sup> Furthermore, proteins have a variety of cofactor such as chlorophyll, metal clusters, chloride ions, hems, quinones, ... These molecules are very different than the amino acids and could have large impact on the dielectric environment of the proteins. For example, the calculated average polarizability of chlorophyll is  $132.3 \text{ \AA}^3$ , with a volume of  $900 \text{ \AA}^3$ , which results in  $\epsilon_r = 5.9$ , while for iron-

sulphur clusters of photosystem I in the oxidized state,<sup>[24]</sup> and its amino acids ligands  $\epsilon_r = 3.2$ .

To study the effect of pH on the dielectric constant, we recalculated the distribution of  $\epsilon_r$  for all proteins by replacing the average polarizabilities  $\alpha_{aa}$  of  $\text{GLU}^-$ ,  $\text{ASP}^-$ , and  $\text{HIS}^0$  with the average polarizabilities of the protonated form  $\text{GLU}^0$ ,  $\text{ASP}^0$ , and  $\text{HIS}^+$  to simulate low pH environment. The mean of the distribution reduced to 3.15 and the standard deviation is unchanged. Because the mean of the  $\epsilon_r$  is changed only by 0.08, it is a reasonable assumption that proteins, which experience pH gradient across different structural motifs have the same dielectric constants.

In conclusions, we developed an empirical method for the calculation of the average polarizabilities of dry proteins based on their amino acids composition. The method is computationally highly efficient and allowed us to calculate the average polarizabilities and dielectric constants of all molecular structures in the PDB. The average dielectric constant for more than 150,000 proteins is  $\epsilon_r = 3.23$ , with a very small standard deviation of 0.04, due to the strong correlation between the average polarizability and the molecular volume.

However, organic and inorganic cofactors could alter the dielectric environment of the proteins significantly. Thus, in order to understand the chemical reactions in proteins, the correct dielectric environment should be implemented in the biochemical/biophysical calculations.

We point out that the current approach does not take into account the molecules shape, which is valid for the scalar average polarizability, see also (1). For the computation of tensorial properties advanced, more expensive methods have to be employed.<sup>[16]</sup>

## Supporting Information

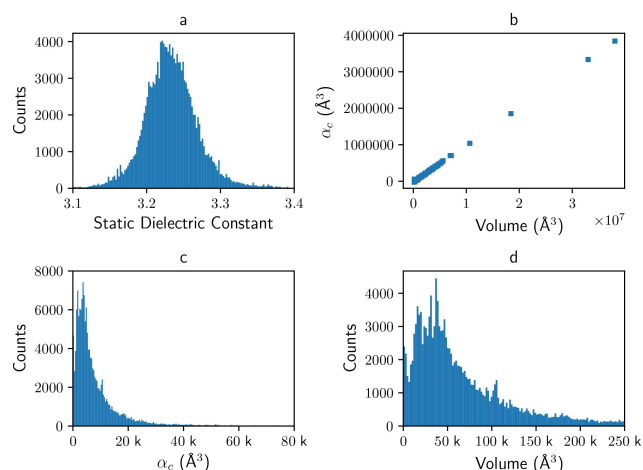
We provide a compressed text file in comma-separated-value format that contains the polarizabilities, the volumes, and the dielectric constants for all structures in PDB (as of 01. August 2019).

## Acknowledgement

This work has been supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) through the Consolidator Grant COMOTION (614507) and by the Deutsche Forschungsgemeinschaft through the Cluster of Excellence "Advanced Imaging of Matter" (AIM, EXC 2056, ID 390715994).

## Conflict of Interest

The authors declare no conflict of interest.



**Figure 1.** a) Histogram of the static dielectric constants  $\epsilon_r$  of all proteins in the PDB database. b) Correlation between the average polarizabilities and volumes of the proteins. c) Histogram of the average polarizabilities of the proteins. d) Histogram of the molecular volume of the proteins.

**Keywords:** proteins · dielectric constants · semi-empirical methods · electrostatic interactions · molecular polarizabilities

- [1] M. H. V. Huynh, T. J. Meyer, *Chem. Rev.* **2007**, *107*, 5004–5064.
- [2] M. Amin, L. Vogt, W. Szejgis, S. Vassiliev, G. W. Brudvig, D. Bruce, M. Gunner, *J. Phys. Chem. B* **2013**, *119*, 7366–7377.
- [3] C. M. Armstrong, B. Hille, *Neuron* **1998**, *20*, 371–380.
- [4] F. J. Sigworth, *Quart. Rev. Biophys.* **1994**, *27*, 1–40.
- [5] A. Keramidasa, A. J. Moorhousea, P. R. Schofield, P. H. Barry, *Prog. Biophys. Mol. Biol.* **2004**, *86*, 161–204.
- [6] H. B. Gray, J. R. Winkler, *Annu. Rev. Biochem.* **1996**, *65*, 537–561.
- [7] F. B. Sheinerman, R. Norel, B. Honig, *Curr. Opin. Struct. Biol.* **2000**, *10*, 153–159.
- [8] D. Rosen, *Trans. Faraday Soc.* **1963**, *59*, 2178–2191.
- [9] S. Takashima, H. P. Schwan, *J. Phys. Chem.* **1965**, *69*, 4176–4182.
- [10] P. Kukic, D. Farrell, L. P. McIntosh, B. García-Moreno, E. K. S. Jensen, Z. Toleikis, K. Teilum, J. E. Nielsen, *J. Am. Chem. Soc.* **2013**, *135*, 16968–16976.
- [11] M. K. Gilson, B. H. Honig, *Biopolymers* **1986**, *25*, 2097–2119.
- [12] L. Li, C. Li, Z. Zhang, E. Alexov, *J. Chem. Theory Comput.* **2013**, *9*, 2126–2136.
- [13] S. T. Hazra, A. U. Siwen, W. Emil, A. Zhao, *J. Math. Chem.* **2019**, 2282–2294.
- [14] T. Simonson, D. Perahia, *Pneumonol. Alergol. Pol.* **1995**, *92*, 1082–1086.
- [15] L. Jansen, *Phys. Rev.* **1958**, *112*, 434–444.
- [16] M. Amin, H. Samy, J. Küpper, *J. Phys. Chem. Lett.* **2019**, *10*, 2938–2943, arXiv:1904.02504 [physics].
- [17] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 10.1093/nar/28.1.235.
- [18] K. J. Miller, J. Savchik, *J. Am. Chem. Soc.* **1979**, *112*, 7206–7213.
- [19] M. L. Mansfield, J. F. Douglas, E. J. Garboczi, *Phys. Rev. A* **2001**, *64*, 061401–061416.
- [20] N. Voss, M. Gerstein, T. Steitz, P. Moore, *J. Mol. Biol.* **2006**, *360*, 893–906.
- [21] M. Schneider, C. Post, *Biochemistry* **1995**, *34*, 16574–16584.
- [22] G. Schneider, M. Blatter, A. Mueller, *Protein Data Bank website* **2018**.
- [23] T. Chang, F. Hsieh, M. Zebisch, K. Harlos, J. Elegheert, E. Jones, *Elife* **2015**, *4*, e06554.
- [24] J. H. Golbeck, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1992**, *43*, 293–324.

Manuscript received: April 19, 2020

Revised manuscript received: May 11, 2020