Electronic Supplementary Material (ESI) for Nanoscale.

Supporting Information: Hybrid quantum-classical polarizability model for single molecule biosensing

Ekaterina Zossimova,*,†,‡ Johannes Fiedler,¶ Frank Vollmer,‡ and Michael Walter†

†*Freiburg Center for Interactive Materials and Bioinspired Technologies (FIT), University of Freiburg, D-79110 Freiburg, Germany*

‡*Department of Physics and Astronomy, Living Systems Institute, University of Exeter, EX4 4QD, Exeter, UK*

¶*Department of Physics and Technology, University of Bergen, Allégaten 55, 5007 Bergen, Norway*

E-mail: ez216@exeter.ac.uk

Contents

S1 Tabulated polarizabilities

Table S1: PCM polarizability α^{PCM} of solvated amino acids, calculated using the finite-field method. The background medium is described by the PCM with the optical permittivity of water, $\varepsilon_{\mathrm{opt}} = 1.77$. Units of α^{PCM} and the cavity volume V are Å^3. The fractional anisotropy (FA) is dimensionless number between 0 (isotropic) and 1 (anisotropic). q specifies the net charge of the molecule ($q = 0$ if left blank).

	Molecule		q	$\hat{\alpha}^{\mathrm{PCM}}$	α_1^{PCM}	α_2^{PCM}	α_3^{PCM}	$\mathrm{FA}^{\mathrm{PCM}}$	V
A	Alanine	$C_3H_7NO_2$		10.2	11.9	10.9	7.9	0.20	132
R	Arginine	$C_6H_{14}N_4O_2$		22.1	28.2	21.4	16.8	0.25	243
		$C_6H_{15}N_4O_2^+$	+1	20.3	26.0	19.4	15.4	0.26	248
N	Asparagine	$C_4H_8N_2O_3$		14.3	17.4	15.3	10.3	0.25	177
D	Aspartic acid	$C_4H_6NO_4^-$	−1	15.2	17.9	16.0	11.6	0.21	171
		$C_4H_7NO_4$		13.2	14.9	13.2	11.7	0.12	173
C	Cysteine	$C_3H_7NO_2S$		14.2	17.3	14.1	11.1	0.22	162
E	Glutamic acid	$C_5H_8NO_4^-$	−1	17.1	20.7	17.5	13.2	0.21	193
		$C_5H_9NO_4$		15.4	19.6	15.7	10.9	0.27	194
Q	Glutamine	$C_5H_{10}N_2O_3$		16.5	20.4	16.8	12.2	0.24	199
G	Glycine	$C_2H_5NO_2$		8.3	10.4	8.5	5.9	0.27	109
N1-H	D-Histidine	$C_6H_9N_3O_2$		18.6	24.1	18.5	13.2	0.29	210
N3-H	E-Histidine	$C_6H_9N_3O_2$		18.4	20.8	19.1	15.4	0.15	210
H	Histidine	$C_6H_{10}N_3O_2^+$	+1	16.7	18.9	17.5	13.7	0.16	210
I	Isoleucine	$C_6H_{13}NO_2$		16.6	17.9	17.2	14.7	0.10	200
L	Leucine	$C_6H_{13}NO_2$		16.8	19.0	16.7	14.6	0.13	201
K	Lysine	$C_6H_{14}N_2O_2$		18.7	22.6	18.6	15.1	0.20	215
		$C_6H_{15}N_2O_2^+$	+1	17.4	21.3	17.4	13.4	0.22	222
M	Methionine	$C_5H_{11}NO_2S$		18.4	21.9	18.0	15.3	0.18	208
F	Phenylalanine	$C_9H_{11}NO_2$		22.5	29.7	22.4	15.2	0.31	238
P	Proline	$C_5H_9NO_2$		13.2	15.3	13.2	11.1	0.16	165
S	Serine	$C_3H_7NO_3$		11.1	12.4	11.7	9.1	0.16	144
T	Threonine	$C_4H_9NO_3$		13.2	14.9	13.9	10.9	0.16	167
W	Tryptophan	$C_{11}H_{12}N_2O_2$		28.4	36.7	29.9	18.4	0.32	281
Y	Tyrosine	$C_9H_{11}NO_3$		23.7	32.4	23.0	15.7	0.34	250
V	Valine	$C_5H_{11}NO_2$		14.4	16.0	14.9	12.4	0.13	177

S2 Finite field versus linear response TDDFT

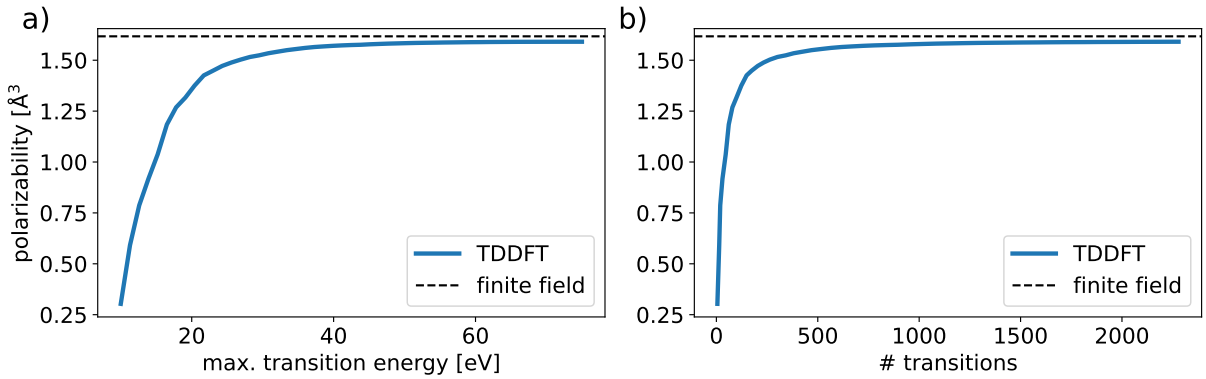

Figure S1: Convergence of TDDFT polarizability of H_2O with a) energy range of transitions and b) number of transitions.

The polarizability can be expanded using transition energies and dipole matrix-elements, where in principle all dipole transitions have to be summed up.[S1] These transitions can be calculated from time-dependent density functional theory (TDDFT). Figure S1 shows that a large number of transitions is needed to converge the TDDFT polarizability $\alpha(0)$ to the result of a calculation within the finite field approach. Since we do not require frequency-dependent polarizabilities, the finite field method is the more computationally efficient way to obtain the static dipole polarizability of a molecule without compromising the accuracy of the calculation.

S3 Excess versus gasphase polarizability

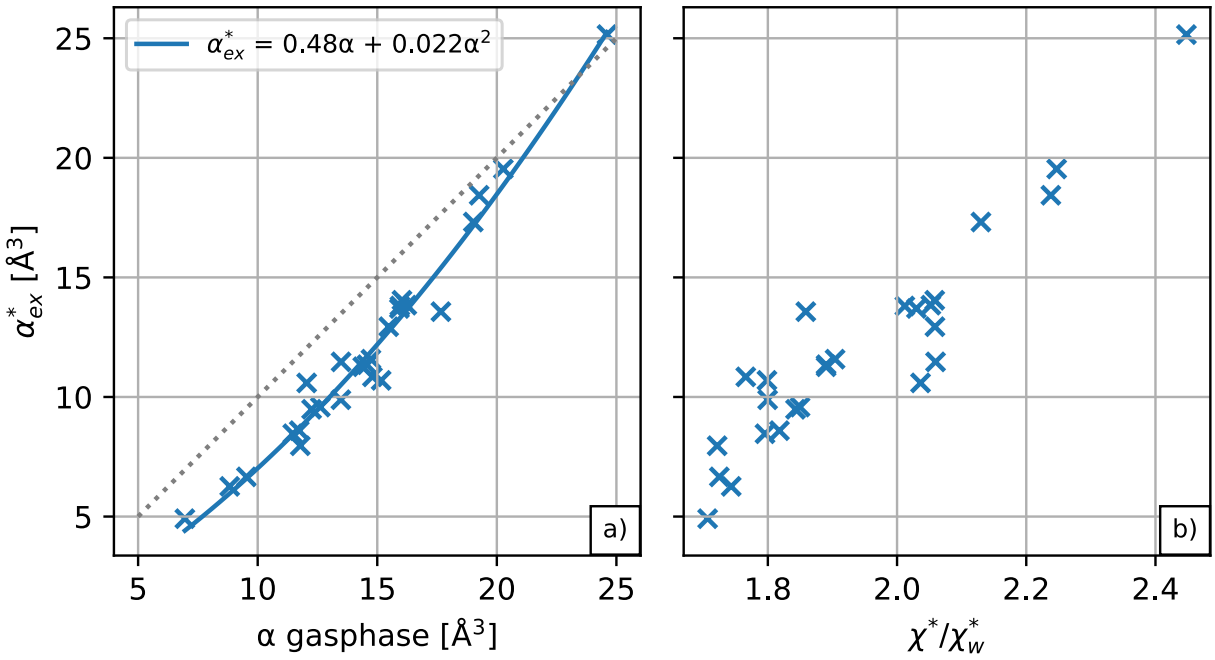

Figure S2: (a) Excess polarizability versus gasphase polarizability of amino acids. (b) Excess polarizability versus the polarizability density χ^* scaled by the polarizability density of water χ^*_w.

S4 Protonation states

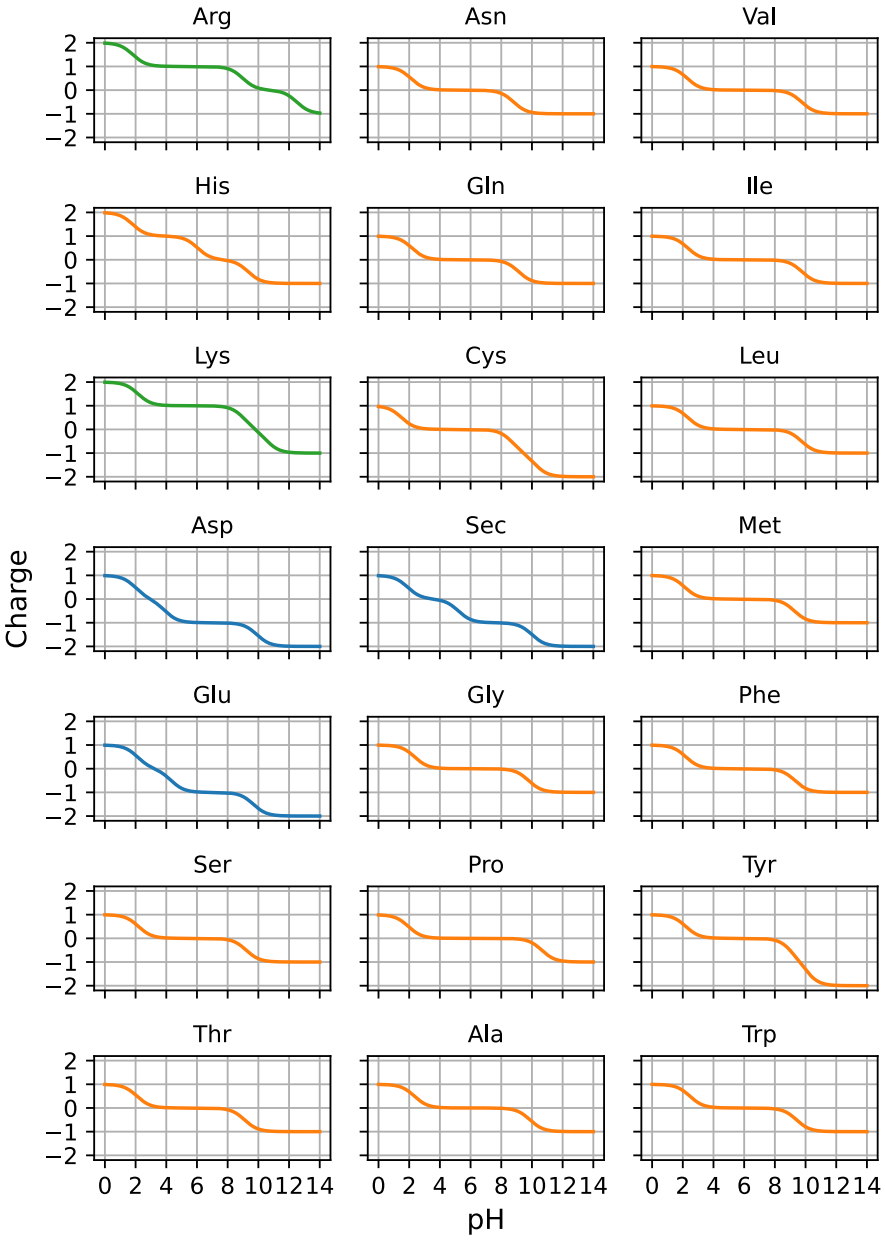

Figure S3: Protonation states of 21 amino acids as a function of pH, calculated using the Henderson-Hasselbalch equation and pK_a data from Ref.[S2] (see Chapter 5: Dissociation Constants of Organic Acids and Bases). The line colour corresponds to the protonation state of the molecule at physiological pH (cation: green, zwitterion: orange, anion: blue).

The dissociation constant of an ionizable group is known as its K_a value and it is related to the strength of the acid by $pK_a = -\log_{10} K_a$. Amino acids have at least two pK_a values, corresponding to the deprotonation of the carboxyl group and deprotonation of the amine group. For amino acids with an ionizable side chain, there is also a third pK_a value corresponding to the deprotonation of the side chain. The Henderson-Hasselbalch equation[S3–S5] can be used to find the ratio between conjugate acid-base pairs, e.g., the ratio between zwitterion and cation forms of an amino acid for a given pH,

$$\mathrm{pH} = pK_1 + \log_{10} \frac{[\sim \mathrm{COO^-}]}{[\sim \mathrm{COOH}]} \tag{S1}$$

where $[\sim \mathrm{COO^-}]$ is the concentration of zwitterions and $[\sim \mathrm{COOH}]$ is the concentration of cations. We can write a similar equation for the ratio between the anion and zwitterion forms of an amino acid,

$$\mathrm{pH} = pK_2 + \log_{10} \frac{[\sim \mathrm{NH_2}]}{[\sim \mathrm{NH_3^+}]} \tag{S2}$$

where $[\sim \mathrm{NH_2}]$ is the concentration of anions and $[\sim \mathrm{NH_3^+}]$ is the concentration of zwitterions (this time highlighting the charge of the amine group). The pK_a values for amino acids can be found in the *CRC Handbook of Chemistry and Physics*.[S2] The fraction, f_i of of the ith ionizable group in the charged or ionized state can be found by rearranging the above equations. This is then multiplied by the charge of the respective ionizable group to get the net charge of the molecule as a function of pH (Ref.[S5] is useful for calculations). The results for 21 proteinogenic amino acids are shown in SI Figure S3. The plots are colour coordinated according to their protonation state at physiological pH.

S5 Conformer effects on the polarizability

S5.1 Gasphase conformers

Table S2: Comparison of the relative gasphase energies of glycine conformers, calculated using FHI-aims code (data collated by Ropo et al.[S6]) and GPAW code (this work). The conformers are labelled according to their filenames in the Ropo et al. dataset, in descending order of energetic stability. The energy of the most stable conformer is taken as the reference energy. The probabilities are calculated using the Boltzmann distribution with temperature $T = 298.15$ K.

	Relative energy (eV)		Boltzmann probability	
Conf.	FHI-aims[S6]	GPAW	FHI-aims	GPAW
1	0.000	0.000	0.748	0.892
2	0.034	0.060	0.201	0.086
3	0.088	0.117	0.024	0.009
4	0.098	0.122	0.017	0.008
5	0.117	0.141	0.008	0.004
6	0.155	0.172	0.002	0.001
7	0.225	0.235	0.000	0.000
8	0.230	0.241	0.000	0.000

Ropo et al. identified 8 stable glycine conformers in the gasphase[S6] and calculated the relative energies of the different conformers with FHI-aims code. The energy of the most stable conformer is taken as the reference energy. Table S2 compares the relative energies calculated with GPAW code, against the original results collated by Ropo et al. in the NOMAD repository,[S7] where good agreement is found. The table shows that the energetic stability of the glycine conformers decreases from $1 \rightarrow 8$.

We use the Boltzmann distribution to calculate the relative probability of glycine conformers. The probability distribution is a function of the relative energy, defined as $E_{\mathrm{rel}} = (E_i - E_{\mathrm{ref}})$, where E_i is the energy of the ith conformer and E_{ref} is the minimal element of the set $\{E_i\}$. The probability of the ith conformer can be written as

$$p_i = \frac{1}{Z} e^{-\beta(E_i - E_{\mathrm{ref}})} \tag{S3}$$

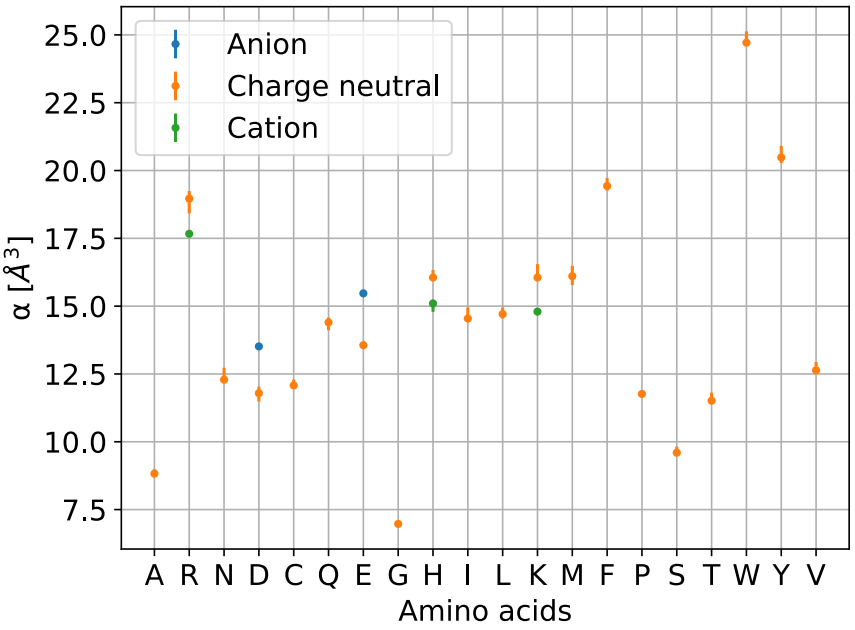

Figure S4: Expectation values for the polarizability of amino acid conformers in the gas phase. Relative energies of gasphase conformers are taken from the Ropo et al. dataset. Probabilities of the conformers are calculated using the Boltzmann distribution. We calculate the expectation value of the molecular polarizability using the conformers that have more than 1% probability of occurring.

where Z is the partition function, defined as

$$Z = \sum_{i}^{N} e^{-\beta(E_i - E_{\text{ref}})} \tag{S4}$$

and $\beta = k_B T$ where $k_B = 8.617 \times 10^{-5}$ eV K^{-1} is the Boltzmann constant and T is temperature (in Kelvin). Here, we consider the relative gas-phase energies of $N = 8$ glycine conformers in Table S2 for $T = 298.15$ K. The corresponding Boltzmann probabilities are listed in Table S2. Conformer 1 has a probability of $\approx 82 \pm 7\%$ and conformer 2 has a probability of $\approx 14 \pm 6\%$. Conformers $3 \rightarrow 8$ have a very low probability of occurring. We repeat this analysis for the other gasphase molecules in the Ropo et al. dataset and obtain the expectation values for the gas phase dipole polarizabilities, as shown in SI Figure S4.

The figure shows that the effects from different conformers on the gasphase polarizability is negligible. However, solvent effects can dramatically change the energy landscape of amino

acids. For example, none of the gasphase conformers listed in Table S2 contain charged functional groups, whereas in solution, glycine exists predominantly as a zwitterion.[S8] The zwitterion form of glycine is approximately 0.3 eV more stable than the the non-ionized form of glycine (energy difference calculated with GPAW code).

S5.2 Solvated conformers

In the following, we test whether the conformer variation is also small for solvated amino acids. We select arginine (R) as an example (an extreme case) because it has the largest number of conformers and the largest deviation in the gasphase polarizability (see Fig. S4). Fig. S5 compares the Boltzmann weighted distribution of R polarizabilities α in the gasphase (dashed lines) and solution (solid lines) relative to $\delta(\alpha) = \alpha/\bar{\alpha}$ -1, where $\bar{\alpha}$ is their expectation value. Part (a) shows the distribution for R cations and part (b) shows the distribution for net charge neutral R molecules. The peaks in the distribution correspond to groups of R conformers with a similar polarizability.

In the gasphase, there is a dominant peak, centered around the expectation value at $\delta(\alpha) = 0$. The associated R conformers are compact, curled up structures (labelled as G1a and G1b). There is greater polarizability dispersion in part b compared to part a because the G1b structure has many energetically similar conformers that have differences in their polarizability.[S6,S9] However, overall, there is a negligible difference between the polarizability of the lowest energy gasphase structure and the expectation value.

In the solvent, there are multiple peaks in the distribution, especially in part (b). This occurs when there are several groups of conformers that are energetically similar but have differences in their polarizability. By analysing the structures associated with each peak, we find that the polarizability groups are related to the elongation of the molecule. The S1a structure is completely unfolded and is the most energetically favourable conformer in water (neutral pH). This is in good agreement with the fact that R is a hydrophilic amino acid. The polarizability of S1a is close to the expectation value (less than 1% difference).

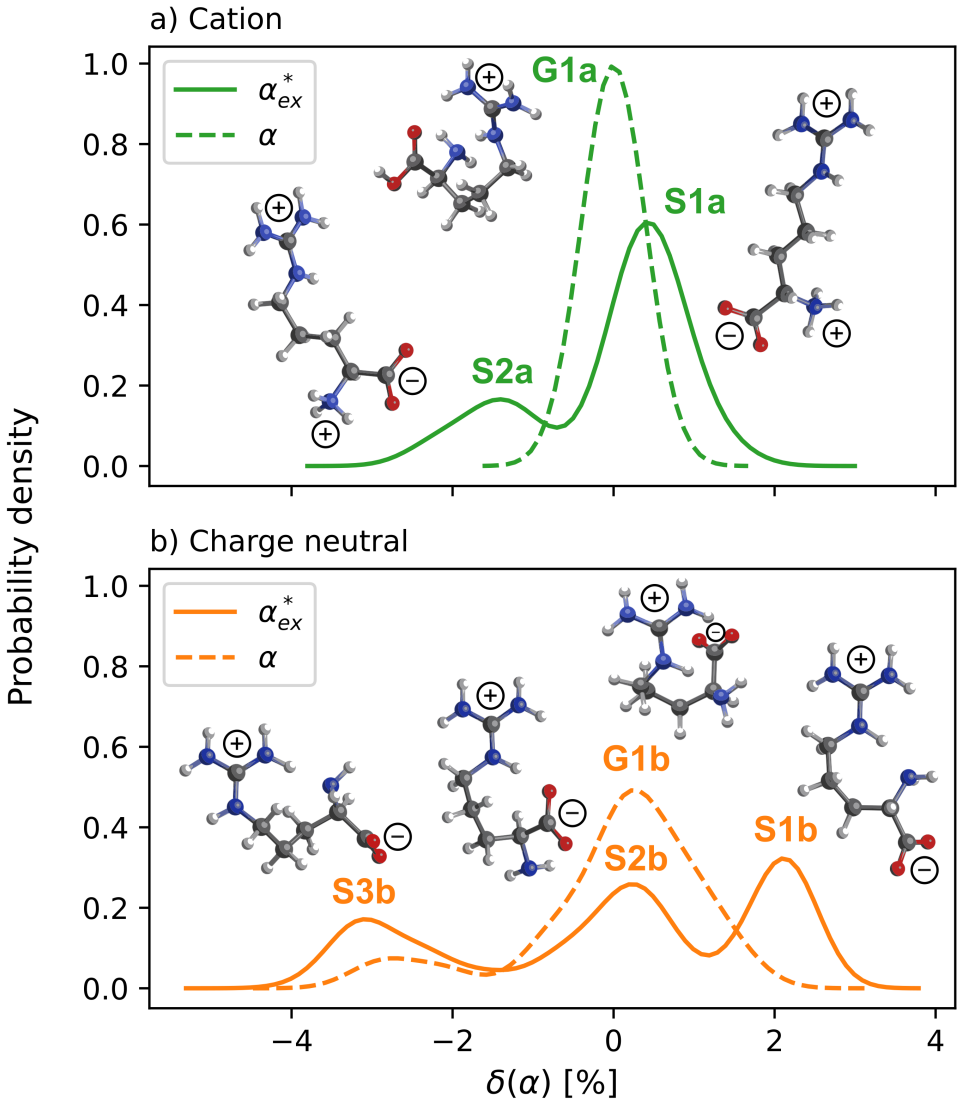

Figure S5: Conformational landscape of arginine (R): (a) cation structures and (b) net charge neutral structures. $\delta(\alpha)$ describes the percentage variation in the polarizability of R in the gasphase (dashed lines) and solvent (solid lines). The percentages $\delta(\alpha)$ are calculated relative to the expectation values of the solvent and gasphase polarizabilities, respectively. Peaks in the probability density distribution correspond to groups of R conformers with a similar polarizability. The associated structures represent a typical R conformer from each polarizability group. The gasphase structures are labelled using the letter 'G' and the solvated structures with the letter 'S'. In the gasphase, R prefers a compact, curled-up conformation (G1). In the solvent, the extended structures (S1) are preferred, but there are also contributions from stable "in-between" structures (S2 and S3), leading to a more distributed conformational landscape. The solvated structures (S1b, S2b and S3b) are energetically similar but vary in their polarizability, leading to a greater variation in $\delta(\alpha)$.

The S1b structure is also elongated, but not unfolded like S1a. This is likely due to the absence of the third hydrogen atom on the amine group, making an intramolecular hydrogen bond more energetically favourable. Here, we observe that the S2b and S3b structures become progressively more compact, affecting the polarizability of these conformers. We find that, overall, the variation is $\delta(\alpha)$ is minimal (on the order of a few percent). The lowest energy S1b conformer over-predicts the expected value of the excess polarizability by 2.2%. We note that in the case of hydrophobic amino acids, the lowest energy conformer may under-predict the polarizability by a similar amount. Based on this information, we conclude that using the lowest energy conformer to calculate molecular polarizability is a reasonable approximation for the amino acids considered in this study.

S6 Dipole moments based on Wannier orbitals.

Table S3: Permanent dipole moment μ and polarizability α of single water molecules in various environments determined by different methods.

quantity	value	method
μ gasphase	0.386 |e|Å	experimental vapour phase[S2]
μ gasphase	0.381 |e|Å	ours
μ gasphase	0.392 |e|Å	ours using Wannier orbitals
μ in water	0.614 |e|Å	explicit water using Wannier orbitals[S10]
μ in water	0.498 |e|Å	ours in $\varepsilon = \varepsilon_{\text{stat}} = 78$ PCM
μ in water	0.266 |e|Å	ours in PCM $\varepsilon = \varepsilon_{\text{stat}} = 78$ using Wannier orbitals
α in water	1.44 Å^3	CCSD of water clusters[S11]
α^{PCM}	2.03 Å^3	ours $\varepsilon = \varepsilon_{\text{stat}} = 78$
α^{PCM}	2.86 Å^3	ours $\varepsilon = \varepsilon_{\text{stat}} = 78$ using Wannier orbitals
α^{PCM}	1.69 Å^3	ours $\varepsilon = \varepsilon_{\text{opt}} = 1.77$
α^{PCM}	2.57 Å^3	ours $\varepsilon = \varepsilon_{\text{opt}} = 1.77$ using Wannier orbitals

S7 Data records

The self-consistent calculation of polarizabilities within the solvent and excess polarizabilities will be included in a new version of GPAW code (later than GPAW 24.1.0),[S12–S14] along with an associated documentation page and examples. Additional input files and structure files to reproduce the main results in this paper are available on GitLab: (https://gitlab.com/ag_walter/amino-acid-structures-and-polarizability).

References

(S1) Walter, M.; Moseler, M. Ab Initio Wavelength-Dependent Raman Spectra: Placzek Approximation and Beyond. *Journal of Chemical Theory and Computation* **2020**, *16*, 576–586.

(S2) Haynes, W. M., Ed. *CRC Handbook of Chemistry and Physics*; CRC Press, 2014.

(S3) Henderson, L. J. Concerning the relationship between the strength of acids and their capacity to preserve neutrality. *American Journal of Physiology-Legacy Content* **1908**, 173–179, Publisher: American Physiological Society.

(S4) Hasselbalch, K. *Die Berechnung der Wasserstoffzahl des Blutes aus der freien und gebundenen Kohlensäure desselben, und die Sauerstoffbindung des Blutes als Funktion der Wasserstoffzahl*; Julius Springer, 1916.

(S5) Sims, P. A. Use of a Spreadsheet To Calculate the Net Charge of Peptides and Proteins as a Function of pH: An Alternative to Using "Canned" Programs To Estimate the Isoelectric Point of These Important Biomolecules. *Journal of Chemical Education* **2010**, *87*, 803–808, Publisher: American Chemical Society.

(S6) Ropo, M.; Schneider, M.; Baldauf, C.; Blum, V. First-principles data set of 45, 892 isolated and cation-coordinated conformers of 20 proteinogenic amino acids. *Scientific Data* **2016**, *3*, 160009.

(S7) Ropo, M.; Baldauf, C.; Blum, V. NOMAD Repository Entry. 2015; `https://nomad-lab.eu/prod/v1/gui/dataset/doi/10.17172/NOMAD/20150526220502`.

(S8) Aikens, C. M.; Gordon, M. S. Incremental Solvation of Nonionized and Zwitterionic Glycine. *Journal of the American Chemical Society* **2006**, *128*, 12835–12850.

(S9) Maksić, Z. B.; Kovačević, B. Neutral vs. zwitterionic form of arginine—an ab initio study. *Journal of the Chemical Society, Perkin Transactions 2* **1999**, 2623–2629.

(S10) Silvestrelli, P. L.; Parrinello, M. Water Molecule Dipole in the Gas and in the Liquid Phase. *Phys. Rev. Lett.* **1999**, *82*, 3308–3311, Publisher: American Physical Society.

(S11) Hammond, J. R.; Govind, N.; Kowalski, K.; Autschbach, J.; Xantheas, S. S. Accurate dipole polarizabilities for water clusters n=2–12 at the coupled-cluster level of theory and benchmarking of various density functionals. *The Journal of Chemical Physics* **2009**, *131*, 214103, Publisher: American Institute of Physics.

(S12) Mortensen, J. J.; Hansen, L. B.; Jacobsen, K. W. Real-space grid implementation of the projector augmented wave method. *Physical Review B* **2005**, *71*, 035109.

(S13) Enkovaara, J.; Rostgaard, C.; Mortensen, J. J.; Chen, J.; Dułak, M.; Ferrighi, L.; Gavnholt, J.; Glinsvad, C.; Haikola, V.; Hansen, H. A.; Kristoffersen, H. H.; Kuisma, M.; Larsen, A. H.; Lehtovaara, L.; Ljungberg, M.; Lopez-Acevedo, O.; Moses, P. G.; Ojanen, J.; Olsen, T.; Petzold, V.; Romero, N. A.; Stausholm-Møller, J.; Strange, M.; Tritsaris, G. A.; Vanin, M.; Walter, M.; Hammer, B.; Häkkinen, H.; Madsen, G. K. H.; Nieminen, R. M.; Nørskov, J. K.; Puska, M.; Rantala, T. T.; Schiøtz, J.; Thygesen, K. S.; Jacobsen, K. W. Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. *Journal of Physics: Condensed Matter* **2010**, *22*, 253202.

(S14) Mortensen, J. J.; Larsen, A. H.; Kuisma, M.; Ivanov, A. V.; Taghizadeh, A.; Peterson, A.; Haldar, A.; Dohn, A. O.; Schäfer, C.; Jónsson, E. O.; Hermes, E. D.; Nilsson, F. A.; Kastlunger, G.; Levi, G.; Jónsson, H.; Häkkinen, H.; Fojt, J.; Kangsabanik, J.; Sødequist, J.; Lehtomäki, J.; Heske, J.; Enkovaara, J.; Winther, K. T.; Dulak, M.; Melander, M. M.; Ovesen, M.; Louhivuori, M.; Walter, M.; Gjerding, M.; Lopez-Acevedo, O.; Erhart, P.; Warmbier, R.; Würdemann, R.; Kaappa, S.; Latini, S.; Boland, T. M.; Bligaard, T.; Skovhus, T.; Susi, T.; Maxson, T.; Rossi, T.; Chen, X.; Schmerwitz, Y. L. A.; Schiøtz, J.; Olsen, T.; Jacobsen, K. W.; Thyge-

sen, K. S. GPAW: open Python package for electronic-structure calculations. 2023; http://arxiv.org/abs/2310.14776, arXiv:2310.14776 [cond-mat, physics:physics].