

Terahertz absorption of lysozyme in solution

Daniel R. Martin¹ and Dmitry V. Matyushov²¹Department of Physics, Arizona State University, PO Box 871504, Tempe, Arizona 85287²Department of Physics and School of Molecular Sciences, Arizona State University, PO Box 871504, Tempe, Arizona 85287^{a)}

Absorption of radiation by solution is described by its frequency-dependent dielectric function and can be viewed as a specific application of the dielectric theory of solutions. For ideal solutions, the dielectric boundary value problem separates the polar response into the polarization of the void in the liquid, created by the solute, and the response of the solute dipole. In the case of a protein as a solute, protein nuclear dynamics do not project on significant fluctuations of the dipole moment in the terahertz domain of frequencies and the protein dipole can be viewed as dynamically frozen. Absorption of radiation then reflects the interfacial polarization. Here we apply an analytical theory and computer simulations to absorption of radiation by an ideal solution of lysozyme. Comparison with experiment shows that Maxwell electrostatics fails to describe the polarization of the protein-water interface and the “Lorentz void”, which does not anticipate polarization of the interface by the external field (no surface charges), better represents the data. An analytical theory for the slope of the solution absorption against the volume fraction of the solute is formulated in terms of the cavity field response function. It is calculated from molecular dynamics simulations in good agreement with experiment. The protein hydration shell emerges as a separate sub-ensemble, which, collectively, is not described by the standard electrostatics of dielectrics.

I. INTRODUCTION

Inserting a solute into a molecular liquid leads to a nonlinear perturbation of its density.¹ The density profile formed around a solute is a sensitive function of the strength of the intermolecular potential. Weak dewetting^{2,3} or, alternatively, collapse of hydration shells⁴⁻⁶ occurs for hydration when the solute-water interaction is either weaker (dewetting) or stronger (collapse) than the water-water interaction.

In contrast to the diversity of scenarios for the density profile of hydration layers, the orientational structure of the solute-solvent interface is typically viewed as a specific case of the standard dielectric boundary-value problem. In this picture, the dipoles of water orient along the lines of the solute’s electric field according to the rules of electrostatics.⁷ There are a number of fundamental reasons why this simplistic picture can break down,⁸ particularly in the case of patterned interfaces of nano-scale dimension.⁹ By the patterned interface we mean a substrate with its polarity or charge changing on the length-scale comparable to the correlation length between the molecular dipoles in the surrounding water. The most illuminating example is the patchwork of positive and negative charges, each orienting the nearest water molecules along the field of the surface charge. If the distance between different charges is comparable with the size of the surface polarized domain, this spatial arrangement will lead to the orientational frustration of the water molecules at the domain boundaries. One can anticipate that the response of this patchwork of frustrated

domains can potentially be different from the response of a homogeneous dielectric.

The picture of a patchwork of positive and negative charges distributed over a nanometer-scale surface is a good match of the protein-water interface. Proteins possess a large density of surface ionized residues, which provide the free energy required to stabilize a protein in solution. The distribution of positive and negative charges at the surface of a globular protein is nearly uniform, with the average distance between them $\sim 10 \text{ \AA}$.^{10,11} This distance is comparable to the length of a chain of 3–4 water molecules. When two waters at the ends of the chain are pinned by the protein charges, the one-two molecules in the middle must be strongly frustrated orientationally. They resolve frustration by joining one of the domain with the corresponding orientation, thus enhancing the overall orientational order. The hydration shell becomes a patchwork of polarized domains. Their low-temperature statistics is reminiscent of that of another material showing domain formations, the relaxor ferroelectric close to the glass transition.¹²

The key question relevant to this phenomenology is that of the depth of domains’ extension into the bulk, or, in other words, of the number of water molecules involved into the hydration layer with properties distinct from the bulk. Given that $\sim (300 - 500)$ molecules of water can be counted only in the first hydration layer of a typical globular protein, changes in the orientational structure of even the second hydration layer will produce an ensemble of water molecules sufficiently large to be characterized as a separate mesophase (~ 1063 water molecules within a 6 Å shell around lysozyme¹²). Such an ensemble will carry new properties requiring characterization.

Absorption of radiation in the THz domain of frequencies has recently appeared as a novel technique to study

^{a)}Electronic mail: dmitrym@asu.edu

hydration shells of proteins.^{13–22} The absorption of light as the means of probing the solvation structure has a significant advantage compared to broad-band dielectric spectroscopy since the sensitivity of the response is improved by a factor roughly equal to the dielectric constant for the same magnitude of the liquid polarization or by a factor of dielectric constant squared for the same magnitude of the external electric field.²³ The difference is due to the transverse geometry of light absorption compared to the longitudinal geometry of the dielectric experiment.

Further advantage of the THz domain spectroscopy over absorption and dielectric spectroscopy at lower frequencies is that the response of the protein dipole (relaxation time ~ 9 ns for lysozyme²⁴) is dynamically frozen at the THz frequency. While there are a number of vibrational modes in the THz region belonging to the vibrational density of states of a protein,^{25,26} they do not project on the protein dipole to produce sufficient THz absorption.²⁷ The protein is therefore mostly transparent in the THz domain and main absorption comes from water in solution. The challenge is how to separate the hydration shell signal from the dominant bulk absorption.

The analysis of the experimental data in terms of simplistic mixing model,^{13,15,16,28,29} disregarding non-additivity of the interfacial polarization (see discussion after Eq. (5) below), suggested the possibility of an extended hydration shell around a protein, with absorption distinct from the bulk.^{16,28} It was later shown that the change in the absorption of solutions relative to bulk water can be related to the cross-correlation between the protein dipole and the dipoles of water molecules in the hydration shell.²⁷ Specifically, the alteration of the dipolar susceptibility of the solution relative to the bulk is directly proportional to the solute-solvent dipolar correlation function $\chi_{0s}(\omega)$

$$\Delta\chi(\omega) \propto \chi_{0s}(\omega) \quad (1)$$

(see Eq. (11) below for a complete form). When the cross-correlation $\chi_{0s}(\omega)$ is calculated for hydration shells of varying thickness, its length of saturation to the bulk value turns out to be in the range of 20–40 Å^{27,30} from the protein surface. Therefore, deviations from the ideal behavior with increasing solute concentration¹⁶ are likely related to changes of χ_{0s} from its infinite-dilution value.

As expected from the general arguments presented above, simulations have shown that orientational correlations of water dipoles in the hydration layer are significantly different from those in a homogeneous dielectric. The major question, still posed to the field, is whether these differences can be reliably extracted from absorption of THz radiation with its wavelength far exceeding the length-scale of any structural correlations in the solution. Since the wavelength of radiation does not provide spatial resolution, theory is required to interpret the data.

The first question to ask is what would the standard electrostatics predict for the absorption coefficient

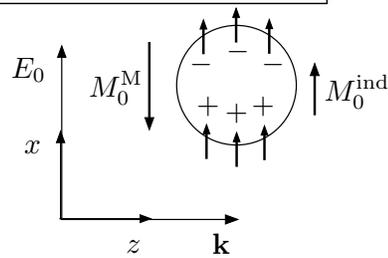


Figure 1. Schematic representation of the polarization of the solute-solvent interface by the uniform field of radiation propagating with the wave-vector \mathbf{k} along the z -axis and polarized along the x -axis. We assume the slab geometry, with \mathbf{E}_0 parallel to the plane of the slab. The short arrows, aligned along the field, indicate dipoles of water polarized by the external field. They create surface charge density with the positive and negative lobes at the opposite sides of the cavity. The surface charge density obtained by solving the standard dielectric boundary value problem⁷ results in the interface dipole M_0^M given by Eq. (2). The dipole M_0^{ind} , induced at the solute by the external field, is the sum of the electronic and nuclear components, Eq. (3). The permanent (nuclear) solute dipole is strongly affected by the radiation frequency and is effectively zero for a protein affected by the electric field of the THz radiation.

of solutions. This problem can in fact be addressed by Maxwell electrostatics of dielectrics, by which we mean the application of the Maxwell boundary conditions to the solution of the Poisson equation.⁷ The problem at hand is illustrated in Fig. 1. The polarization of the interface of a spherical void approximating the hydrated protein integrates to an effective dipole assigned to the void, which is anti-parallel to the external field

$$M_0^M = -\Omega_0 P \frac{3\epsilon_s}{2\epsilon_s + 1}. \quad (2)$$

Here, P is the polarization of the bulk caused by the uniform field of the electromagnetic wave, Ω_0 is the volume of the spherical solute and ϵ_s is the dielectric constant of water; the superscript “M” specifies the solution of the Maxwell boundary value problem. The Maxwell electrostatics thus predicts a decrease in the effective polarity of the solution compared to that of a homogeneous liquid due to the combination of removing polar solvent from the solute volume Ω_0 and an additional polarization of the interfacial dipoles as dictated by Maxwell’s boundary conditions (Fig. 1).

Simulations^{27,31} and absorption measurements for hydrated lysozyme³² and for some amino acids¹⁸ have suggested that the polarization of the interface can significantly deviate from the Maxwell scenario. We indeed show below that the so-called Lorentz void,³³ instead of the Maxwell void (Eq. (2)), provides a better representation of the collective interfacial polarization. However, this conclusion cannot be reached without taking into account the induced dipole of the protein itself.

The negative projection of the interface Maxwell dipole is counterbalanced by a positive induced solute dipole

M_0^{ind} . The latter is composed of the instantaneous (on the THz time-scale) electronic induced dipole M_0^e and the average permanent dipole $\langle M_0 \rangle_E$ established along the field for an ensemble of proteins in solution

$$M_0^{\text{ind}} = M_0^e + \langle M_0 \rangle_E. \quad (3)$$

The permanent dipole of course depends on the frequency of the signal and is expected to be dynamically frozen in the THz domain. The average over an ensemble of randomly oriented permanent dipoles of the proteins in solution then yields a negligible net value of $\langle M_0 \rangle_E$. Nevertheless, dynamical freezing depends on the radiation frequency and should not be blindly assumed. The induced dipole cannot be generally neglected, as we indeed show below. Given the mutual compensation between the anticipated negative interface dipole and the solute induced dipole, a quantitative theory is required to incorporate both components. Here, we construct such a theoretical description based on previous theoretical advances.^{23,27} Our main focus here is on bringing the theory to the level permitting direct applications to experiment.

The permanent protein dipole induced by the external field, $\langle M_0 \rangle_E$ in Eq. (3), is supplied by molecular dynamics (MD) simulations. These results confirm that it is dynamically frozen in the THz frequency window and can be neglected. The main compensation is therefore between the positive induced electronic dipole and the negative interface dipole. We compare the results of our analysis to recent THz absorption data reported for lysozyme solutions.³²

II. FORMALISM

The absorption coefficient for electromagnetic radiation at frequency ω is fully defined in terms of the real and imaginary parts of the dielectric function of solution $\epsilon(\omega) = \epsilon'(\omega) + i\epsilon''(\omega)$

$$\alpha_{\text{abs}}(\omega) = \frac{\omega}{c} \frac{\epsilon''(\omega)}{\sqrt{\epsilon'(\omega)}}. \quad (4)$$

The problem is therefore reduced to calculating $\epsilon(\omega)$ in terms of the complex-valued dielectric function of the solvent $\epsilon_s(\omega)$ (water in our case) and the properties of the solute (protein). A general formalism for $\epsilon(\omega)$ for an ideal solution was developed in Refs. 23 and 27. Here we summarize the main steps of this theory relevant to our simulations and comparison to experiment.

We first consider a fictitious solute, which does not produce any dipolar response and thus can be approximated as a void in a polar liquid. This approximation is relevant to many THz experiments when the solute dipole is too slow and is dynamically frozen. Computer simulations presented below specifically address the question of whether this approximation is applicable in the case of hydrated lysozyme. Independently of the outcome, this limit is the simplest conceptual point of departure for our discussion.

A. Voids in water

A void removes polar material from its volume thus reducing the overall dipole moment of the liquid from its homogeneous magnitude $M_x^{\text{liq}} = PV$ to the value $M_x^{\text{liq}} - N_0\Omega_0P$. Here, we have adopted the geometry of an absorption experiment sketched in Fig. 1 in which light is propagated with the wave-vector along the z -axis of the laboratory frame and its electric field polarized along the x -axis. Further, the solution contains N_0 solutes, each having the volume Ω_0 , $\eta_0 = (N_0\Omega_0)/V$ is the volume fraction.

Removing polarized medium from the void's volume is not sufficient to calculate the polarization of the solution. Any discontinuous interface inside a dielectric is polarized by the field, and there will be an additional interfacial polarization, which requires solving the dielectric boundary value problem.⁷ In the case of a spherical void, the transverse geometry of the absorption experiment leads to the following relation²³

$$M_x = P(V - N_0\Omega_0) - N_0\Omega_0P \frac{\epsilon_s - 1}{2\epsilon_s + 1}. \quad (5)$$

Here, the last summand accounts for an additional polarization of the interface not accounted for by the removal of the polarized liquid.

The first term in Eq. (5) describes the reduction of absorption produced by removing water from the volume of the protein. This is the water term in the two-component analysis often used in the literature.^{13,15,16,28,29} In this approach, the absorption coefficient of the solution is represented as the volume-fraction weighted sum of the protein (p) and water (w) components: $\alpha_{\text{abs}} = (1 - \eta_0)\alpha_w + \eta_0\alpha_p$. This approximation can produce negative protein absorption when applied to analyzing experimental data.¹⁹ This seemingly astonishing result is in fact trivial when compared to Eq. (5): an additional negative contribution, which is assigned to absorption by the protein in the two-component analysis, is caused by polarization of the void in addition to the removal of water (the second summand in Eq. (5)).

The two-component analysis often leads to seemingly reasonable results, but this comes not from the soundness of this approximation, but instead from the combination of the breakdown of the Maxwell dielectric picture for the heterogeneous protein-water interface (see below) and a potential compensation between the interface dipole and the dipole induced at the solute (M_0^{ind} in Eq. (3) and in Fig. 1).²⁷ Deviations between the two-component model and observations have been often assigned to the third component, the hydration water. Given that such deviations, according to Eq. (5), are predicted even within the standard dielectric picture, any conclusions based on such an analysis are unfounded. Numerically, the additional polarization of the interface [second summand in Eq. (5)] is a significant effect, amounting to about a half, at $\epsilon_s \gg 1$, of the polarization loss from the direct expulsion of water from the solute volume.

The appearance of the dielectric constant of bulk water ϵ_s in Eq. (5) is a consequence of using the standard dielectric boundary conditions at the discontinuous surface separating the void from the polar liquid. Hydration water surrounding proteins is strongly altered compared to the bulk and it is a priori not clear if the standard dielectric boundary conditions apply. In fact, we have shown before²⁷ and demonstrate more conclusively here that polar response of the protein-water interface is qualitatively different from that of the standard dielectric interface. In order to account for these differences, we introduce the parameter $\alpha(\omega)$, which allows us to re-write Eq. (5) in the following form

$$M_x(\omega) = P(\omega)(V - N_0\Omega_0) - N_0\Omega_0\alpha(\omega)P(\omega)\frac{\epsilon_s(\omega) - 1}{2\epsilon_s(\omega) + 1}. \quad (6)$$

Here, we have indicated the dependence of all components on the radiation frequency ω . It has to be stressed that, similarly to the dielectric function $\epsilon_s(\omega)$, the parameter $\alpha(\omega) = \alpha'(\omega) + i\alpha''(\omega)$ is a complex-valued function, as will be clear from the following discussion. At $\alpha(\omega) = 1$ one restores the standard Maxwell solution in Eq. (5) and at $\alpha(\omega) = 0$ the interface is not polarized, which we dub here as the Lorentz scenario. Any value of $\alpha(\omega)$ distinct from these two limits is obviously allowed.

From Eq. (6), one can directly calculate the change in the susceptibility of the solution $4\pi\chi(\omega) = \epsilon(\omega) - 1$ relative to the susceptibility of the bulk liquid $4\pi\chi_s(\omega) = \epsilon_s(\omega) - 1$

$$\frac{\Delta\chi_v(\omega)}{\chi_s(\omega)} = -\eta_0 \left[1 + \alpha(\omega)\frac{\epsilon_s(\omega) - 1}{2\epsilon_s(\omega) + 1} \right], \quad (7)$$

where we used the subscript “v” to stress that this susceptibility change is due to the ideal solution of voids. In the case of the Maxwell dielectric, $\alpha(\omega) = 1$, the solution susceptibility decreases linearly with the volume fraction η_0

$$\frac{\Delta\chi_v(\omega)}{\chi_s(\omega)} = -\eta_0 \frac{3\epsilon_s(\omega)}{2\epsilon_s(\omega) + 1}. \quad (8)$$

The slope of the linear scaling with η_0 is equal to the the product of $\epsilon_s(\omega)$ with the cavity field susceptibility of dielectric theories^{34,35}

$$\chi_c^M(\omega) = \frac{3}{2\epsilon_s(\omega) + 1} \quad (9)$$

The susceptibility $\chi_c = E_c/E_0$ is the ratio of the field inside the void (cavity) E_c to the uniform external field E_0 (see below), the superscript “M” indicates the Maxwell solution for this property.^{7,33}

Equation (7) does not accomplish much unless a connection between $\alpha(\omega)$ and parameters accessible by either simulations and/or experiment can be established. It was previously suggested²⁷ that the effective dipole at the surface of a spherical void polarized by the uniform

external field can be alternatively calculated from the response to a dipole placed inside the void. The connection is provided by the relation

$$1 + \alpha(\omega)\frac{\epsilon_s(\omega) - 1}{2\epsilon_s(\omega) + 1} = -\frac{3\epsilon_s(\omega)}{2(\epsilon_s(\omega) - 1)}\frac{\chi_{0s}(\omega)}{\chi_{00}(\omega)}. \quad (10)$$

Equation (7) is therefore re-written in the following form

$$4\pi\Delta\chi_v(\omega) = \frac{3}{2}\eta_0\epsilon_s(\omega)\frac{\chi_{0s}(\omega)}{\chi_{00}(\omega)}. \quad (11)$$

Equation (11) provides a significant advantage for the theory-experiment connection since it directly yields the THz slope (α_{abs} vs η_0), instead of absorption at a given concentration as calculated from simulations of the total dipole of the mixture.^{22,36} The required input is two susceptibility functions, $\chi_{00}(\omega)$ and $\chi_{0s}(\omega)$. They are accessible by computer simulations in the limit of a single solute in the simulation cell, thus bypassing the difficulties of simulating mixtures. The susceptibilities $\chi_{00}(\omega)$ and $\chi_{0s}(\omega)$ arise from, correspondingly, the “self”, $\propto \langle \delta\mathbf{M}_0(t) \cdot \delta\mathbf{M}_0(0) \rangle$, and “cross”, $\propto \langle \delta\mathbf{M}_s(t) \cdot \delta\mathbf{M}_0(0) \rangle$, time correlation functions of the fluctuating dipole moments of the solute, $\mathbf{M}_0(t)$, and solvent, $\mathbf{M}_s(t)$. Being linear response functions, $\chi_{00}(\omega)$ and $\chi_{0s}(\omega)$ represent the permanent dipole moment induced by the external electric field at the solute (subscript “00”) and the dipole moment caused in the hydration layer by the solute dipole (subscript “0s”).

The frequency-dependent susceptibilities are Laplace-Fourier ω -transforms of the corresponding time correlation functions appearing in the dynamic linear response theory.³⁷ One can define the normalized time correlation functions

$$S_{0a}(t) = [\langle \delta\mathbf{M}_0 \cdot \delta\mathbf{M}_a \rangle]^{-1} \langle \delta\mathbf{M}_0(t) \cdot \delta\mathbf{M}_a(0) \rangle, \quad (12)$$

where $\delta\mathbf{M}_a(t) = \mathbf{M}_a(t) - \langle \mathbf{M}_a \rangle$ and $a = 0, s$. These functions are Laplace-Fourier transformed to yield the functions $\tilde{S}_{0a}(\omega)$, which then enter the corresponding linear response functions

$$\chi_{0a}(\omega) = (\beta/3\Omega_0)\langle \delta\mathbf{M}_0 \cdot \delta\mathbf{M}_a \rangle \left[1 + i\omega\tilde{S}_{0a}(\omega) \right]. \quad (13)$$

The fitting procedure using a multi-exponential decay with amplitudes A_i ($\sum_i A_i = 1$) and relaxation times τ_i leads to

$$\tilde{S}_{0a}(\omega) = \sum_i \frac{A_i\tau_i}{1 - i\omega\tau_i} \quad (14)$$

and

$$\chi_{0a}(\omega) \propto \sum_i A_i [1 - i\omega\tau_i]^{-1}. \quad (15)$$

The cross correlation $\propto \langle \delta\mathbf{M}_s(t) \cdot \delta\mathbf{M}_0(0) \rangle$ is typically poorly converged and is a challenge to calculate from simulations. In anticipation of the final results used to

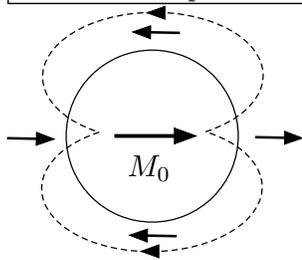


Figure 2. Schematic representation of the solvent dipoles oriented along the electric field lines (dashed) of the solute dipole M_0 . The axial dipoles are aligned parallel to the solute dipole, and the equatorial dipoles are aligned antiparallel to the solute dipole. Since there are more equatorial than axial dipoles, the overall correlation function $\chi_{0s}(\omega)$ is typically negative.

calculate the absorption coefficients, we note that the response function

$$\chi_0^d(\omega) = \chi_{00}(\omega) + \chi_{0s}(\omega) \quad (16)$$

enters the final expressions. This function is based on the time correlation function $\propto \langle \delta \mathbf{M}_0(t) \cdot \delta \mathbf{M}(0) \rangle$ between the solute dipole \mathbf{M}_0 and the entire dipole moment of the sample \mathbf{M} . This correlation function is typically a faster converging property.

Since $\chi_{00}(\omega)$ is always positive, Eq. (11) clearly shows that the sign of slope of $\Delta\chi_v$ vs η_0 is determined by the sign of the cross, solute-solvent dipolar correlations. The solute dipole and the dipoles of the solvent molecules are expected to be anti-correlated, thus resulting in a usually observed^{13,14,32} negative THz slope. The reason for such negative cross correlations is illustrated in Fig. 2, which shows the dipoles of the solvent arranged around a solute dipole. While axial solvent dipoles tend to orient in parallel, the equatorial dipoles tend to orient antiparallel. Since there are always more equatorial dipoles than axial dipoles, the overall cross-correlation is negative. This picture is, however, based on the assumption that interfacial dipoles can freely change their orientations along the field lines of the solute dipole. Hydration shells of proteins are densely packed^{38,39} and orientationally frustrated,^{6,12} thus potentially affecting the ability of hydration waters to polarize along the field of the solute. While most solutes still show a negative slope of $\Delta\chi_v$ vs η_0 , solutions of some amino acids (glycine and serine¹⁸) produce a positive THz slope. The latter result might be, at least partially, related to the compensating positive dipole moment induced at the solute (Eq. (3)), which we consider below.

In the analysis below, we use the input for the frequency-dependent dielectric function of water from experiment⁴⁰ and model the electronic induced dipole of the protein by its refractive index. The force fields typically used in MD simulations miss polarizability of at least one of these two components. It is not a priori clear if the force fields typically used in condensed phase simulations are suitable for modeling absorption. To illustrate

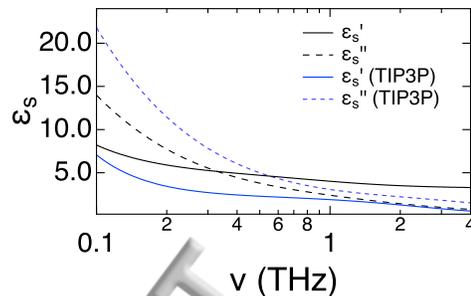


Figure 3. Real, $\epsilon'_s(\nu)$, and imaginary, $\epsilon''_s(\nu)$, dielectric functions of water⁴⁰ and of TIP3P force-field model of water in the THz range of frequencies. MD simulations of TIP3P water at 300 K performed for this study were fitted with two Debye processes with dielectric increments $\Delta\epsilon_1 = 101.0$ and $\Delta\epsilon_2 = 2.2$ and relaxation times $\tau_1 = 7.05$ ps and $\tau_2 = 0.07$ ps.

this point, Fig. 3 compares the real and imaginary parts of $\epsilon_s(\nu)$ ($\omega = 2\pi\nu$) for TIP3P water (dashed lines) with experiment⁴⁰ (solid lines). The differences between experiment and TIP3P water are significant. In order to establish whether these differences affect absorption of radiation, one needs a formalism operating in terms of solvent parameters accessible by both experiment and simulations. Equation (11) presents this opportunity, and we discuss this analysis below. It turns out that the use of the frequency-dependent dielectric function of TIP3P water significantly deteriorates the agreement between calculations and experiment within the analytical (Lorentz void) framework, but does not strongly affect the results directly obtained from MD simulations.

B. Solute dipole

The solute dipole can include both the permanent and electronic components. As is typically done in mean-field theories of polarizable dielectrics,^{41,42} it is convenient to combine them in one polar density parameter²⁷

$$y_0(\omega) = (4\pi/3)\chi_{00}(\omega) + y_e, \quad (17)$$

where $y_e = (4\pi/3)(\alpha_0/\Omega_0)$ and α_0 is the electronic polarizability of the solute. Clausius-Mossotti equation can be used to estimate y_e

$$y_e = \frac{1}{\eta_0^p} \frac{n_p^2 - 1}{n_p^2 + 2}. \quad (18)$$

Here, n_p is the refractive index of the protein and η_0^p is the packing fraction of proteins in the powder used to measure n_p . We will assume $\eta_0^p \simeq 1$ in our estimates below, which is also the assumption made by Onsager for molecular polarizability.⁴³ Corrections can be introduced if this parameter is known independently.

The frequency-dependent dipolar response of the solute, adds to the solution susceptibility, which is now defined as

$$\Delta\chi = \Delta\chi_v + \Delta\chi_0. \quad (19)$$

Here, the first summand, due to the void, is given by Eq. (11). The second terms, the change in the susceptibility due to dipolar response of the solutes, is given by the following equation²⁷

$$\frac{4\pi\Delta\chi_0}{\eta_0 y_0} = 3\epsilon_s\chi_c + J_0(\eta_0) [3\epsilon_s\chi_c - \epsilon_s - 2], \quad (20)$$

where here and in Eq. (19) the dependence on frequency was dropped for brevity. The frequency-dependent response function $\chi_c(\omega)$ is

$$\chi_c(\omega) = \frac{\chi_0^d(\omega)}{\chi_{00}(\omega)} = 1 + \frac{\chi_{0s}(\omega)}{\chi_{00}(\omega)}, \quad (21)$$

where $\chi_0^d(\omega)$ given by Eq. (16) is the total susceptibility of the solute dipole in response to the uniform polarizing external field.^{44,45} The physical meaning of χ_0^d/χ_{00} is that it is equal²⁷ to the ratio, E_c/E_0 , of the cavity field E_c experienced by the solute dipole from the polarized solvent to the field of the external charges E_0 (the result of Maxwell electrostatics is given by Eq. (9)). Although a direct measurement of the cavity field is hardly possible, it enters a number of spectroscopic observables.⁴⁶ Correspondingly, the response function $\chi_c(\omega)$ represents the ratio of the cavity field to the external field at a given frequency.

The first summand in Eq. (20) is the infinite-dilution contribution of the solute dipole to the solution susceptibility. The second summand, which includes the function $J_0(\eta_0)$, represents the non-ideal effect of polarizing the void by the permanent dipoles of other solutes in the solution. Therefore, $J_0(0) = 0$, and it is given for an arbitrary η_0 by an equation involving the density structure factor of the solutes, $S_0(k) = N_0^{-1} \sum_{i,j} \exp[i\mathbf{k} \cdot \mathbf{r}_{ij}]$, where \mathbf{r}_{ij} is the vector connecting two solutes in the solution and \mathbf{k} is the wavevector. The density structure factor determines mutual correlations of the positions of the solutes and $J_0(\eta_0)$ is given as²³

$$J_0(\eta_0) = \frac{3}{\pi} \int_0^\infty [j_1(x/2)]^2 [S_0(x) - 1] dx. \quad (22)$$

Here, $j_1(x)$ is the spherical Bessel function of the first order⁴⁷ and integration is performed over the dimensionless variable $x = k\sigma_0$, where σ_0 is the solute diameter. While $S_0(k)$ is experimentally available from small-angle scattering, we estimated it here based on Percus-Yevick solution for a fluid of hard spheres.⁴⁸ These estimates (see $J_0(\eta_0)$ tabulated in Ref. 23) show that $J_0(\eta_0)$ can be dropped from the final equations at the low volume fraction $\eta_0 < 0.1$. When $J_0 = 0$, one can apply Maxwell electrostatics ($\alpha = 1$) as a consistency check of Eq. (20). One

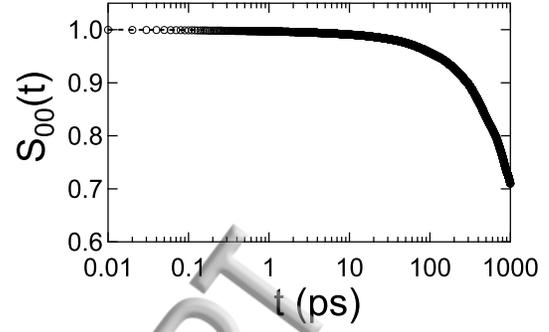


Figure 4. Time correlation function $S_{00}(t)$ [Eq. (12)] calculated from MD simulations (points). The simulation results are fitted by three decaying exponents (dashed line): $S_{00}(t) = \sum_{i=1}^3 A_i \exp[-t/\tau_i]$: $A_1 = 0.004$, $A_2 = 0.006$, $A_3 = 0.990$, $\tau_1 = 0.77$ ps, $\tau_2 = 24.6$ ps, $\tau_3 = 2988$ ps.

then gets Eq. (9) for χ_0^d/χ_{00} and for the static ($\omega = 0$) response

$$\Delta\chi_0 = N_0 \frac{\beta \langle \delta \mathbf{M}_0^2 \rangle}{3V} \frac{3\epsilon_s}{2\epsilon_s + 1}. \quad (23)$$

This is the standard linear-response result³⁴ for an ideal solution of impurities experiencing the cavity field $E_c = 3\epsilon_s E / (2\epsilon_s + 1)$ from the surrounding dielectric (cf. to Eq. (9), E is the Maxwell field in the bulk).

III. SIMULATION RESULTS

All-atom MD simulations were performed using the initial crystallographic structure of lysozyme (PDB entry 1AIK) solvated with TIP3P water. The simulation cell consisted of a total of 87050 atoms, with 28361 water molecules. The total charge of the lysozyme protein was -7 e. The constant pressure temperature equilibration simulations were done using the Langevin temperature-pressure control with the damping coefficient of 5 ps^{-1} , a piston pressure of 1 atm, a piston decay time of 50 fs, a piston oscillation period of 200 fs, and at temperature of 300 K. A cutoff radius of 12 Å and full electrostatics using the particle mesh Ewald technique at every simulation step were employed. NAMD 2.10⁴⁹ with the CHARMM27 force field was used to produce the MD trajectories. An initial optimization of the simulation cell was performed by conjugate gradient minimization for 2000 steps, followed by a 5 ns NPT equilibration simulation. Proceeding from equilibration run, production NVT simulations were carried out for 10 ns.

No counterions were used in the simulation setup. The simulation cell therefore carries a total non-zero charge equal to the charge of the protein. According to the standard practice of employing Ewald sums for the long-range electrostatics, this setup is equivalent to placing

a uniform compensating background charge in the periodic Wigner lattice.⁵⁰ This background charge will affect the free energy of a charge placed within the simulation cell,^{50,51} but will cancel out from the deviation of the dipole moment $\delta\mathbf{M}_a$ from its average value. Only $\delta\mathbf{M}_a$ enter the calculation of susceptibilities in Eq. (13) and the background charge does not affect our results. In terms of comparison of our results to experiment, the relaxation time of the electrolyte buffer used in the typical experimental setup exceeds the picosecond time-scale relevant to THz experiment.⁴⁸ One can therefore assume that the electrolyte is dynamically frozen in the THz experiment.

The main reason for a relatively short length of the simulation trajectory used here is that the simulations (both the equilibration NPT and the production NVT) were performed with a relatively short timestep of 0.5 fs and flexible hydrogen bonds of the protein. The trajectory was saved every 10 fs. This focus on the short-time dynamics has allowed us to calculate the time correlation functions with a 10 fs time resolution required for the modeling of the THz response. Exponential fits of the correlation function were done on the time window of 1 ns. Those were Laplace-Fourier transformed to obtain Eq. (14) for the correlation functions and Eq. (15) for the response functions.

IV. COMPARISON TO EXPERIMENT

A significant result of our present and previous²⁷ computer simulations is that $(4\pi/3)\chi_{00}$ can be neglected in Eq. (17) relative to y_e in the THz domain. The nuclear motions of the protein do not produce sufficient fluctuations of its dipole and absorption due to the protein permanent dipole is negligible. Only the electronic dipole induced at the protein makes a non-negligible contribution. Note that, in contrast with the previous simulations,²⁷ the current setup allowed vibrations of protons in the protein, along with a significantly smaller integration step and more frequent savings. Nevertheless, $S_{00}(t)$ is nearly flat at $t \simeq 0.1 - 1$ ps (Fig. 4) and, correspondingly, $\chi_{00}(\omega)$ is very small for $\nu = \omega/(2\pi)$ in the THz domain of frequencies.

With this simplification and after dropping the term proportional to J_0 from Eq. (20), one arrives at a simple relation

$$4\pi\Delta\chi(\omega) = -\frac{3}{2}\eta_0\epsilon_s(\omega) \left[1 - \chi_c(\omega) \frac{3n_p^2}{n_p^2 + 2} \right], \quad (24)$$

where the Clausius-Mossotti relation between the protein refractive index n_p and y_e was applied [$\eta_0^p = 1$ in Eq. (18)]. In the limit of the Maxwell electrostatics (superscript “M”) one obtains

$$4\pi\Delta\chi^M(\omega) = -\frac{3}{2}\eta_0\epsilon_s(\omega) \left[1 - \frac{9n_p^2}{(2\epsilon_s(\omega) + 1)(n_p^2 + 2)} \right]. \quad (25)$$

At $n_p = 1$, this relation converts into Eq. (8) for the susceptibility of a Maxwell void.

One can additionally consider the limit in which no surface charges, originating from introducing a dividing surface into the dielectric,⁷ are created at the interface of a void (no “+” and “-” at the void’s surface in Fig. 1). This limit corresponds to $\alpha(\omega) = 0$ in Eq. (7) and the assumption that the only result of creating the void is the expulsion of the polarized water from the void’s volume. Such a void is known in the theory of dielectrics as the virtual, or Lorentz, cavity.³⁴ We have dubbed this limit as the “Lorentz scenario”.³³ The ratio χ_0^d/χ_{00} becomes equal to the Lorentz cavity field,²⁷ $E_c/E_0 = (\epsilon_s(\omega) + 2)/(3\epsilon_s(\omega))$, and one obtains from Eq. (24) (superscript “L” is for the Lorentz scenario)

$$4\pi\Delta\chi^L(\omega) = -\frac{3}{2}\eta_0\epsilon_s(\omega) \left[1 - \frac{\epsilon_s(\omega) + 2}{\epsilon_s(\omega)} \frac{n_p^2}{n_p^2 + 2} \right]. \quad (26)$$

Not surprisingly, there is no change to the solvent response when the virtual void is filled with the same material as the solvent, $\Delta\chi^L = 0$ at $\epsilon_s = n_p^2$. Note that the Lorentz cavity field is used in deriving the Clausius-Mossotti equation, and that fact leads to the complete cancellation of the interface and electronically induced solute dipoles at $\epsilon_s = n_p^2$. This does not happen in the Maxwell case in Eq. (25) since the cavity field of dielectric theories,³⁴ instead of the Lorentz field, is used in the Maxwell electrostatics.

In application to analyzing experimental results, the complex-valued dielectric function of bulk water was taken from recent measurements⁴⁰ covering the range of frequencies 5.9 – 1120 GHz. The experimental data³² for the absorption coefficient of lysozyme were obtained by averaging absorption over the interval of frequencies 0.38 – 0.92 THz. There is therefore a sufficient frequency overlap between the data for bulk water⁴⁰ and solutions.³² The refractive index of the protein, $n_p \simeq 1.55$, is from Ref. 52.

Figure 5 presents the calculations of the change in the absorption coefficient of the solution with increasing the volume fraction of the solute according to Eq. (4). Specifically, we normalize the change in absorption by the absorption of bulk water

$$\Delta\bar{\alpha}_{\text{abs}}(\omega) = \alpha_{\text{abs}}(\omega)/\alpha_{\text{abs}}^s(\omega) - 1, \quad (27)$$

where the absorption of water $\alpha_{\text{abs}}^s(\omega)$ is calculated from Eq. (4) with the water dielectric function $\epsilon_s(\omega)$ used in place of the solution dielectric function $\epsilon(\omega)$. The solution dielectric function is calculated as $\epsilon(\omega) = \epsilon_s(\omega) + 4\pi\Delta\chi(\omega)$, with three scenarios presented by Eqs. (24)–(26) for $4\pi\Delta\chi(\omega)$.

Both the Maxwell and Lorentz voids are sensitive to the choice of the protein refractive index. The dashed lines in Fig. 5 refer to the calculations neglecting the induced dipole of the protein ($M_0^{\text{ind}} = 0$ in Eq. (3)), which is achieved by putting $n_p = 1$. There is, however, no physical reason to adopt this assumption and

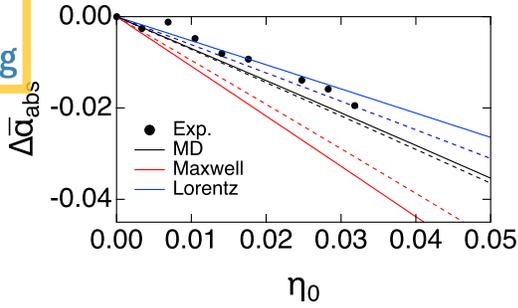


Figure 5. Relative change of the absorption coefficient of solution $\Delta\bar{\alpha}_{\text{abs}}(\omega)$ [Eq. (27)] vs the solute volume fraction η_0 . Calculations are based on $\chi_c(\omega)$ from MD (black, Eq. (24)), and analytical equations based on Maxwell (red, Eq. (25)) and Lorentz (blue, Eq. (26)) scenarios. The points are experimental data³² and the calculations are made at $\nu = 0.65$ THz. The solid lines refer to $n_p = 1.55$ ⁵² and the dashed lines correspond to the assumption of a non-polarizable protein, $n_p = 1$. Experimental⁴⁰ $\epsilon_s(\omega)$ is used in all calculations. The cavity field susceptibility in Eq. (21) is based on exponential fits of the simulation data: $\chi_0^d(\omega) = \sum_{i=1}^2 a_i(1 - i\omega\tau_i)^{-1}$ with $a_1 = 0.018$, $a_2 = 23.806$, $\tau_1 = 1.74$ ps, $\tau_2 = 2641$ ps and $\chi_{00}(\omega) = \sum_{i=1}^3 b_i(1 - i\omega\tau_i)^{-1}$ with $b_1 = 0.086$, $b_2 = 0.140$, $b_3 = 23.575$, $\tau_1 = 0.77$ ps, $\tau_2 = 24.6$ ps, $\tau_3 = 2988$ ps. The susceptibilities $\chi_{0a}(\omega)$ are unitless per Eqs. (13) and (16).

this limit is shown here as a mere warning that a “good agreement” with experiment can follow from unphysical approximations (the Lorentz scenario is in perfect agreement with experimental data in this case, blue dashed line). It is also clear that MD simulations of $\chi_c(\omega)$ provide a good account of the cavity field dynamics and are consistent with experiment. We additionally show in Fig. 6 the comparison of the calculations employing the experimental⁴⁰ dielectric function $\epsilon_s(\omega)$ (solid lines) and the same function obtained from MD simulations of TIP3P water (dashed lines). There is little sensitivity to the choice of the water model when the cavity response function $\chi_c(\omega)$ is calculated from MD simulations. In contrast, the Maxwell and Lorentz equations for the cavity field are sensitive to the choice of the frequency-dependent dielectric function. The agreement with experiment is much worse when $\epsilon_s(\omega)$ for TIP3P water is used Eqs. (25) and (26). This is not surprising given large discrepancies in the dielectric properties between the experimental and TIP3P water shown in Fig. 3.

V. CONCLUSIONS

Absorption of radiation provides access to the orientational structure of the hydration shell in terms of the dipolar solute-solvent (cross) correlation function [Eqs. (1) and (11)]. The alternative representation is in terms of the cavity field susceptibility [Eq. (24)]. Either of these functions is the only information about the solution ac-

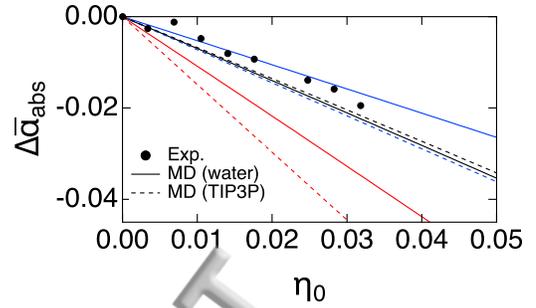


Figure 6. Relative change of the absorption coefficient of solution $\Delta\bar{\alpha}_{\text{abs}}(\omega)$ [Eq. (27)] vs the solute volume fraction η_0 . Calculations are based on $\chi_c(\omega)$ from MD (black, Eq. (24)), Maxwell (red, Eq. (25)) and Lorentz (blue, Eq. (26)) scenarios. The points are experimental data³² and the calculations are made at $\nu = 0.65$ THz. The solid lines refer to $n_p = 1.55$ ⁵² and the experimental results for and the dielectric function of water.⁴⁰ The dashed lines refer to the same conditions, but $\epsilon_s(\omega)$ from MD simulations of TIP3P water (this study).

cessible from the absorption experiment. It does not allow any spatial resolution of the interface and requires theoretical interpretation to draw the connection between these susceptibilities and the interfacial structure. The analytical model^{23,27} presented here calculates the slope of the absorption coefficient vs the solute volume fraction in terms of the cavity susceptibility at the radiation frequency. Simulations of a single lysozyme in solution performed here are capable of producing this function and lead to a good agreement with experiment³² (Fig. 5).

Nuclear motions of the protein do not contribute to the signal, and protein itself is effectively transparent in the THz domain. The only contribution to protein’s absorption comes from the electronic induced dipole and the corresponding refractive index. Therefore, it is predominantly the protein-water interface that determines the change in the THz absorption by solutions relative to the bulk.

The structure of water can be disrupted to a different extent depending on the solute. The extent of disruption, and the resulting structure of the hydration shell, affect the decay of the solute-solvent correlations into the bulk.²⁷ The decay length can be in the range of 20 – 40 Å.^{27,30} The extended hydration shell anticipated in the past¹⁶ in fact implies extended dipolar cross-correlations. Those are electrostatic correlations, for which long range of decay is anticipated. Absorption by ideal solutions provides no direct input into either the structure of the hydration shell (spatial arrangement of the nuclei) or its extent into the bulk.

In the absence of direct simulations, two analytical limits, corresponding to the Maxwell and Lorentz voids, can be derived [Eqs. (25) and (26)]. We find, in agreement with previous calculations,³⁰ that the Maxwell void gives a poor representation of the data, while the Lorentz void

give a better account of experiment. One still has to realize that the appearance of the Lorentz limit for the total polarization is not consistent with the standard electrostatics, but is an emergent consequence of the orientational and density restructuring of the protein hydration layer compared to bulk water. It is therefore the deviation from the Maxwell limit that should be considered as the specific effect of the protein hydration shell. The Lorentz scenario is an effective-medium representation of the complex structure of the protein-water interface, which requires modification of the standard dielectric boundary-value problem.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation (CHE-1464810) and through XSEDE resources (TG-MCB080116N).

REFERENCES

- ¹D. Chandler, Phys. Rev. E **48**, 2898 (1993).
- ²D. Chandler, Nature **437**, 640 (2005).
- ³S. Sarupria and S. Garde, Phys. Rev. Lett. **103**, 037803 (2009).
- ⁴I. Brovchenko and A. Oleinikova, J. Phys. Chem. C **111**, 15716 (2007).
- ⁵I. Brovchenko and A. Oleinikova, *Interfacial and confined water* (Elsevier, Amsterdam, 2008).
- ⁶A. D. Friesen and D. V. Matyushov, J. Phys. Chem. Lett. **3**, 3685 (2012).
- ⁷J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1999).
- ⁸D. V. Matyushov, J. Chem. Phys. **140**, 224506 (2014).
- ⁹N. Giovambattista, C. F. Lopez, P. J. Rossky, and P. G. Debenedetti, Proc. Natl. Acad. Sci. **105**, 2274 (2008).
- ¹⁰D. J. Barlow and J. M. Thornton, Biopolymers **25**, 1717 (1986).
- ¹¹S. Takashima, J. Non-Crystal. Solids **305**, 303 (2002).
- ¹²D. R. Martin and D. V. Matyushov, J. Phys. Chem. Lett. **6**, 407 (2015).
- ¹³C. Zhang and S. M. Durbin, J. Phys. Chem. B **110**, 23607 (2006).
- ¹⁴J. Xu, K. W. Plaxco, and S. J. Allen, Prot. Science **15**, 1175 (2006).
- ¹⁵U. Heugen, G. Schwaab, E. Bründermann, M. Heyden, X. Yu, D. M. Leitner, and M. Havenith, Proc. Natl. Acad. Sci. U.S.A. **103**, 12301 (2006).
- ¹⁶S. Ebbinghaus, S. J. Kim, M. Heyden, X. Yu, U. Heugen, M. Gruebele, D. M. Leitner, and M. Havenith, Proc. Natl. Acad. Sci. **104**, 20749 (2007).
- ¹⁷B. Born, S. J. Kim, S. Ebbinghaus, M. Gruebele, and M. Havenith, Faraday Disc. **141**, 161 (2009).
- ¹⁸G. Niehues, M. Heyden, D. A. Schmidt, and M. Havenith, Faraday Discuss. Chem. Soc. **150**, 193 (2011).
- ¹⁹N. Q. Vinh, S. J. Allen, and K. W. Plaxco, J. Am. Chem. Soc. **133**, 8942 (2011).
- ²⁰G. Acbas, K. A. Niessen, E. H. Snell, and A. G. Markelz, Nat. Comm. **5**, 3076 (2014).
- ²¹N. Yamamoto, K. Ohta, A. Tamura, and K. Tominaga, J. Phys. Chem. B **120**, 4743 (2016).
- ²²J. M. Moix, J. E. Parker, and I. Echchgadda, J. Phys. Chem. B **121**, 3024 (2017).
- ²³D. V. Matyushov, Phys. Rev. E **81**, 021914 (2010).
- ²⁴D. R. Martin and D. V. Matyushov, J. Chem. Phys. **141**, 22D501 (2014).
- ²⁵X. Yu, J. Park, and D. M. Leitner, J. Phys. Chem. B **107**, 12820 (2003).
- ²⁶A. Lerbret, A. Hédoux, B. Annighöfer, and M.-C. Bellissent-Funel, Proteins: Structure, Function, and Bioinformatics **81**, 326 (2012).
- ²⁷D. V. Matyushov, J. Chem. Phys. **136**, 085102 (2012).
- ²⁸T. Q. Luong, P. K. Verma, R. K. Mitra, and M. Havenith, Biophys. J. **101**, 925 (2011).
- ²⁹O. Sushko, R. Dubrovka, and R. S. Donnan, J. Chem. Phys. **142**, 055101 (2015).
- ³⁰M. Heyden, D. J. Tobias, and D. V. Matyushov, J. Chem. Phys. **137**, 235103 (2012).
- ³¹D. R. Martin, A. D. Friesen, and D. V. Matyushov, J. Chem. Phys. **135**, 084514 (2011).
- ³²F. Novelli, S. Ostovar Pour, J. Tollerud, A. Roozbeh, D. R. T. Appadoo, E. W. Blanch, and J. A. Davis, J. Phys. Chem. B **121**, 4810 (2017).
- ³³D. R. Martin and D. V. Matyushov, Europhys. Lett. **82**, 16003 (2008).
- ³⁴C. J. F. Böttcher, *Theory of Electric Polarization*, Vol. 1 (Elsevier, Amsterdam, 1973).
- ³⁵H. Fröhlich, *Theory of dielectrics* (Oxford University Press, Oxford, 1958).
- ³⁶M. Heyden, E. Bründermann, U. Heugen, D. M. Leitner, and M. Havenith, J. Am. Chem. Soc. **130**, 5773 (2008).
- ³⁷R. Kubo, Rep. Prog. Phys. **29**, 255 (1966).
- ³⁸M. Gerstein and C. Chothia, Proc. Natl. Acad. Sci. USA **93**, 10167 (1996).
- ³⁹D. I. Svergun, S. Richard, M. H. J. Koch, Z. Sayers, S. Kuprin, and G. Zaccai, Proc. Natl. Acad. Sci. **95**, 2267 (1998).
- ⁴⁰N. Q. Vinh, M. S. Sherwin, S. J. Allen, D. K. George, A. J. Rahmani, and K. W. Plaxco, J. Chem. Phys. **142**, 164502 (2015).
- ⁴¹M. S. Wertheim, Annu. Rev. Phys. Chem. **30**, 471 (1979).
- ⁴²G. Stell, G. N. Patey, and J. S. Høye, Adv. Chem. Phys. **48**, 183 (1981).
- ⁴³L. Onsager, J. Am. Chem. Soc. **58**, 1486 (1936).
- ⁴⁴H. A. Stern and S. E. Feller, J. Chem. Phys. **118**, 3401 (2003).
- ⁴⁵A. D. Friesen and D. V. Matyushov, Chem. Phys. Lett. **511**, 256 (2011).

- ⁴⁶D. Topytgin, J. Fluoresc. **13**, 201 (2003).
- ⁴⁷M. Abramowitz and I. A. Stegun, eds., *Handbook of Mathematical Functions* (Dover, New York, 1972).
- ⁴⁸J.-P. Hansen and I. R. McDonald, *Theory of simple liquids*, 4th ed. (Academic Press, Amsterdam, 2013).
- ⁴⁹J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten, J. Comput. Chem. **26**, 1781 (2005).
- ⁵⁰G. Hummer, L. R. Pratt, and A. E. García, J. Chem. Phys. **107**, 9275 (1997).
- ⁵¹P. H. Hünenberger and J. A. McCammon, J. Chem. Phys. **110**, 1856 (1999).
- ⁵²B. Cervelle, F. Cesbron, and J. Berthou, Acta Cryst. **A30**, 645 (1974).

ACCEPTED MANUSCRIPT

