Supporting Information:

Understanding THz Spectra of Aqueous Solutions: Glycine in Light and Heavy Water

Jian Sun,[†,¶] Gudrun Niehues,[‡] Harald Forbert,[†] Dominique Decka,[‡]

Gerhard Schwaab,[‡] Dominik Marx,[†] and Martina Havenith[*,‡]

Lehrstuhl für Theoretische Chemie, Ruhr–Universität Bochum, 44780 Bochum, Germany, and Lehrstuhl für Physikalische Chemie II, Ruhr–Universität Bochum, 44780 Bochum, Germany

E-mail: martina.havenith@rub.de

[*]To whom correspondence should be addressed

[†]Lehrstuhl für Theoretische Chemie, Ruhr–Universität Bochum, 44780 Bochum, Germany

[‡]Lehrstuhl für Physikalische Chemie II, Ruhr–Universität Bochum, 44780 Bochum, Germany

[¶]Present Address: Theory of Condensed Matter Group, Cavendish Laboratory, University of Cambridge, J J Thomson Avenue, Cambridge CB3 0HE, United Kingdom

Measuring THz Spectra

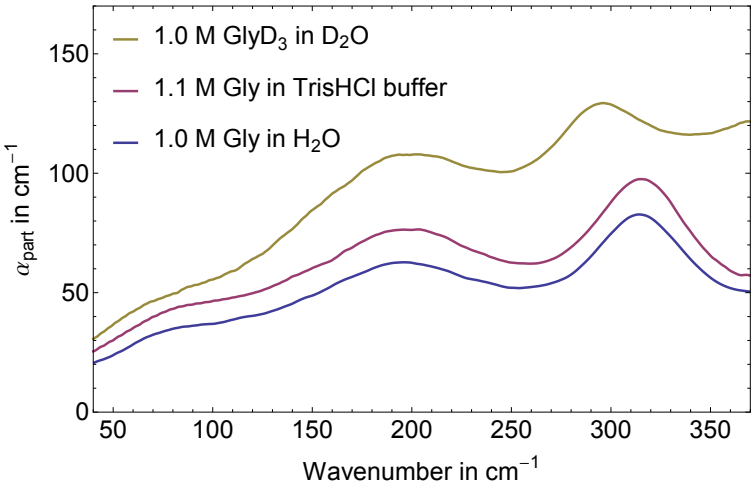


Figure S1: Comparison of the THz spectra of 1.0 M glycine solvated in neat water (blue), 1.1 M glycine in 50 mM TrisHCl buffer (red), and 1.0 M $GlyD_3$ in neat D_2O (yellow).

Normal glycine, Gly (Acros Organics, 99%) was solvated in 50 mM TrisHCl buffer pH 7.4 and deuterated glycine–N,N,O–d_3, $GlyD_3$ (SYNTHON Chemicals GmbH & Co.) in pure deuterium oxide (Deutero GmbH, 99,9%). In both cases, the zwitterionic form of the amino acids is expected to be dominant. To ensure that adding the TrisHCl buffer does not change the THz spectrum, glycine was also solvated in ultrapure water (see Fig. S1). In the spectral range investigated here, the spectral shapes of glycine solvated in neat water and in buffer solution are virtually identical. For the p-Ge system we used a commercial liquid cell (Bruker A145) with z-cut Quartz windows in a distance of 50 μm. All measurements were carried out at (20 ± 1)°C and the humidity was kept below 8 %. For FT–THz spectroscopy of the undeuterated glycine, CVD diamond windows (500 $\pm$ 100 μm thickness; Diamond Materials GmbH) were used with Kapton spacers of 35 μm,[1] while for glycine-N,N,O-d_3 dissolved in D_2O we used TPX (poly(4-methyl-1-pentene) windows (approx. 2.7 mm thickness) and Kapton spacers of 25 μm. We kept the liquid cell under nitrogen purged and temperature-controlled conditions (20 $\pm$ 0.1 °C) in a sample compartment, which was separated by polyethylene flaps from the evacuated rest of the spectrometer. The exact spacing size of both cell types was determined precisely by recording etalon fringes for the empty

cell.

The normal aqueous glycine solutions, Gly(aq), and the heavy solutions of deuterated glycine in D_2O, $GlyD_3$(aqD), were measured for different concentrations using the Bruker Vertex 80v wideband Fourier Transform spectrometer. The normal glycine solutions were also measured with our narrowband p-Ge laser THz difference spectrometer containing the unique beryllium p-doped germanium crystal, using a two channel set up.[2] The wideband low frequency absorbance is collected in the range between 50 and 350 cm^{-1} using the FT–THz spectrometer equipped with a liquid-He cooled silicon bolometer for detection and a mercury arc lamp as radiation source. Each spectrum was measured with resolutions of 2 cm^{-1} for Gly(aq) and 1 cm^{-1} for $GlyD_3$(aqD) and averaged over 128 scans. A normalized Gaussian function with $\sigma = 15$ cm^{-1} was used as a weighting function for a moving average in order to smoothen the spectra. In a separate set-up, our p-Ge laser difference spectrometer[4] records the averaged THz absorption in the frequency range from 2.3 to 2.8 THz (73–93 cm^{-1}) using a photoconductive germanium detector cooled with liquid helium.[3]

The total THz absorption follows Lambert-Beer's law

$$I(\omega) = I_0(\omega) \exp[-\alpha(\omega)\, d] \ , \tag{S1}$$

with d being the solution layer thickness. In the following 'solvent' refers to 50 mM TrisHCl buffer or D_2O in case of Gly(aq) and $GlyD_3$(aqD), respectively. After determination of the effective concentration $c_{\mathrm{solvent}}(c)$ for a given solute concentration c by density measurement, we could determine the partial absorption coefficient using

$$\alpha_{\mathrm{part}}(\omega, c) = \alpha_{\mathrm{sample}}(\omega, c) - \frac{c_{\mathrm{solvent}}(c)}{c_0} \alpha_{\mathrm{solvent}}(\omega) \ , \tag{S2}$$

where c_0 is the concentration of the neat solvent in moles per dm^3.

Special care was taken to minimize the influence of changes in transmission due to a change in index of refraction. Precise absorption spectra were obtained by measuring light

and heavy water (H_2O and D_2O) at the same temperature T using two spacer sizes. Assuming that the large water absorption allows one to neglect frequency-dependent two-beam interference (standing waves), all index of refraction effects due to reflection at air–window and window–water interfaces contribute in the same way to both measurements and the absorption coefficient can be directly retrieved from the logarithm of the ratio of the two measurements. For the measurements of Gly(aq), we additionaly obtained the absorption spectrum $\alpha_{\text{buffer}}(\omega)$ of the neat buffer solution by using bulk water as a reference. Finally, the absorption coefficients of glycine solutions at different concentrations were referenced to the neat solvent solution: 50 mM TrisHCl buffer in case of Gly(aq) and D_2O for $GlyD_3$(aqD).

Simulation and Assignment of THz Spectra

The AIMD simulations[5] are performed via Born–Oppenheimer propagation using the `Quickstep` electronic structure module in the `CP2k` package.[6] The PBE functional and dual–space norm–conserving pseudo potentials are employed together with a plane wave cutoff energy of up to 400 Ry and a TZV2P basis set.[6] A periodic cubic ≈ 9.85 Å supercell containing one glycine and 30 water molecules is used; doubling the number of water molecules for a similar molecule leads to an overall intensity decrease of α as in experiment but does not affect the lineshape function. The system was carefully equilibrated using massive Nosé–Hoover chain thermostats[5] at a rescaled temperature of 400 K to approximately counterbalance its systematic underestimation by about 20–30 % as done before.[7–9] Subsequently, 80 statistically independent initial conditions have been taken from canonical NVT trajectories and sampled as starting points of microcanonical NVE trajectories of 20 ps each using a timestep of 0.5 fs. Every 2 fs maximally localized Wannier functions have been computed (see Sec. 7.2.4 in Ref.[5]) and molecular dipole moments $\boldsymbol{\mu}_I(t)$ were constructed using the Wannier centers (see Sec. 7.2.6 in Ref.[5]). The data shown have been averaged over all NVE trajectories thus yielding proper canonical averages.

All force field molecular dynamics (FFMD) simulations were done with the GROMACS program package[10,11] in which one glycine molecules represented by the OPLS–AA force field[12,13] surounded by 510 flexible TIP4P water molecules[14] was hosted in a periodic 15.6 Å cubic box using a 0.2 fs timestep. The partial charge of the C_α atom was adjusted to $+0.09|e|$ as to yield an overall neutral zwitterionic glycine molecule. For each isotopic case, one NVT trajectory was generated from which every 40 ps initial conditions were sampled for subsequent 80 ps long NVE trajectories in order to compute from FFMD the THz spectra at NVT conditions.

In our so-called AIMD/pc approach, the same point charges (pc) as used in the FFMD simulations were used for calculating the dipole moment and thus the dipole autocorrelation function, but assigning these partial charges to the particle positions according to the trajectories that have been obtained previously from AIMD simulations. This method allows one to qualitatively probe, with reference to AIMD data, the impact of (electronic) polarizability on properties since the same trajectories are used in both cases, whereas the full self–consistent electronic structrure is used only in the AIMD approach to compute the total dipole moment (obtained from the Wannier centers as explained above).

It is well known that the total IR absorption cross section is given by the Fourier transform of the dipole auto-correlation function,

$$\alpha(\omega) = \frac{1}{n(\omega)} \frac{1}{6\epsilon_0 V c} \; \frac{\omega^2}{k_{\rm B}T} \; \int_{-\infty}^{+\infty} dt \;\; \langle \boldsymbol{M}(0)\boldsymbol{M}(t)\rangle \, e^{-i\omega t} \;\; , \tag{S3}$$

where $\boldsymbol{M}(t)$ is the total dipole moment of the sample of volume V at temperature T and $n(\omega)$ is its refractive index. In molecular liquids the total dipole moment can be decomposed into a sum of molecular dipole moments, $\boldsymbol{M}(t) = \sum_{I=1}^{N} \boldsymbol{\mu}_I(t)$, which are computed in AIMD from the centers of maximally localized Wannier functions, see e.g. Sec. 7.2 in Ref.[5] The refractive index is obtained after decomposing the product $\alpha(\omega)n(\omega)$, which is the quantity directly obtained from AIMD, into $\alpha(\omega)$ and $n(\omega)$ using the Kramers-Kronig relation[15] as

explained below in detail.

The computation of total vibrational spectra of aqueous solutions along these lines has made lively progress in the last decade and is by now a standard tool. However, the analysis of such spectra, i.e. the assignment of vibrational bands to atomic motions, is by no means straightforward when it comes to *low frequency modes* including their strong coupling to solvation shell dynamics, both being crucial to understand measured THz spectra of aqueous solutions. Important steps were the idea to impose a localization criterion in the frequency domain to define "effective modes"[16,17] followed by the introduction of so–called "generalized normal coordinates" that allow one to include large-amplitude motion, permutational symmetries and different conformations.[18,19]

Here, we generalize these ideas significantly, which enables us to pinpoint the individual contributions of correlated atomic motion to the IR reponse of aqueous solutions down the low–frequency regime and thus to assign THz spectra. A first step is to include n specific interfacial water molecules, which define what we call a "supermolecular solvation complex" (SSC), denoted in the following as Gly(n H_2O), out of all N water molecules in the sample. It is important to realize that the SSC model, including the associated supermolecular dipole moment computed from the electronic structure via Wannier analysis, is exclusively introduced to *analyze* the trajectories of the solution, in distinct difference to any microsolvation or continuum approach.

For the present inter-molecular liquid-phase technique, we exclusively use Cartesian velocities in a molecule–oriented reference frame instead of internal coordinate velocities. To directly obtain IR spectra and not just power spectra, we also include the centers of maximally localized Wannier functions as pseudo atoms and instead of mass weighting, multiply all velocities by the effective charge q_i of the corresponding center. This approach yields a dipole velocity cross-correlation matrix $C_{i\zeta,j\xi}(\omega)$ between charge-weighted Cartesian velocity

components ζ of center i and ξ of center j,

$$C_{i\zeta,j\xi}(\omega) = \int_{-\infty}^{+\infty} dt\; q_i q_j \,\langle v_{i\zeta}(0)\; v_{j\xi}(t)\rangle\, e^{-i\omega t} \;\; , \tag{S4}$$

where $v_{i\zeta}(t)$ is the ζ–component of the velocity for the i-th center at time t rotated into the molecule–oriented reference frame (of glycine). The dipole velocity cross-correlation matrix is then subsequently diagonalized in the sense of minimizing the power in the off-diagonal spectral elements,

$$\mathrm{D} \cdot \frac{\mathrm{C}(\omega)}{6\epsilon_0 V c\; n(\omega)\; k_\mathrm{B} T} \cdot \mathrm{D}^\dagger = \Lambda(\omega) \approx \mathrm{diag}(\Lambda_{k,k}(\omega)) \;\; , \tag{S5}$$

analogously to the generalized normal mode approach mentioned above, that is the transformation matrix D is given by

$$\mathrm{D} = \arg\min \sum_{k\neq\ell} \int_0^\infty |\Lambda_{k,\ell}(\omega)|^2 \, d\omega \;\; . \tag{S6}$$

The total IR absorption can then be expressed as a sum over mode-specific absorption coefficients $\alpha_k(\omega)$ and a remaining term describing the off-diagonal correlations

$$\alpha(\omega) = \sum_k \alpha_k(\omega) + \alpha_\mathrm{cross}(\omega), \tag{S7}$$

with

$$\alpha_k(\omega) \;\; = \;\; \Lambda_{k,k}(\omega)\; (\mathbf{d}_k \cdot \mathbf{d}_k) \;\; , \tag{S8}$$

$$\alpha_\mathrm{cross}(\omega) \;\; = \;\; \sum_{k\neq\ell} \Lambda_{k,\ell}(\omega)\; (\mathbf{d}_k \cdot \mathbf{d}_\ell) \;\; . \tag{S9}$$

The dipole direction vectors $\mathbf{d}_k$ of the modes k are determined by the transformation matrix D and have Cartesian components ξ given by $d_k^{(\xi)} = \sum_i D_{k,i\xi}$. Note, that the vectors $\mathbf{d}_k$ are not normalized and also contribute to the overall mode strength, though they do not

depend on the frequency, so the shape of a particular mode k is solely determined by the approximately diagonalized matrix $\Lambda(\omega)$. Obviously, in the ideal case, i.e. for $\Lambda_{k\neq\ell}(\omega) = 0$, the cross correlation term $\alpha_{\mathrm{cross}}(\omega)$ also vanishes, yielding an IR spectrum perfectly decoupled into separate modes. Since velocities for the Wannier centers were not available, we used in Eq. [S4] finite differences for all degrees of freedom and corrected $\mathrm{C}(\omega)$ in Eq. [S5] by multiplying with $\omega^2/(2\sin(\omega\delta/2))^2$, where δ is the time difference between steps. The sum of the molecular dipole moments of the remaining water molecules not belonging to the specified SSC is included in the analysis as a single additional coordinate triplet. Thereby, the dimensionality of the dipole velocity cross-correlation matrix $C_{i\zeta,j\xi}(\omega)$ is increased by three and thus the correlation of the total dipole moment of the cell is included within $C_{i\zeta,j\xi}(\omega)$. When we assume mirror symmetry with respect to the average plane of the glycine heavy atoms, it is possible to improve the statistics of the in total necessary 141 auto-correlation and 9870 cross-correlation spectra in the present case. The influence of the choice of the glycine–oriented reference frame was observed to be far smaller than the intrinsic statistical noise when switching between glycine–oriented reference frame and fixed–lab frame leaving both total dipole and glycine dipole spectra virtually unchanged.

First of all, we can base a mode–specific kinetic energy decomposition scheme on this formalism in order to quantify the relative contributions of hindered translations, rotations and vibrations of the entire SSC and its fragments, i.e. solvent and solute molecules, to particular THz modes. The total kinetic energy of a particular mode k of the entire SSC, $E_{\mathrm{tot}} = \sum_{i\xi} m_i v_{i\xi}^2/2$, is modeled using velocities $v_{i\xi} = D_{k,i\xi}/q_i$ given by the corresponding THz mode vector. The total sum can be split into contributions by water molecules ("Water") and solute in the SSC, and the solute modes can furthermore be decomposed into a center–of–mass part corresponding to hindered translations ("Gly–CoM"), rigid–body–like hindered rotations ("Gly–Rot"), and the purely intra–molecular vibrations ("Gly–Vib") as defined in detail below.

More importantly, the total absorption cross section of the aqueous solution can be recov-

ered systematically by this approach upon summing up the IR contributions of all the individual modes, $\alpha_k(\omega)$, plus the sum over properly weighted leftover (minimized) crossterms, see Eqs. [S7]–[S9]. This technique allows us, in particular, to include also selected solvent degrees of freedom and thus inter-molecular motion, namely the three water molecules that are strongly hydrogen-bonded to the protonated amino group of zwitterionic glycine in the present case. "Synthesizing" the THz spectra step-by-step, i.e. adding the frequency-dependent intensity contribution of one mode after the other, turns out to be a systematic approach to reconstruct the lineshape and to come to an assignment of the observed THz absorption features of Gly(aq).

Calculation of n(ω) via Kramers-Kronig relations

It is well known that the complex dielectric function, $\varepsilon_r(\omega) = \varepsilon'_r(\omega) + i\varepsilon''_r(\omega)$, obeys the Kramers-Kronig relation

$$\varepsilon'_r(\omega) = 1 + \frac{2}{\pi}\, P\!\int_0^\infty \frac{\omega' \varepsilon''_r(\omega')}{\omega'^2 - \omega^2} d\omega' \tag{S10}$$

for $\omega \geq 0$ and P denoting the principal value of the integral. The complex dielectric function is also connected to the complex refractive index, $\hat{n}(\omega) = n(\omega) + i\kappa(\omega)$, via

$$\hat{n}^2(\omega) = \varepsilon_r(\omega)\mu_r(\omega) \ . \tag{S11}$$

Assuming that the relative magnetic permeability, $\mu_r(\omega)$, is constant, i.e. unity, and using the connection $2\omega\kappa(\omega) = c\alpha(\omega)$, we can relate

$$\varepsilon''_r(\omega) = \frac{c\ \alpha(\omega) n(\omega)}{\omega} = \frac{1}{6\epsilon_0 V}\ \frac{\omega}{k_{\mathrm{B}}T}\ \int_{-\infty}^{+\infty} dt\ \ \langle \boldsymbol{M}(0)\boldsymbol{M}(t)\rangle\, e^{-i\omega t}\ , \tag{S12}$$

where we have used in the second step Eq. [3] from the manuscript. Since we do not obtain the electronic excitation spectrum, we need to split the integral in Eq. [S10] into three ranges.

We can calculate the integral up to a maximal frequency, $\omega_{\max} \approx 8300$ cm^{-1}, determined by the time resolution of our AIMD simulation, from $\omega_{\max}$ up to the lowest electronic excitation frequency, ω_{el}, the imaginary dielectric function, $\varepsilon_r''(\omega')$, is vanishingly small, thus the integral in that range is also close to zero. In the leftover range starting at ω_{el} we have $\omega' \gg \omega$ and therefore $\omega'^2 - \omega^2 \approx \omega'^2$. Thus, this range of the integral only contributes a nearly constant offset, which we determine from the experimental value of $n(\omega = 15000\ \mathrm{cm}^{-1}) \approx 1.33$ of bulk water assuming that the solution has a similar value there. The whole procedure is similar to previous approaches, see for example Refs.[15,20]

Once the real part of the dielectric function is determined, we can again with the help of Eq. [S11] get $\alpha(\omega)$ and $n(\omega)$ separately as

$$n(\omega) = \sqrt{\frac{\varepsilon_r'(\omega) + \sqrt{\varepsilon_r'^2(\omega) + \varepsilon_r''^2(\omega)}}{2}} \quad \text{and} \tag{S13}$$

$$\alpha(\omega) = \frac{\omega}{c}\sqrt{2\left(\sqrt{\varepsilon_r'^2(\omega) + \varepsilon_r''^2(\omega)} - \varepsilon_r'(\omega)\right)}\ . \tag{S14}$$

Kinetic energy decomposition into mode-specific contributions

To derive contributions for the various types of motion, the mode vector $D_{k,i\xi}/q_i$ for the mode k is taken as a velocity vector with components $v_{i\xi} = D_{k,i\xi}/q_i$. Under this assumption, the total kinetic energy of the k-th mode is then given by

$$E_{\mathrm{tot}} = \sum_{i\xi} \frac{1}{2} m_i v_{i\xi}^2\ , \tag{S15}$$

where m_i denotes the mass of the i-th center. This energy is then subdivided into several types of contributions. The water contribution has the same form as E_{tot}, but the sum is

restricted to water degrees of freedom,

$$E_{\mathrm{Water}} = \sum_{i\xi \text{ of water}} \frac{1}{2} m_i v_{i\xi}^2 \ . \tag{S16}$$

The remaining glycine contribution is split into three parts, namely a center–of–mass part (Gly–CoM), a rigid–body rotation contribution (Gly–Rot) and an intra–molecular vibration term (Gly–Vib), as described in the following. The center–of–mass momentum of glycine,

$$\boldsymbol{p}_{\mathrm{Gly-CoM}} = \sum_{i \text{ glycine}} m_i \boldsymbol{v}_i \ , \tag{S17}$$

defines the center–of–mass contribution to the kinetic energy,

$$E_{\mathrm{Gly-CoM}} = \frac{1}{2M} \boldsymbol{p}_{\mathrm{Gly-CoM}}^2 \ , \tag{S18}$$

where M is the total glycine mass

$$M = \sum_{i \text{ glycine}} m_i \ . \tag{S19}$$

The rigid–body rotation part of the energy,

$$E_{\mathrm{Gly-Rot}} = \frac{1}{2} \, \boldsymbol{L}^T \cdot \mathrm{I}^{-1} \cdot \boldsymbol{L} \tag{S20}$$

is calculated from the total angular momentum of glycine,

$$\boldsymbol{L} = \sum_{i \text{ glycine}} m_i \boldsymbol{r}_i \times \boldsymbol{v}_i \tag{S21}$$

and the inverse of the moment of inertia tensor

$$I_{\zeta,\xi} = \sum_{i \text{ glycine}} m_i (\boldsymbol{r}_i^2 \delta_{\zeta,\xi} - r_{i\zeta} r_{i\xi}) \ , \tag{S22}$$

where the $\boldsymbol{r}_i$ denote the average positions of the atoms of glycine. Finally, the intra–molecular vibration energy is calculated as

$$E_{\mathrm{Gly-Vib}} = E_{\mathrm{tot}} - E_{\mathrm{Water}} - E_{\mathrm{Gly-CoM}} - E_{\mathrm{Gly-Rot}} \ . \tag{S23}$$

The mode contributions reported in the table of the manuscript are then percentage values of the respective kinetic energy contributions relative to E_{tot}.

Solvation of Gly(aq): $-NH_3^+$ *versus* $-COO^-$ group

Radial distribution functions are orientationally averaged quantities, whereas spatial distribution functions unfold correlations in full three-dimensional space. Thus, they are most useful to analyze solvation shell patterns around non-spherical molecules that contain in addition distinct functional groups. The spatial distribution function of zwitterionic glycine in water is depicted in Fig. S2 using two different isovalues to represent the number density stemming from solvation water molecules as given by the positions of the corresponding oxygen nuclei.

The solvation shell of the $-NH_3^+$ group is very well structured in space as a result of 3 hydrogen bonds that are donated by the 3 protons of this cationic group, which leads to 3 pronounced peaks of high density (see left panel in Fig. S2). The average coordination number of the $-NH_3^+$ group amounts to about 3.4 as reported earlier,[9] which is due to additionally incoming interfacial water molecules as a result of thermal fluctuations in the solvation sphere. The situation is distinctly different for the $-COO^-$ group. In this case there are formally 5 lone electron pairs that would be available to accept 5 hydrogen bonds from closeby water molecules. However, the reported[9] coordination number of ≈ 4.4 is *reduced* with respect to the ideal value and, moreover, there are only 3 peaks visible in the spatial distribution function of oxygens. In addition, one of the peaks turns out to be weak, which is the one that is sterically hindered by the solvation shell around the $-NH_3^+$ group, whereas the

peak corresponding to water molecules pointing away from the entire molecule is pronounced (see left panel in Fig. S2). Still, there are on average about 4.4 molecules in the first solvation shell of the carboxylate group. This solvation pattern could be called "frustrated" (in view of only 3 inequivalent peaks for 4.4 solvating water molecules to be hosted), which results in vivid interchanges of water molecules w.r.t. the carboxylate oxygens to which they donate hydrogen bonds *within* the first hydration sphere of the anionic group. In conclusion, the solvation shell of the $-COO^-$ group is much less well defined than that of the $-NH_3^+$ group both in terms of structure and dynamics.

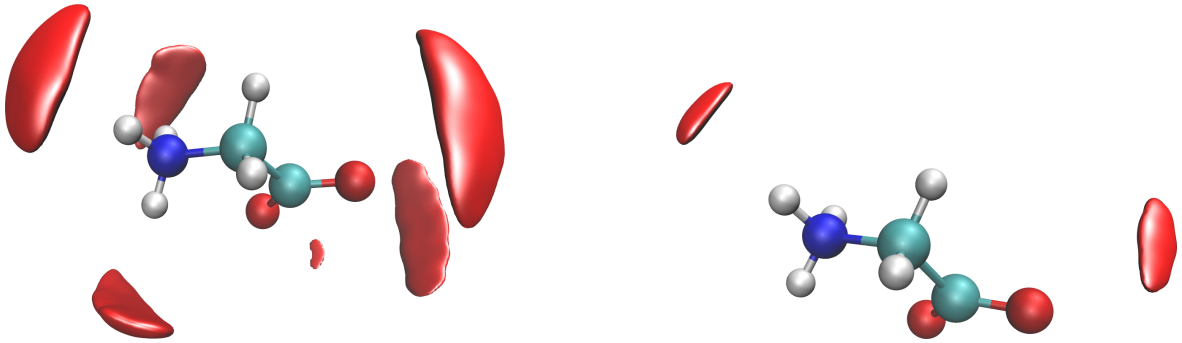


Figure S2: Spatial distribution function of water around solvated glycine, Gly(aq). The red isosurfaces corresponding to the oxygen density are presented at "low" and "high" isovalues in the left and right panels, respectively, in order to visualize the different solvation patterns of the cationic and anionic groups (see text). The solute–fixed reference frame is defined by the glycine molecule as shown where the heavy atoms are coplanar.

Furthermore, it is of interest to analyze the limiting solvation behavior in the sense of asking the question which site of the cationic and anionic functional group of the amino acid is the "most strongly solvated" one in a statistical sense. The answer is clearly seen by inspecting the right panel of Fig. S2 where a larger isovalue than in the left panel has been chosen in order to provide a stricter look at the hydration spheres of these two groups. Interestingly, it is that oxygen of the $-COO^-$ group and that proton of the $-NH_3^+$ group which are most distant from each other accross the averaged solvation complex and the oxygen sites of both water molecules are coplanar with the heavy atoms of glycine.

This analysis of the spatial distribution function provides us with important information

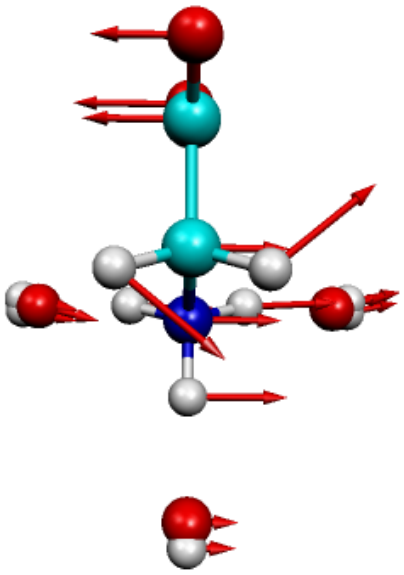
in order to define SSC models, Gly(n H_2O), that are at the same time scientifically meaningful and computationally practical. The SSC concept relies on the idea that individual water molecules are not only bound "sufficiently long" to a particular solute site, but they must also remain "sufficiently well localized" w.r.t. the solute molecule in a solute–fixed reference frame in order to provide "sufficiently robust" structural motifs for subsequent mode analysis. For the $-NH_3^+$ group it turned out that it is possible to include simultaneously all three solvating water molecules, which defines our so–called Gly(3 H_2O) model on which the comprehensive vibrational analysis is carried out using the methods introduced above. This leads to the THz modes for which the dynamical animations are included in this Supporting Information, whereas their mode–specific absorption cross sections as well as the kinetic energy decomposition are presented in Fig. 2 and Table I of the manuscript, respectively.

Mode analysis of this ill–defined solvation shell of the $-COO^-$ group nevertheless turns out to be possible upon considering only its most strongly bound solvating water molecule as easily identified with the help of the right panel of Fig. S2. This so–called Gly(1 H_2O) model provides us with the mode–specific absorption cross sections as well as the kinetic energy decomposition for all its THz modes as presented in Fig. S3 and Table S6 of this Supporting Information, *vide infra*.

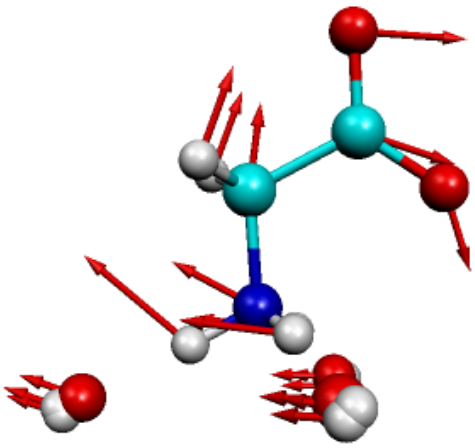
THz modes of glycine with 3 water molecules at the $-NH_3^+$ group

On the following pages Tables S1 to S4 contain the dynamical animations of all THz modes (see text) as obtained from analyzing the Gly(3 H_2O) supermolecular solvation complex, whereas Table S5 contains the two intra–molecular THz modes obtained from analyzing only the solute (Gly–only, see text). The animations can be activated and manipulated separately using the control panels below each figure.

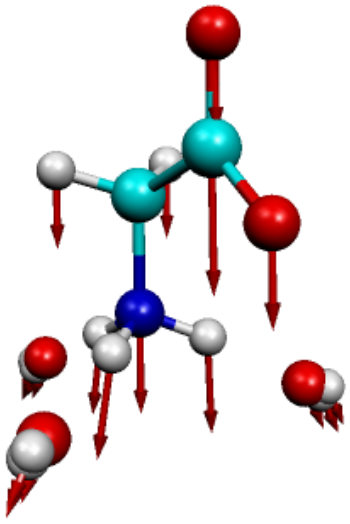
Table S1: THz modes (< 400 cm^{-1}) of zwitterionic Glycine + 3 Water molecules (part I).



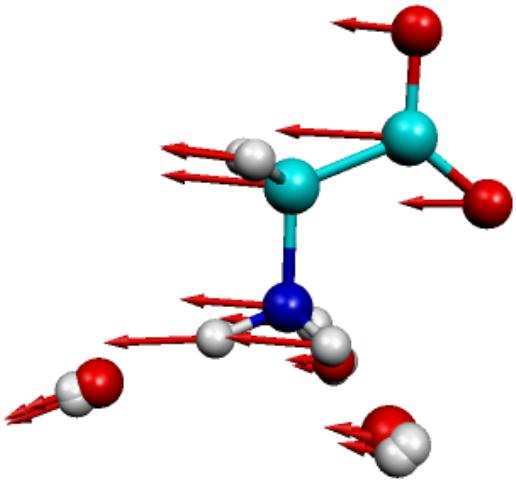
cage libration I : 82 cm^{-1}



cage libration II : 90 cm^{-1}

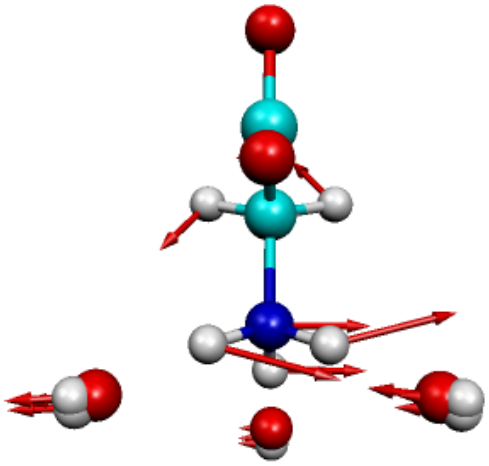


cage rattling I : 62 cm^{-1}

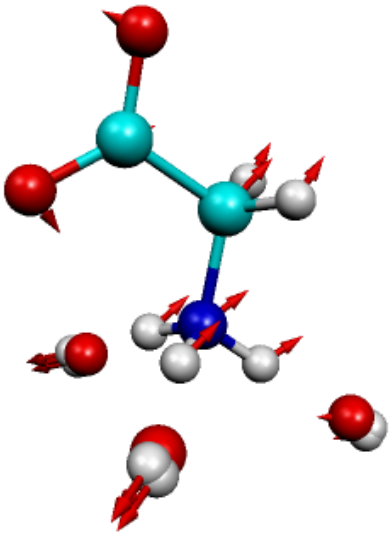


cage rattling II : 73 cm^{-1}

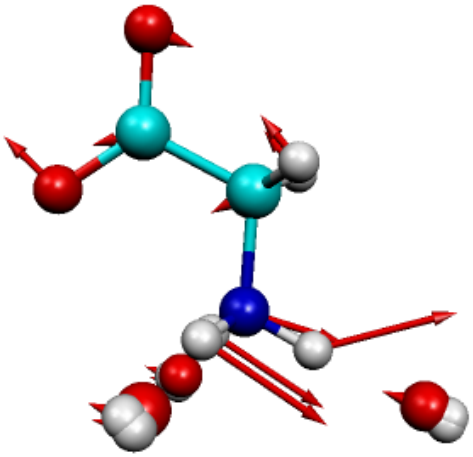
Table S2: THz modes (< 400 cm^{-1}) of zwitterionic Glycine + 3 Water molecules (part II).



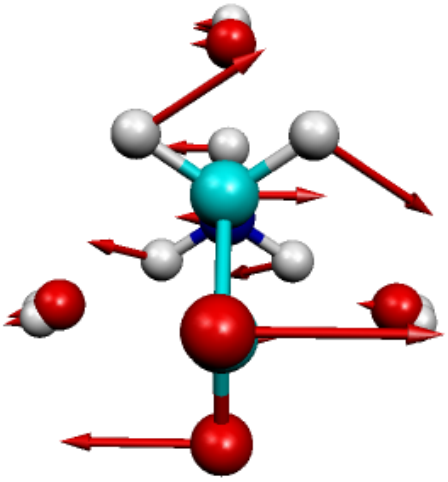
HB stretch + C-C twist: 247 cm^{-1}



HB stretch I : 210 cm^{-1}

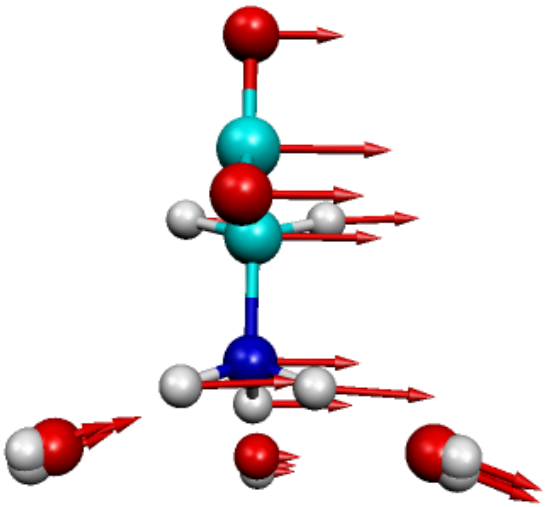


N-C-C-O open/close : 304 cm^{-1}

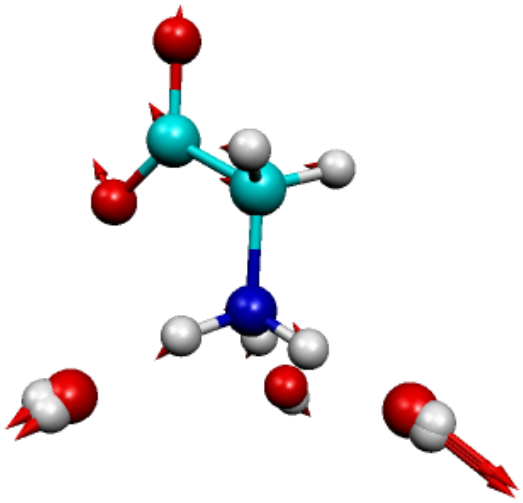


cage libration III : 89 cm^{-1}

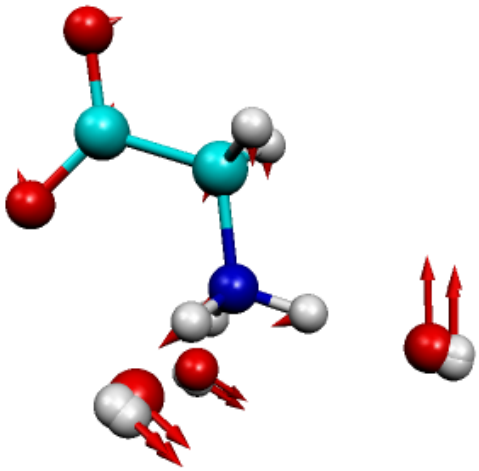
Table S3: THz modes (< 400 cm^{-1}) of zwitterionic Glycine + 3 Water molecules (part III).



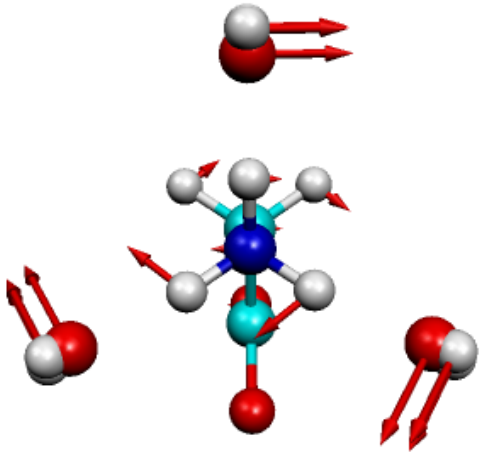
cage rattling III : 64 cm^{-1}



HB stretch II : 218 cm^{-1}

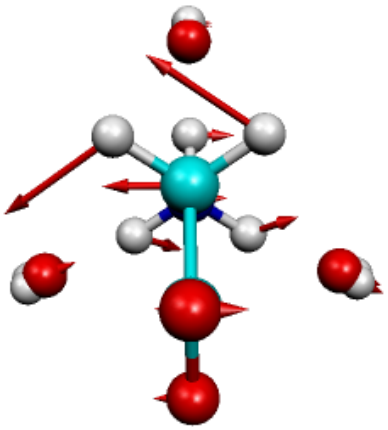


HB bend I : 102 cm^{-1}



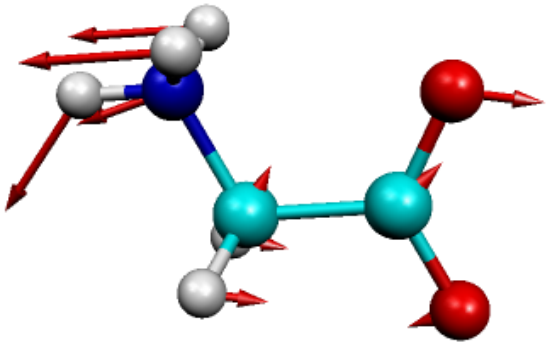
HB bend II : 90 cm^{-1}

Table S4: THz modes (< 400 cm^{-1}) of zwitterionic Glycine + 3 Water molecules (part IV).

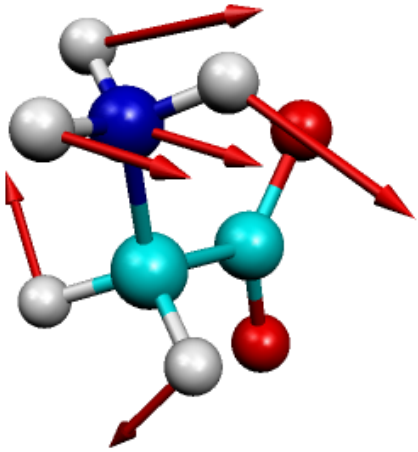


C_α out-of-plane : 125 cm^{-1}

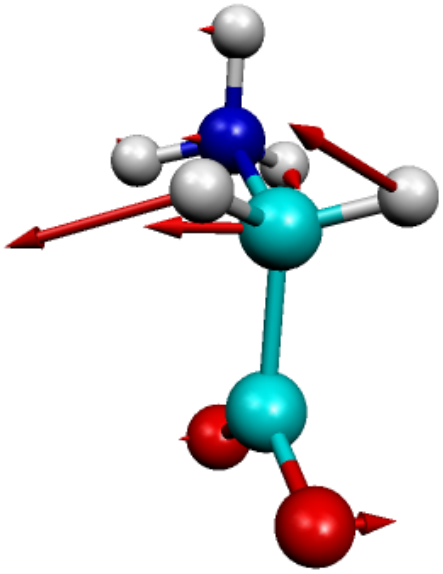
Table S5: Purely intra-molecular THz modes ($<$ 400 cm^{-1}) of zwitterionic Glycine.



N-C-C-O open/close : 305 cm^{-1}



C-C twist : 170 cm^{-1}



C_α out-of-plane : 137 cm^{-1}

Analysis of contributions due to the –COO^- group

The comprehensive analysis of the Gly(1 H_2O) model, carried out in full analogy to that of the Gly(3 H_2O) model, is presented in this section. The mode–specific absorption cross sections depicted in Fig. S3 herein can be compared one–to–one to those obtained from the Gly(3 H_2O) model as presented in Fig. 2 of the manuscript. It is remarkable to observe that the overall structure of the THz reponse is very comparable to that obtained from the larger SSC model, Gly(3 H_2O). In particular, we identify several cage rattling and libration modes at low frequencies (noting that "cage libration I" is not present in the Gly(1 H_2O) model) as well as the prominent "N-C-C-O opening/closing" mode all at similar frequencies and of comparable relative intensities compared to those extracted from the Gly(3 H_2O) model (see Table S6 for numbers and direct comparison of the two SSC models). The similarity of the different classes of modes is quantified by the decomposition of the mode–specific kinetic energy into contributions due to water motion ("Water") and solute motion, i.e. hindered translations ("Gly–CoM"), hindered rotations ("Gly–Rot") and intra–molecular vibrations ("Gly–Vib"), according to the analysis compiled in Table S6. Obviously, the "Water" contributions are less pronounced in the SSC that contains only one water molecule, Gly(1 H_2O), compared to the larger Gly(3 H_2O) model. The "C–C twisting" mode is also present, albeit being much broader and, in view of the geometric nature of the Gly(1 H_2O) model, evidently without any coupling to HB stretching motion (see Table S6). Again, peak intensity and position (when considering the average frequencies rather than those at the maximum as indicated by the two entries following the stars in Table S6) of this mode are very similar according to the two SSC models.

Adressing the THz signature of HB modes, we again find inter–molecular bending and stretching modes when using the Gly(1 H_2O) model where the –COO^- group is solvated. They are located at similar frequencies as in the larger SSC, i.e. below 100 cm^{-1} and around 200 cm^{-1}, respectively. The former contributions are very weak and thus negligible in terms of THz intensity in full accord with the results obtained from the Gly(3 H_2O) model. How-

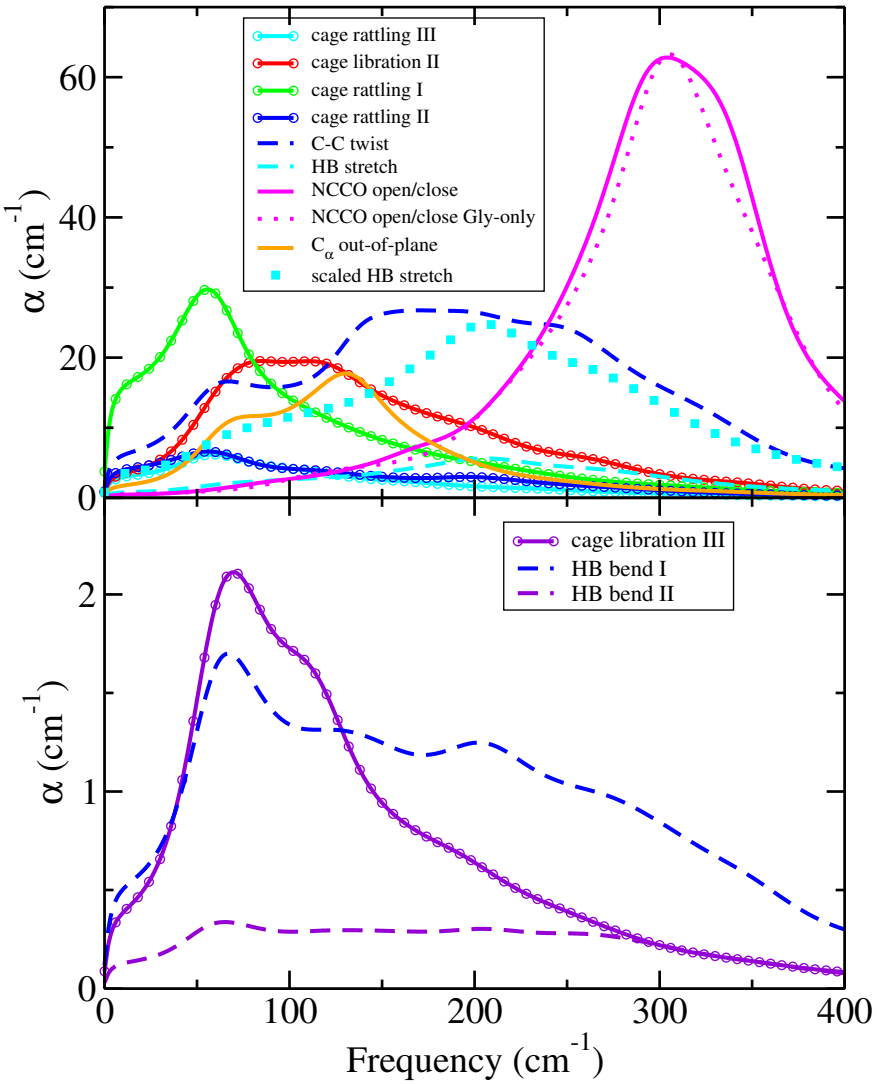


Figure S3: Intensities of THz modes of Gly(aq) according to the Gly(1 H_2O) model. All mode-specific absorption cross sections $\alpha_k(\omega)$ of the Gly(1 H_2O) supermolecular solvation complex having intensity maxima below 400 cm^{-1} with high and low absolute intensities are compiled in (A) Top and (B) Bottom to be directly compared to Fig. 2 from the manuscript obtained using the Gly(3 H_2O) SSC model (note that some curves have been shifted from the bottom to the top panel due to intensity increase while keeping the same labeling and color code). The "scaled HB stretch" mode (light-blue squares) is obtained by scaling the corresponding "HB stretch" mode (light-blue dashed line) using the average coordination number 4.4 of the $-COO^-$ group as explained in the text together with Fig. S4. See labels and text for mode assignments and Table S6 for energy decomposition. Note that the dotted line in panel (A) is not obtained from the SSC but the N-C-C-O opening/closing mode of Gly–only.

ever, at variance with the Gly(3 H_2O) model where the $-NH_3^+$ group is completely solvated using three water molecules, we use only the most strongly solvating water molecule for the carboxylate group. This certainly implies an underestimation of the contribution of water motion and, in particular, of HB stretching modes to the overall THz response in the frequency window around 200 cm^{-1}. The following procedure allows us to approximately correct for the missing intensity contribution.

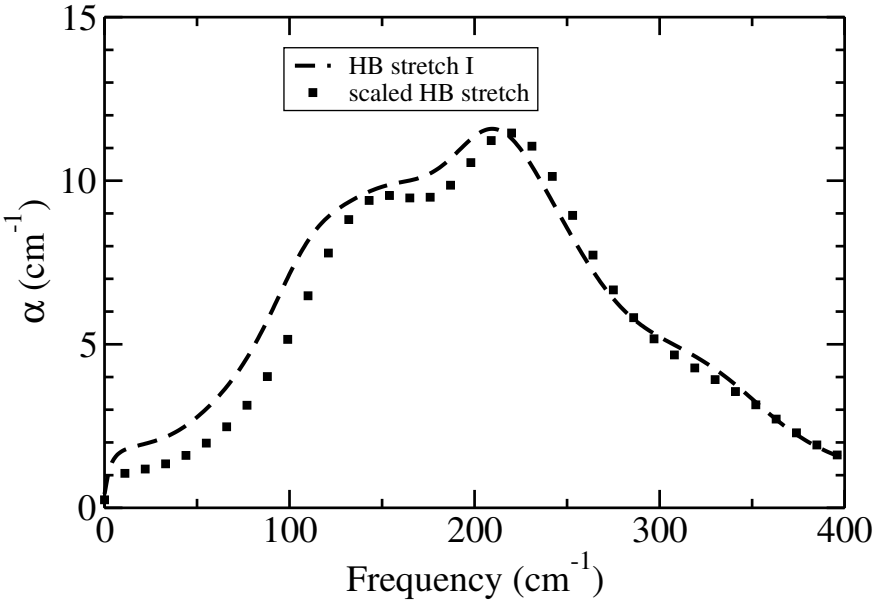


Figure S4: Benchmarking the additivity approximation of the Gly(1 H_2O) SSC model using the $-NH_3^+$ group of Gly(aq) in the Gly(3 H_2O) SSC model as a reference. The mode–specific absorption cross section of the "HB stretch I" mode according to rigorous analysis of the Gly(3 H_2O) model (taken from Fig. 2 of the manuscript) is shown using the dashed line. The squares show the mode–specific absorption cross section due to HB streching obtained by taking into account only a single water molecule in the solvation shell of the $-NH_3^+$ group and multiplying its cross section with the coordination number of the $-NH_3^+$ group, i.e. 3.4 as determined earlier.[9] The specific water molecule is the most strongly solvating water molecule in the $-NH_3^+$ solvation shell according to the spatial analysis of Fig. S2 and resides in the plane spanned by the heavy atoms of glycine. Both characteristics are identical for the specific water molecule that is considered in the solvation shell of $-COO^-$ according to the Gly(1 H_2O) model that leads to the cross section marked with squares in the top panel (A) of Fig. S3 after scaling the "HB stretch" cross section of the Gly(1 H_2O) model by the coordination number of the $-COO^-$ group, i.e. 4.4 as determined earlier.[9]

In order to estimate the HB contribution of the fully solvated carboxylate group, it might be reasonable to neglect the HB bending contributions in view of their vanishingly small intensities (cf. Table S6) and to assume an approximate additivity of HB stretching modes

which scales linearly with the number of contributing water molecules in the solvation shell. The latter number is given by the usual coordination number of the respective functional group. Before applying this idea to the $-COO^-$ group, however, this approximation is tested for the $-NH_3^+$ group where the dipolar response of the full solvation shell is available from the Gly(3 H_2O) SSC model. The comparison carried out in Fig. S4 supports the usefulness of this approximation, where the dashed line is the reference mode–specific THz absorption cross section using the full solvation shell of the cationic group from the Gly(3 H_2O) model. This compares favorably to the squares resulting from invoking the additivity assumption based on a single water molecule that is hydrogen bonded to the $-NH_3^+$ group followed by coordination number scaling of that mode–specific cross section (see caption for details).

Table S6: Decomposition (given in percent, deviations from 100 % due to rounding errors) of the kinetic energy of the THz modes of the Gly(3 H_2O) and Gly(1 H_2O) SSC models of the $–NH_3^+$ and $–COO^-$ functional groups, respectively, reported mode–by–mode in the first and second rows, respectively (see text). The peak frequencies in the rows marked with a star (*) have been obtained by averaging the corresponding mode–specific absorption cross sections to take broadness into account, in addition to reporting as usual their frequencies at the peak maximum provided in the preceding rows. Modes without entries in the second row do not exist in the smaller Gly(1 H_2O) SSC model.

Mode	Frequency	Intensity	Gly–CoM	Gly–Rot	Gly–Vib	Water
cage rattling I	62 cm^{-1}	21 cm^{-1}	85	< 1	5	11
water at $–COO^-$	56 cm^{-1}	30 cm^{-1}	86	< 1	5	8
cage rattling II	73 cm^{-1}	7 cm^{-1}	80	< 1	4	15
water at $–COO^-$	56 cm^{-1}	7 cm^{-1}	89	1	6	4
cage rattling III	64 cm^{-1}	1 cm^{-1}	78	< 1	2	19
water at $–COO^-$	59 cm^{-1}	6 cm^{-1}	75	19	4	3
cage libration I	82 cm^{-1}	41 cm^{-1}	5	74	4	17
water at $–COO^-$	—	—	—	—	—	—
cage libration II	90 cm^{-1}	9 cm^{-1}	6	68	4	22
water at $–COO^-$	85 cm^{-1}	20 cm^{-1}	7	80	6	7
cage libration III	89 cm^{-1}	< 1 cm^{-1}	2	91	4	3
water at $–COO^-$	70 cm^{-1}	2 cm^{-1}	1	93	5	1
NCCO open/close	304 cm^{-1}	59 cm^{-1}	2	1	92	5
water at $–COO^-$	304 cm^{-1}	63 cm^{-1}	1	< 1	99	< 1
C_α out-of-plane	125 cm^{-1}	2 cm^{-1}	4	6	75	15
water at $–COO^-$	130 cm^{-1}	18 cm^{-1}	12	32	56	1
HB stretch + C-C twist	247 cm^{-1}	37 cm^{-1}	13	30	20	37
*	299 cm^{-1}					
water at $–COO^-$	170 cm^{-1}	27 cm^{-1}	15	46	39	< 1
*	283 cm^{-1}					
HB stretch I	210 cm^{-1}	12 cm^{-1}	9	21	5	66
water at $–COO^-$	206 cm^{-1}	6 cm^{-1}	8	12	2	77
HB stretch II	218 cm^{-1}	2 cm^{-1}	17	4	6	74
water at $–COO^-$	—	—	—	—	—	—
HB bend I	102 cm^{-1}	< 1 cm^{-1}	< 1	8	< 1	92
water at $–COO^-$	67 cm^{-1}	2 cm^{-1}	5	< 1	< 1	95
HB bend II	90 cm^{-1}	< 1 cm^{-1}	< 1	< 1	2	98
water at $–COO^-$	65 cm^{-1}	< 1 cm^{-1}	2	4	1	94

This reassuring outcome allows one to apply the additivity approximation also to the –COO^- group that is hydrogen bonded to a single water molecule according to the Gly(1 H_2O) model in order to estimate the THz reponse that is expected from its *full solvation shell* consisting on average of 4.4 water molecules. It is important to note that the water molecule hydrogen bonded to the –COO^- group in the Gly(1 H_2O) model is similar to the one used to test the additivity approximation in that it is also coplanar with the heavy atoms of glycine and that it is also the most strongly solvating one in the anionic hydration sphere as evidenced by the right panel of Fig. S2. This procedure provides us with an estimate of the *total* HB stretching contribution due to *full* solvation of the –COO^- group of Gly(aq) as shown by the squares in the top panel of Fig. S3 herein (reproduced in Fig. 2 of the manuscript). Note also that the decomposition of the kinetic energy of the underlying HB stretching mode according to the Gly(1 H_2O) SSC model is very similar to that of the corresponding mode (i.e. "HB stretch I") of the Gly(3 H_2O) model according to the data compiled in Table S6, which also holds for the much less intense HB bends. In summary, the spectral range around 200 cm^{-1}, intermediate to the cage and intra–molecular frequencies below and above, respectively, gains considerable intensity after adding the dipolar response of the full hydration sphere around the carboxylate group. Apart from providing additional insights, this agreement when it comes to similar classes of modes in the Gly(1 H_2O) and Gly(3 H_2O) SSC models strongly supports both the basic idea and the usefulness of the SSC concept for assigning the THz response of solvated molecules in terms of modes. Most importantly, this finding also suggests that the *total THz response* of large molecules, such as biopolymers, can be approximatively constructed by computing all *incremental contributions* within the additivity approximation. This amounts to computing the response due to cage motion, intra–molecular modes, and inter–molecular HB dynamics around strongly solvated functional groups. The local response due to individual functional groups can certainly be addressed using *ab initio* simulations, whereas suitably parameterized polarizable force fields will be required to address large–scale motion.

References

(1) Schmidt, D. A.; Birer, Ö.; Funkner, S; Born, B. P.; Gnanasekaran, R.; Schwaab, G. W.; Leitner, D. M.; Havenith, M. *J. Am. Chem. Soc.* **2009**, *131*, 18512–18517.

(2) Born, B.; Weingärtner, H.; Bründermann, E.; Havenith, M. *J. Am. Chem. Soc.* **2009**, *131*, 3752–3755.

(3) Funkner, S; Niehues, G.; Schmidt, D. A.; Heyden, M.; Schwaab, G.; Callahan, K. M.; Tobias, D. J.; Havenith, M. *J. Am. Chem. Soc.* **2012**, *134*, 1030–1035.

(4) Bergner, A.; Heugen, U.; Bründermann, E.; Schwaab, G.; Havenith, M.; Chamberlin, D. R.; Haller, E. E. *Rev. Sci. Instrum.* **2005**, *76*, 063110-1–5.

(5) Marx, D.; Hutter, J. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*; Cambridge University Press: Cambridge, 2009.

(6) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. *Comp. Phys. Comm.* **2005**, *167*, 103–128.

(7) Heyden, M.; Sun, J.; Funkner, S.; Mathias, G.; Forbert, H.; Havenith, M.; Marx, D. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 12068–12073.

(8) Heyden, M.; Sun, J.; Forbert, H.; Mathias, G.; Havenith, M.; Marx, D. *J. Phys. Chem. Lett.* **2012**, *3*, 2135–2140.

(9) Sun, J.; Bousquet, D.; Forbert, H.; Marx, D. *J. Chem. Phys.* **2012**, *133*, 114508-1–10.

(10) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comp. Phys. Comm.* **1995**, *91*, 43–56.

(11) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

(12) Jorgensen, W. L.; Maxwell, D. S.; Tirado–Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(13) Kaminski, G. A.; Friesner, R. A.; Tirado–Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.

(14) Lawrence, C. P.; Skinner, J.L. *Chem. Phys. Lett.* **2003**, *372*, 842–847.

(15) Iftimie, R.; Tuckerman, M. E. *J. Chem. Phys.* **2005**, *122*, 214508.

(16) Martinez, M.; Gaigeot, M. P.; Borgis, D.; Vuilleumier, R. *J. Chem. Phys.* **2006**, *125*, 144106.

(17) Gaigeot, M. P.; Martinez, M.; Vuilleumier, R. *Mol. Phys.* **2007**, *105*, 2857–2878.

(18) Mathias, G.; Baer, M. D. *J. Chem. Theory. Comput.* **2011**, *7*, 2028–2039.

(19) Mathias, G.; Ivanov, S. D.; Witt, A.; Baer, M. D.; Marx, D. *J. Chem. Theory. Comput.* **2012**, *8*, 224–234.

(20) Johnson, D. W. *J. Phys. A - Math. Gen.* **1975**, *8*, 490–495.