

Capturing the Effects of Explicit Waters in Implicit Electrostatics Modeling: Qualitative Justification of Gaussian-Based Dielectric Models in DelPhi

Arghya Chakravorty,* Shailesh Panday, Swagata Pahari, Shan Zhao, and Emil Alexov*



Cite This: *J. Chem. Inf. Model.* 2020, 60, 2229–2246



Read Online

ACCESS |



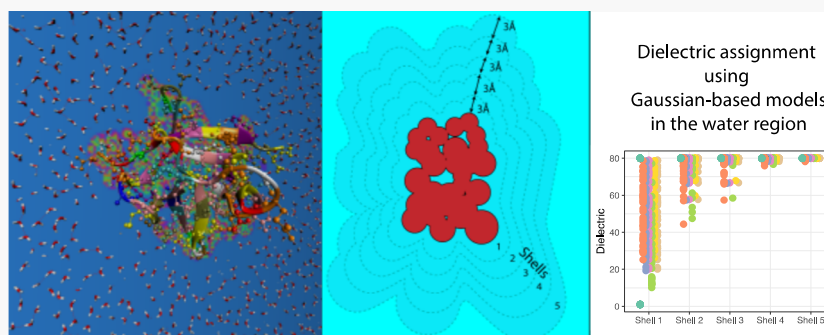
Metrics & More



Article Recommendations



Supporting Information



ABSTRACT: Our group has implemented a smooth Gaussian-based dielectric function in DelPhi (*J. Chem. Theory Comput.* 2013, 9 (4), 2126–2136) which models the solute as an object with inhomogeneous dielectric permittivity and provides a smooth transition of dielectric permittivity from surface-bound water to bulk solvent. Although it is well-understood that the protein hydrophobic core is less polarizable than the hydrophilic protein surface, less attention is paid to the polarizability of water molecules inside the solute and on its surface. Here, we apply explicit water simulations to study the behavior of water molecules buried inside a protein and on the surface of that protein and contrast it with the behavior of the bulk water. We selected a protein that is experimentally shown to have five cavities, most of which are occupied by water molecules. We demonstrate through molecular dynamics (MD) simulations that the behavior of water in the cavity is drastically different from that in the bulk. These observations were made by comparing the mean residence times, dipole orientation relaxation times, and average dipole moment fluctuations. We also show that the bulk region has a nonuniform distribution of these tempo-spatial properties. From the perspective of continuum electrostatics, we argue that the dielectric “constant” in water-filled cavities of proteins and the space close to the molecular surface should differ from that assigned to the bulk water. This provides support for the Gaussian-based smooth dielectric model for solving electrostatics in the Poisson–Boltzmann equation framework. Furthermore, we demonstrate that using a well-parametrized Gaussian-based model with a single energy-minimized configuration of a protein can also reproduce its ensemble-averaged polar solvation energy. Thus, we argue that the Gaussian-based smooth dielectric model not only captures accurate physics but also provides an efficient way of computing ensemble-averaged quantities.

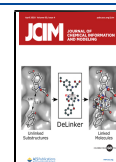
1. INTRODUCTION

Electrostatic interactions and energies are calculated in molecular biology either via explicit or implicit models. Although explicit models provide a more detailed picture of the dynamics of the system and account directly for rearrangement of existing dipoles (and charges) being modeled during the process, they are too time-consuming for use in investigating a large number of cases. In contrast, continuum models offer greater speed and simplicity but overlook many microscopic effects by virtue of macroscopic averaging. For example, the traditional two-dielectric model of an implicit solvent setup divides the total volume into two phases: solute and solvent. It assigns a low dielectric constant to the solute and a high dielectric constant to the solvent, concomitant with their

polarizability. However, proteins are inhomogeneous objects, with an inhomogeneous distribution of polar, charged, and nonpolar residues. Furthermore, the packing in the hydrophobic core is much more dense than that in the protein surface. This motivated us to introduce a new dielectric distribution function into the popular Poisson–Boltzmann equation (PBE) solver into DelPhi. This is a smooth Gaussian-based dielectric function

Received: February 11, 2020

Published: March 10, 2020



ACS Publications

© 2020 American Chemical Society

2229

<https://dx.doi.org/10.1021/acs.jcim.0c00151>
J. Chem. Inf. Model. 2020, 60, 2229–2246

that treats the entirety of the solvated macromolecular volume as an object with an inhomogeneous dielectric distribution, assigning lower dielectric values to the highly packed nonpolar hydrophobic core and relatively larger values to the loosely packed hydrophilic surface.^{1,2} A second and equally important consideration that inspired the development of the Gaussian-based approach in DelPhi was the treatment of the solvent or the water phase. Here, we focus on the water phase with the goal of showing that the Gaussian-based approach can capture the effects of solvation as seen in solvated protein systems with explicit water.

Water, by virtue of its abundance and active participation in biomolecular phenomena, has become the most well-studied solvent. With roles ranging in specificity, there is a vast variety of scenarios in which water interacts with biomolecules and modulates their thermodynamic, kinetic, and hydrodynamic properties.³ Despite this versatility, some features of the aqueous medium in a solvated biomolecular system are common and of biological significance. These features centrally pertain to the alterations of water's thermodynamic and kinetic properties, the effects of which are both microscopic and macroscopic.^{3,4} Microscopically, water's structural order is perturbed/altered in the presence of biomolecules in that it leads to the breakage of interwater hydrogen bonds to create space for itself which then prods the proximal or the "bound" water molecules (typically referred to as biological water^{3,5–7}) to adopt new hydrogen bond networks. The free energy change associated with this process arises from the change in entropy and the change in the nonpolar component of enthalpy. At the same time, the biomolecular charges (leading typically to a large dipole moment) induce polarization effects in water. As a result, a reaction field in the aqueous medium is generated. The interaction of the biomolecule with this field accounts for the polar component of enthalpy and is largely influenced by the dielectric value of water. Collectively, these effects result in an increased residence time and decreased diffusion rate of biological water, by affecting both their temporal and spatial properties.^{8–12} An important addition to this category of water is that found in the interstitial cavities of the proteins. Cavity-bound waters (or any internal water) play a significant role in stabilizing protein folds.¹³ They can affect the pK_a values of ionizable groups lining the cavities,¹⁴ aid in proton transport through the core of a protein,^{15–18} and influence the binding affinity of specific ligands, small molecules, and other proteins by mediating interactions between the protein and the partner.^{13,19,20} They also coordinate with metals or ions in active sites of several enzymes.²¹ This functionally important localization of cavity waters is equally conducive to altering their tempo-spatial properties and thus their dielectric response.

An implicit solvent model, on account of its continuum depiction of the solvent region, neglects some of the local microscopic effects associated with it. By integrating out the contribution of the solvent degrees of freedom, it overlooks the different relaxation and solvation dynamics of the biological and bulk water along with their altered structural order.^{11,22} It also discards the multimodal distribution of the re-orientation response of biological water.²³ Moreover, the definition of a solute–solvent surface (e.g., solvent accessible surface, solvent-excluded surface, and van der Waals surface) also affects the dielectric assignment.²⁴ The choice of a surface (and its implementation through an algorithm) can also affect the treatment of internal cavities, which has the geometric and chemical properties to host solvent molecules. All these factors

can lead to misrepresentation of the naturally occurring local polarization effects which are significant to the solvation energetics. In other words, the traditional continuum model does not succeed in capturing the distinct dielectric properties of the bulk and biological water which has been observed from various NMR and dielectric relaxation method studies (see ref 7 for a list of relevant references).

To account for the nonuniform dielectric response of the solute and water media, several heterogeneous dielectric models for the implicit solvent framework have been developed.^{2,25–28} One such model is the Gaussian-based smooth dielectric model.^{1,2,29} Its simplistic and physically appealing design is based on the idea that loosely packed regions (typically made of charged polar residues and bound or internal waters) of the biomolecule exhibit more conformational flexibility than tightly packed ones (typically nonpolar hydrophobic residues). This should be conducive to a better ability to re-orient, resulting in a higher dielectric permittivity. It assigns a dielectric value at a point based on its proximity to any atom of the solute. A probability function, of an unnormalized Gaussian form, is assigned to each atom "i" of the solute (as shown in eq 1).

$$\rho_i(\vec{r}) = \exp\left(-\frac{(\vec{r} - \vec{r}_i)^2}{\sigma^2 R_i^2}\right) \quad (1)$$

The dielectric value is taken to be inversely proportional to this probability. Farther from any solute atom, the probability function yields a lower value and a higher dielectric value and vice versa. This is accomplished via the mathematical framework presented in eq 2, which uses the collective probability function of the solute atoms and a lower and upper bound of dielectric values (ϵ_{ref} and $\epsilon_{\text{solvent}}$, respectively) to deliver a space-dependent continuous dielectric distribution. This model is implemented in the popular PBE solver package DelPhi.³⁰

$$\begin{aligned} \rho(\vec{r}) &= \prod_i (1 - \rho_i(\vec{r}))\epsilon(\vec{r}) \\ &= \epsilon_{\text{ref}}\rho(\vec{r}) + \epsilon_{\text{solvent}}(1 - \rho(\vec{r})) \end{aligned} \quad (2)$$

This design principle not only allows different regions in the solute's structure to feature diverse dielectric properties but also diversifies the dielectric assignment in the solvent region. Regions farther away from the solute approach bulk-like dielectric properties, while those closer to the solute atoms feature intermediate values. The continuity of the dielectric assignment using the Gaussian-based model also provides flexibility to the value assigned to a region which has an intermediate packing density. As opposed to being strictly classified as a "solute" or "solvent" region depending on the accessibility of a solvent probe (as is the case with the traditional two-dielectric model), the dielectric assigned to a region is reflective of its ability to act as a site occupied by biological water. Examples of such regions include solvent-accessible interstitial cavities and the solute–solvent interface. By virtue of this design principle and the continuity it imparts, the model renders larger cavities with a higher average dielectric than a smaller one and this simple fact can mimic the correlated relationship between the dielectric relaxation rate of water in a region with its rate of exchange with the bulk⁷ as larger cavities are more likely to be able to exchange solvent molecules with the bulk.

This work focuses on using this Gaussian-based smooth dielectric distribution model (and its variants) to elucidate the

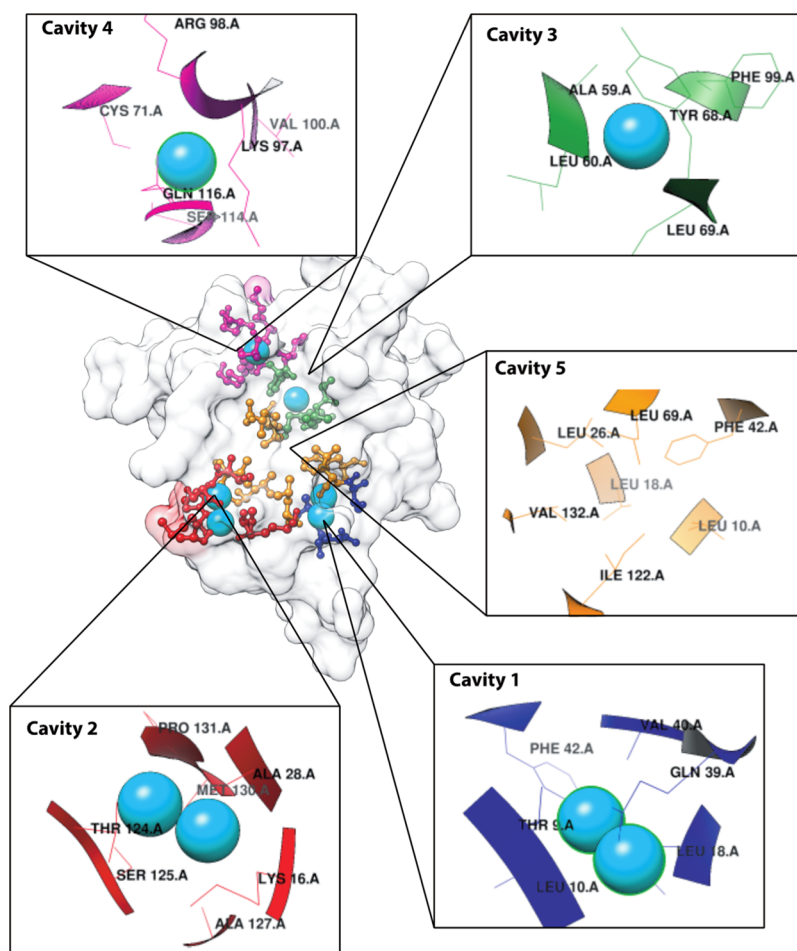


Figure 1. Structure of IL-1 β protein (PDB: 2NVH). The protein's structure is shown using surface representation. Its cavities and the water molecules present therein are shown using ball and stick representations and sphere representations, respectively. Cavity 1 residues are colored in dark blue, cavity 2 in red, cavity 3 in green, cavity 4 in magenta, and cavity 5 in orange. The water oxygen atoms are colored in cyan.

plausible values of the dielectric value in the cavity and the bulk regions of the solvated system, using observations from explicit solvent simulations as benchmarks. The basis of this work is rooted in the fact that this model has been successful in the correct prediction of pK_a values,³¹ effects of mutation on binding free energy,³² and optimum pH and proton transfer analyses,³³ using the time-efficient implicit solvent/Poisson–Boltzmann (PB) framework. We have also qualitatively shown the realism exhibited by this model in contrast to the traditional two-dielectric model.²⁹ However, its basis was mostly attributed to protein flexibility and polarizability and not so much on the water phase, and parametrization has been largely empirical.^{1,31,34,35} Therefore, this work attempts to connect its successful use with the trends of tempo-spatial properties of the water medium observed in explicit water simulations.

For this work, we chose the 1-interleukin beta (IL-1 β) protein (PDB: 2NVH) because of its cavity content, the hydration properties of which have been investigated meticulously.^{36,37} It presents itself as a good model for the observation and study of the diverse tempo-spatial properties of the water that can be present in the internal cavities, the solute–solvent interface, and the bulk region. This protein has also been used previously for a computational study that developed a super-Gaussian-based dielectric model.²⁷ Figure 1 illustrates its structure and shows the location and structure of these cavities. Table S1 lists the water identifiers and the residues in these cavities.

In the following sections, we lay out the outcomes of our work in two main parts. First, the analyses of the tempo-spatial properties of the cavity-bound, interfacial, and bulk water, as observed in the explicit water simulations, are presented. These analyses entail the observations made using two different force-field/water model combinations to evaluate if the observed time scales of water dynamics are affected by this choice. Next, we describe the Gaussian model's ability (and that of its variants) to accurately mimic these dynamics via dielectric assignment. Through qualitative reasoning, the model's portrayal of the different hydration sites is examined and a connection between the explicit and implicit solvent results is discussed. Finally, but importantly, the breakdown of this connection is discussed in the context of the limitations imposed by the small length scales of cavities on the macroscopic definition of the dielectric of a medium.

2. METHODS

2.1. System Setup and Molecular Dynamics Simulations in Explicit Water. Before setting up the protein for a molecular dynamics simulation in explicit water, all 149 water molecules present in the crystal structure were removed, including the 6 that were buried in 4 internal cavities of the protein³⁶ (see Figure 1). The 10 other hetero-atoms were also discarded. The resulting structure was fixed using the PROFIX

tool from the JACKAL package (<http://honig.c2b2.columbia.edu/jackal>) to model the missing C-terminal Ser153 residue.

Explicit water molecular dynamics simulation was performed using GROMACS v5.0.5.³⁸ For these simulations, two different water models (and two different force fields) were used. The AMBER99SB/TIP3P combination was used to simulate the dynamics in the presence of a nonpolarizable water model, while the OPLSAA/TIP4P combination was used for a simulation in the presence of a polarizable water model. For both combinations, a cubic water box was built with the protonated protein at its geometric center. The box also consisted of Na⁺ and Cl[−] ions that were added to neutralize the $-1e$ net charge of the protein while also maintaining a salt concentration of 0.15 M. As the protein was solvated with these water models, no effort was made to ensure that the cavities bereaved of crystallographic waters were occupied by water molecules. Our intention was to allow the system to evolve under isothermal–isobaric (*NPT*) conditions and to examine if water molecules can find a way to these cavities.

Three independent *NPT* simulations, each 30 ns long, were carried out for each of the force-field/water model combinations. These simulations were initiated with different initial velocities and consisted of, at most, 10,000 steps of the steepest descent (SD) minimization followed by a 500 ps long equilibration under constant volume condition (*NVT*) at 300 K and a 2 ns long equilibration under *NPT* conditions at 1 atm pressure. Both of these steps were performed while restraining the position of the protein-heavy atoms with an isotropic force constant of 1000 kJ/mol/nm, and all the covalent bonds were constrained using LINCS. Constant-temperature conditions were ensured by the velocity-rescaling thermostat,³⁹ and the Parrinello–Rahman barostat was used for maintaining constant pressure.

2.2. Internal Cavities and Hydration Shells. As another goal of this work was to outline the differences in the properties of internal cavity waters and bulk waters (including any water molecule located outside of the molecular surface of the protein), residence time of the latter was also computed in a distance-dependent manner. The volume of the bulk was divided into concentric but disjoint regions, as illustrated in Figure 2. Each region, identified as a hydration shell, was 3 Å in

width. These shells were treated as hydration sites in order to calculate different time scales. The same analysis performed on the internal cavities was also performed on the shells. For the volume of the cubic box used for the explicit solvent simulations, the solvent volume could be divided into five shells. Properties observed in “Shell 1” denoted the attributes of the first hydration shell, or in other words, of the solvent juxtaposed with the protein’s surface and interacting dominantly with it. The properties of “Shell 5” were deemed to represent the behavior of bulk water. For each of these hydration shells, a set of representative grid points in DelPhi was selected. These representative points were taken from a surface drawn at a distance, corresponding to a hydration shell using Delphi’s SURFPOT module.⁴⁰ On these points, the dielectric value assigned by any of the models was computed. Their distribution at these points was defined as depicting the distribution at that distance.

2.3. Mean Residence Time. The residence time of water at a given site was computed from the index autocorrelation function (IACF). The IACF measures the probability of detecting a particular water molecule, labeled by some index, in two snapshots separated by some time gap (Δt). The definition of the IACF, represented in this article using $P(\Delta t)$, is very similar to the one used in ref 41 and is essentially a time correlation function. In the literature, many variants of this correlation function exist and are primarily developed in order to compute the time for which a water molecule typically occupies a hydration site (see ref 42). In our work, the IACF was defined as the following

$$P(\Delta t) = \left\langle \frac{\text{length}(I(t_0 + \Delta t) \cap I(t_0))}{\text{length}(I(t_0))} \right\rangle_{t_0} \quad (3)$$

In this equation, $I(t)$ is a set of indices of the water molecules that occupied a site at some time t , and $\text{length}(I(t))$ is the number of elements in this set. For various starting points in time, denoted by t_0 , the IACF is computed as a function of Δt . In a rather simplistic case, if snapshots separated by a given Δt consistently show that no water molecules occupying a site are of the same index (e.g., residue sequence number), then $P(\Delta t) \approx 0$ is likely to be the case. Such results can be expected for snapshots separated by a large value of Δt . On the other hand, if the snapshots share a large number of common water molecules occupying that site, then $P(\Delta t) \approx 1$, which can be expected from snapshots separated by smaller Δt values. The IACF will therefore be bound between 0 and 1.

The computed IACF was fit to a biexponential model that has been used by Makarov et al.⁴¹ to determine the time constants. As they have noted, the biexponential model can provide reasonable insights into short- and long-term residing waters at a site but is not completely able to capture the complex nature of water dynamics. In other words, any site can be thought of as being occupied by two kinds of water—the type that stays for shorter times (short-term water) and the type that stays for infinitely longer times (long-term water). For our analyses, this distinction was sufficient. The model has the following expression

$$P(\Delta t) = a_0((1 - w)e^{-\Delta t/\tau_s} + we^{-\Delta t/\tau_l}) \quad (4)$$

In eq 4, the two terms on the right-hand side represent the contribution of the short- and long-term residing water molecules. The factor a_0 ($\in [0,1]$) denotes the occupancy of the site whose occupancy is being modeled; a value of 1 means

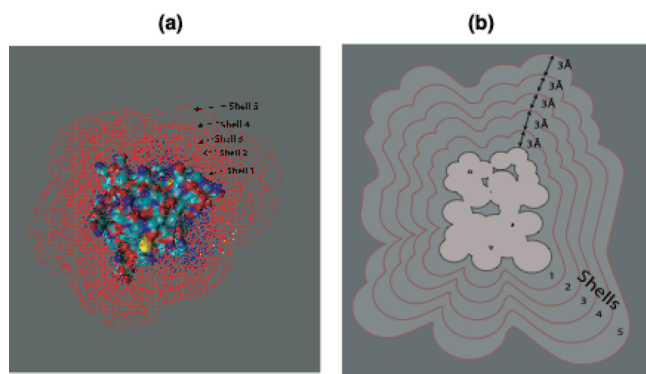


Figure 2. External hydration shells. A cartoon illustration of the five concentric and disjoint external hydration shells surrounding the protein is shown. Each shell is 3 Å in width and labeled in increasing order as a function of distance from the surface of the solute (colored in red). (a) Representative points for the shells around the protein used for dielectric calculations. (b) Simplistic version of the division of the bulk region into five hydration shells around the protein.

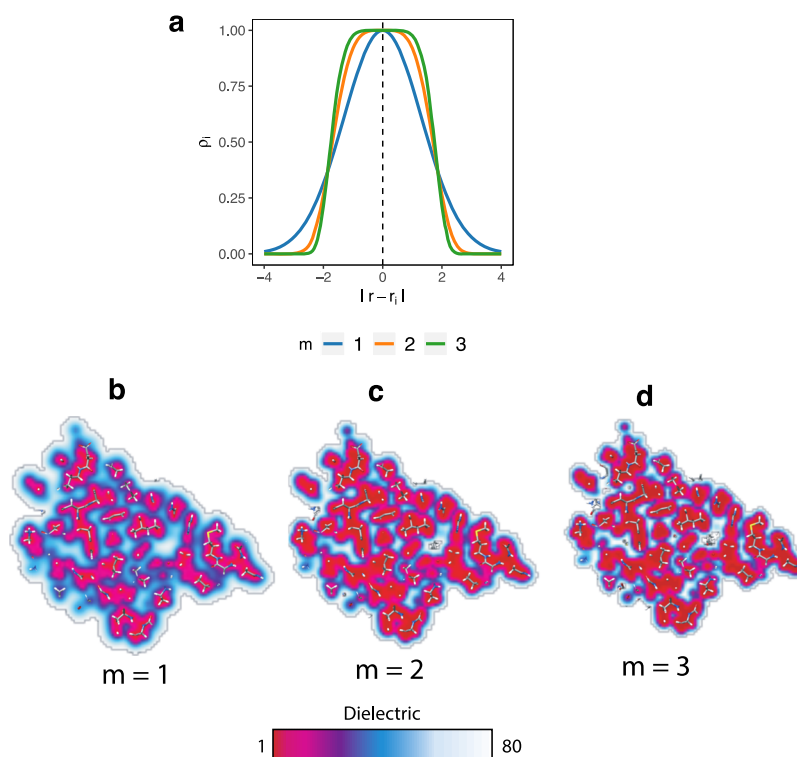


Figure 3. Depiction of the Gaussian and the super-Gaussian models of dielectric distribution. (a) For the various values of “ m ”, the shape of the probability function is shown. (b–d) Dielectric distribution obtained using different values of “ m ” shown for a slice cutting through the center of the crystal structure of IL-1 β protein. Although subtle, the effect of varying “ m ” is evident.

that it is fully occupied by water in all snapshots. The factor w ($\in[0,1]$) is the weight that linearly modulates the contribution of the long-term residing water molecules to the time correlation function. Concurrently, $(1-w)$ is the weight associated with the short-time residing waters. The corresponding time constants are denoted by τ_1 and τ_s .

2.4. Water Orientational Relaxation Time. The orientational relaxation time of water was computed from the rotational autocorrelation function (RACF) of its dipole moment vector. The RACF of the water dipole moment was computed using the definition mentioned in ref 43, which is based on the angle of rotation of the unit dipole moment vector of a water molecule. Expressed as in eq 5, the RACF or $C_2(\Delta t)$ maps the average deviation of the unit vector from some starting configurations after time Δt . The deviation is quantified using the second-order Legendre function of the inner product of the unit vectors (the cosine of the angle between them in configurations separated by time Δt).

$$C_2(\Delta t) = P_2(\hat{\mu}(t_0 + \Delta t) \cdot \hat{\mu}(t_0))_{t_0} \quad (5)$$

Time-dependent unit dipole moment vectors are denoted using $\hat{\mu}(t)$ and $P_2(x) = 3x^2/2 - 1/2$ is the second-order Legendre function. By virtue of this definition of the RACF, any two configurations of a water molecule (or any dipolar entity) separated by some fixed time Δt will contribute a $C_2(\Delta t) = 1$ if their unit vectors are aligned in the same direction and $C_2(\Delta t) = 0$ if the alignment is random.

The RACF was fitted to a single exponential model expressed in eq 6 as opposed to a biexponential⁴³ or stretched exponential model⁴⁴ that has been used in identical analyses. Because our objective was not to determine the precise values of the rotational relaxation time but only to focus on their relative

trend as a function of the micro-environment, a single exponential model serves the purpose.

$$C_2(\Delta t) = a_0 e^{-\Delta t/\tau_r} \quad (6)$$

Based on the fit, the orientational relaxation time or τ_r was determined.

2.5. Super-Gaussian-Based Implicit Solvent Dielectric Models. Recent work by Hazra et al.²⁷ inspired modifications to the Gaussian dielectric model, which is known as the super-Gaussian dielectric model. By increasing the quadratic exponent of the function used in the Gaussian model (eq 1), it allows for the spread of the probability function to be adjusted. This is achieved by introducing an exponent multiplier “ m ” in the probability function (eq 7).

$$\rho_i^m(\vec{r}) = \exp\left(-\left(\frac{(\vec{r} - \vec{r}_i)^2}{\sigma^2 R_i^2}\right)^m\right) \quad (7)$$

The modified probability function can use the formulations to deliver a varied dielectric distribution. When $m = 1$, the original Gaussian model is retrieved. With increasing m , the shape of the probability function changes as the slope of the transition region increases (see Figure 3a). With very high values of m , the distribution asymptotically approaches the two-dielectric model (with σ that retains the van der Waals radius), and along this transition, the average dielectric of the system increases. The super-Gaussian model manipulates the representation of the mobility of each solute atom (depicted by the spread of the Gaussian probability function) while ensuring that the continuity of the dielectric distribution is preserved (Figure 3b–d). This is particularly important in the context of computing solvation energies (or energy of transfer across two

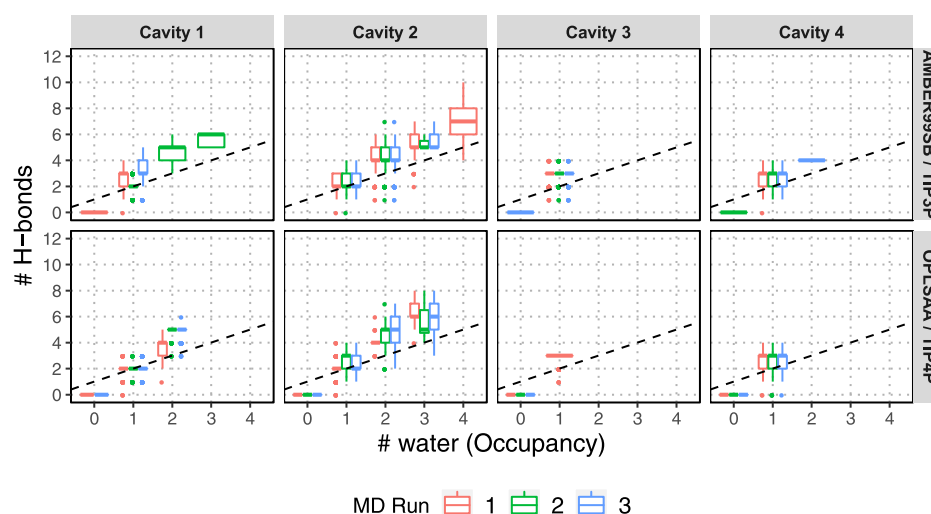


Figure 4. Number of hydrogen bonds as a function of occupancy of a cavity. Plots showing the number of hydrogen bonds formed between the cavity residues and the water molecules visiting them. Each box plot indicates the range of the number of hydrogen bonds formed for a corresponding value of a cavity's occupancy as seen across the MD trajectory. The relation observed across different MD runs is denoted using different colors and that applies to simulations conducted using either force-field/water model combination. The dashed black line is the line of equality (slope = 1 and y-intercept = 0).

different dielectric media) by virtue of the implementation of the finite difference algorithm used by DelPhi to solve the PB equation. For more information regarding the technical details, we refer elsewhere.^{1,29}

2.5.1. Cavity Dielectric Term. The super-Gaussian model²⁷ has another extension which was considered separately in our work. The use of an explicit cavity dielectric or ϵ_{gap} term in their model places an upper bound on the dielectric value that can be assigned to regions in the vicinity of the protein atoms. In its absence, a region's dielectric lies between ϵ_{ref} and $\epsilon_{\text{solvent}}$ (the protein and the solvent dielectric) and is dictated only by ρ , the value of the probability of finding any solute atom there (eqs 1 and 2). With the addition of the ϵ_{gap} term, a limit is placed on the value of the dielectric assigned to a point in space based on its proximity to a representative solute or solvent phase. This modifies the dielectric distribution in the representative solute region which, for the purpose of this work, was delineated by an isosurface defined by some cutoff value of ρ that best describes the van der Waals surface S of the protein.³⁵ Using S , the following dielectric distribution is modeled

$$\epsilon(\vec{r}) = \begin{cases} \epsilon_{\text{ref}} \rho(\vec{r}) + \epsilon_{\text{gap}}(1 - \rho(\vec{r})), & r \text{ is inside } S \\ \epsilon_{\text{ref}} \rho(\vec{r}) + \epsilon_{\text{solvent}}(1 - \rho(\vec{r})), & r \text{ is outside } S \end{cases} \quad (8)$$

The above mentioned scheme works irrespective of the value of “ m ” and was implemented in DelPhi as part of its development for future releases. We note that the dielectric function $\epsilon(\vec{r})$ defined in eq 8 is usually discontinuous across the isosurface S , while $\epsilon(\vec{r})$ of the original super-Gaussian model,²⁷ based on a different implementation, is continuous and has derivatives of all orders. Because of such a difference, the optimal parameter values of the present super-Gaussian results could be different from those in ref 27.

2.6. Computing Pairwise Atomic Electrostatic Interaction Energies. Pairwise electrostatic interaction energy of atom pairs from certain cavities was computed using the FRC module of DelPhi (see ref 45) in the presence and absence of

explicit water molecules. Essentially, one of the atoms from each pair was treated as the “source” of the local electrostatic field and the other one was treated as the “target”. The source atom was charged, while the rest of the protein (including the target atom) was not, in order to obtain the electrostatic potential at the positions of the target without the influence of other charges. From the potential, the pairwise interaction energy was computed by multiplying the charge of the target atom with the potential. Because the effect of the presence of explicit water molecules on this interaction was the main goal of this exercise, a set of calculations was performed by modeling their explicit presence in the cavities. Like the source atom, their charges were also kept intact in these calculations.

3. RESULTS AND DISCUSSION

The primary objective of this work was to infer the differences in the structural and tempo-spatial properties of water bound to the internal cavities and in hydration shells outside of the protein's surface through explicit water MD simulations. Using these inferences, a dielectric model for the implicit solvent PB model is investigated to examine its capacities to capture the solvation effects seen in those simulations. We first discuss the observations made from the explicit water simulations and then switch to capturing the effects using a Gaussian-based dielectric model. We demonstrate the ability of the dielectric model to mimic the differences in the attributes shown by cavity and bulk water and how it serves to provide a deeper insight into solvation effects than the traditional two-dielectric model.

3.1. Occupancy of the Internal Cavities. The five internal cavities in the protein's structure were tested for the number of water molecules that they hosted (known as the occupancy) and the number of hydrogen bonds that existed between them and the cavity residues. A water molecule was said to be occupying a cavity if its oxygen atom was placed no farther than 3 Å from any heavy atom of a cavity residue. Occupancies for each of the five cavities in the structure were computed from three independent runs obtained using AMBER99SB/TIP3P and OPLSAA/TIP4P combinations and the results are illustrated in Figure 4.

Four out of the five cavities were found to be hydrated or occupied for at least a few snapshots, if not for the entirety of the

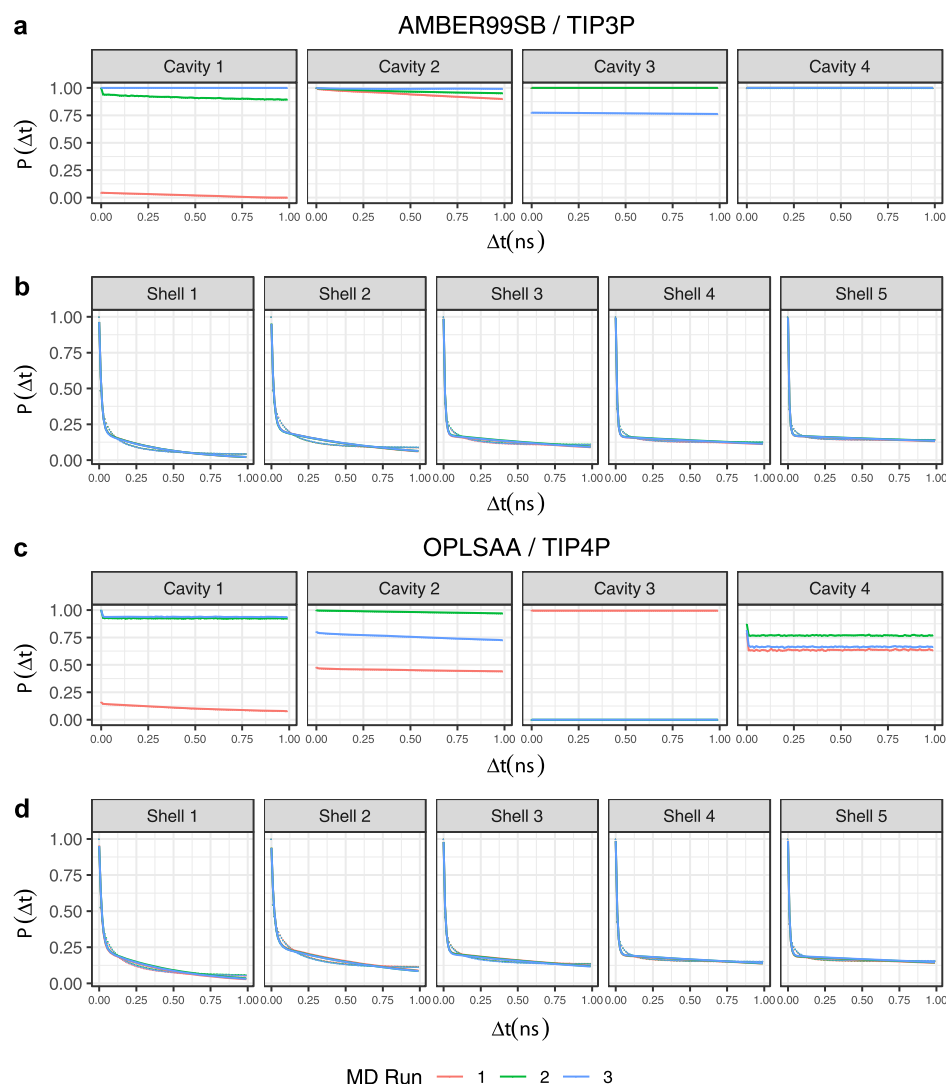


Figure 5. IACF profiles of water in the cavities and the bulk hydration shells from independent runs using different force-field/water model combinations. (a,b) IACF of the four cavities and the hydration shells, respectively, obtained from the trajectories of simulation performed using the AMBER99SB force field and TIP3P model of water. (c,d) IACF of the cavities and hydration shells from trajectories of simulation performed using the OPLSAA force field and TIP4P model of water. The profiles from the independent runs are colored differently (i.e., run 1—orange, run 2—green, and run 3—blue).

production run. Interestingly enough, the fifth cavity, or “cavity 5”, which is as large as the first and second cavities in terms of volume, was found to have zero occupancy across all the runs irrespective of the force-field/water model combination. This is concomitant with the observation reported in ref 36 where this cavity was the only one with no water molecule occupation. As mentioned above, the hydration of this cavity, as observed using different experimental techniques, has yielded conflicting results.³⁷ It has been argued that this cavity is in fact occupied by water molecules but because of their very low residence times or high mobility, they are not detectable using standard crystallographic methods. If that was the case, then based on the sampling frequency in our simulations, the residence time of water visiting cavity 5 would have to be less than 10 ps, which is unusually small.

The typical occupancy of the remaining four cavities matched well with that reported by Quillin et al.³⁶ and correlated positively with their volumes (see Table S1). Cavities 1 and 2 featured a higher occupancy than cavities 3 and 4. The results were seemingly independent of the force-field and water model

combination. However, these observations provided us with configurations of the protein where its cavities were hydrated despite starting with a structure where no water molecule was present in them. This suggests that these cavities are connected to the bulk via some conduit whose conformation may largely depend on the conformation of the protein. Consequently, it permits dynamic exchange of water molecules between the bulk and the cavity sites, which was seen during the simulations (discussed later). In addition, water molecules visiting the cavities engaged in more than one hydrogen bond with the residues lining them, except for a few cases. Generally, a water molecule appeared to be forming ~2–3 hydrogen bonds on average which is suggestive of multiple acceptor/donor sites in the cavities. Such an environment inhibits bulk-like mobility of the water in the cavity not only by virtue of geometric constraints but also by engaging in favorable interactions with the water molecules. On top of that, the favorable interactions compensate for the penalty paid by the water molecules that leave the bulk and enter the cavities. These hydrogen-bonding sites (acceptors/donors) in the cavities and the degree of their participation

Table 1. Mean Residence Times of the Water Molecules in the Internal Cavities^a

cavity	short-term residence time τ_s (ps)		long-term residence time τ_l (ps)		weight factor w		occupancy a_0		goodness of fit R^2	
	TIP3P	TIP4P	TIP3P	TIP4P	TIP3P	TIP4P	TIP3P	TIP4P	TIP3P	TIP4P
1	— ^b	5.51	—	1532.07	—	0.90	—	0.16	—	0.995
	4.52 ^c	3.49	>10,000	>10,000	0.94	0.93	1.00	1.00	0.977	0.961
	— ^d	3.21	—	>10,000	—	0.94	—	1.00	—	0.955
2	8.53	12.33	>10,000	>10,000	0.99	0.98	1.00	0.48	0.998	0.998
	428.56	7.11	>10,000	>10,000	0.97	1.00	1.00	1.00	0.997	1.000
	57.01	16.06	>10,000	>10,000	0.99	0.98	1.00	0.80	0.706	0.999
3	—	0.45	—	>10,000	—	1.00	—	1.00	—	0.909
	—	—	—	—	—	—	—	—	—	—
	0.05	—	>10,000	—	1.00	—	0.78	—	1.000	—
4	—	0.14	—	>10,000	—	0.81	—	0.79	—	0.928
	>10,000	0.42	>10,000	>10,000	1.00	0.88	1.00	0.87	1.000	0.959
	0.43	3.67	>10,000	>10,000	1.00	0.82	1.00	0.82	1.000	0.959

^aFor the four internal cavities, values of the parameters are listed that optimally fit their respective IACF profiles using the biexponential model. The residence times are indicated by τ_s and τ_l which represent the typical residence times of short-term and long-term waters at these cavities. Parameters " w " and " a_0 " respectively, denote the weight of the contribution of the long-term water molecules to the IACF and the water occupancy. A dash (—) indicates that IACF profiles could not be fitted using the biexponential model for these cases. ^bData from MD run 1. ^cData from MD run 2. ^dData from MD run 3.

Table 2. Mean Residence Times of the Water Molecules in the Bulk Hydration Shells^a

shell	short-term residence time τ_s (ps)		long-term residence time τ_l (ps)		weight factor w		occupancy a_0		goodness of fit R^2	
	TIP3P	TIP4P	TIP3P	TIP4P	TIP3P	TIP4P	TIP3P	TIP4P	TIP3P	TIP4P
1	14.55 ^b	16.55	434.27	468.85	0.21	0.26	0.96	0.96	0.969	0.970
	14.66 ^c	17.50	443.23	525.81	0.21	0.26	0.96	0.95	0.970	0.967
	14.87 ^d	17.19	452.25	484.79	0.20	0.26	0.96	0.95	0.969	0.969
2	15.94	19.17	771.85	891.40	0.23	0.29	0.95	0.94	0.951	0.955
	15.61	18.87	806.10	869.00	0.23	0.29	0.95	0.94	0.952	0.956
	15.90	19.55	809.43	875.02	0.23	0.28	0.95	0.94	0.951	0.955
3	9.64	11.55	1510.8	1735.0	0.17	0.22	0.99	0.98	0.968	0.968
	9.69	11.59	1563.7	1736.5	0.18	0.22	0.98	0.98	0.968	0.968
	9.62	11.69	1540.6	1747.5	0.18	0.21	0.98	0.98	0.968	0.968
4	8.15	10.14	2627.6	2862.1	0.16	0.20	0.99	0.99	0.978	0.972
	8.11	10.12	2814.8	2846.4	0.17	0.20	0.99	0.99	0.979	0.973
	8.11	10.08	2768.3	2895.7	0.16	0.20	0.99	0.99	0.979	0.973
5	7.98	9.72	3886.8	3646.6	0.17	0.19	0.99	0.99	0.979	0.973
	7.81	9.84	4234.3	3675.7	0.17	0.19	0.99	0.99	0.980	0.972
	7.86	9.77	4118.4	3753.8	0.17	0.20	0.99	0.99	0.980	0.972

^aFor these shells, values of the parameters are listed that optimally fit their respective IACF profiles using the biexponential model. The meanings of the symbols are described in the footnote of Table 1. ^bData from MD run 1. ^cData from MD run 2. ^dData from MD run 3.

are illustrated in Figure S1. Collectively, these configurations constituted the data set that was used to analyze the distinct tempo-spatial behavior of water molecules in these cavities, as described in the following sections.

3.2. Mean Residence Time. The mean residence time of water molecules in the hydrated cavities was computed using the procedure detailed in Section 2.1.

It is best to compare the visual appearance of the IACF profiles of the internal cavity and bulk waters first in order to obtain a qualitative sense of the difference between them. The stark difference is evident from the plots shown in Figure 5 which show the IACF of the four hydrated internal cavities and the five external hydration shells obtained using both of the force-field/water model combinations. From the IACF plots of the internal cavities, it is clear that the $P(\Delta t)$ values generally tend to be greater than 0.5 for smaller and larger values of Δt (0–1 ns) used in the calculation except for a few special cases where the occupancy was low throughout the simulation (e.g.,

cavity 1 for run 1). On the contrary, the IACF profiles of the shells in the solvent volume show a very different tendency. $P(\Delta t)$ drops from 1 at 0 ns (identical snapshots) to less than 0.10 at 0.50 ns.

Upon fitting the IACF profiles to the biexponential model (eq 4), quantitative analyses of the difference were performed. At this point, it is worthwhile to mention two significant points. First, the model assumes that at a given hydration site, short-term and long-term waters can exist and that the optimal fit of the observed IACF profile yields the corresponding time scales, known as the short-term (τ_s) and long-term residence times (τ_l), respectively. However, it is crucial that they are examined in the light of the weight factor " w " which is an indicator of the "category" of water whose presence dominates (i.e., has greater contribution to the trend of the IACF profile). Second, it is also important to consider the goodness of the fit of the model to the observed IACF. Using the coefficient of determination, $R^2 \in [0,1]$, the quality of the fit was judged and we argue that if the fit

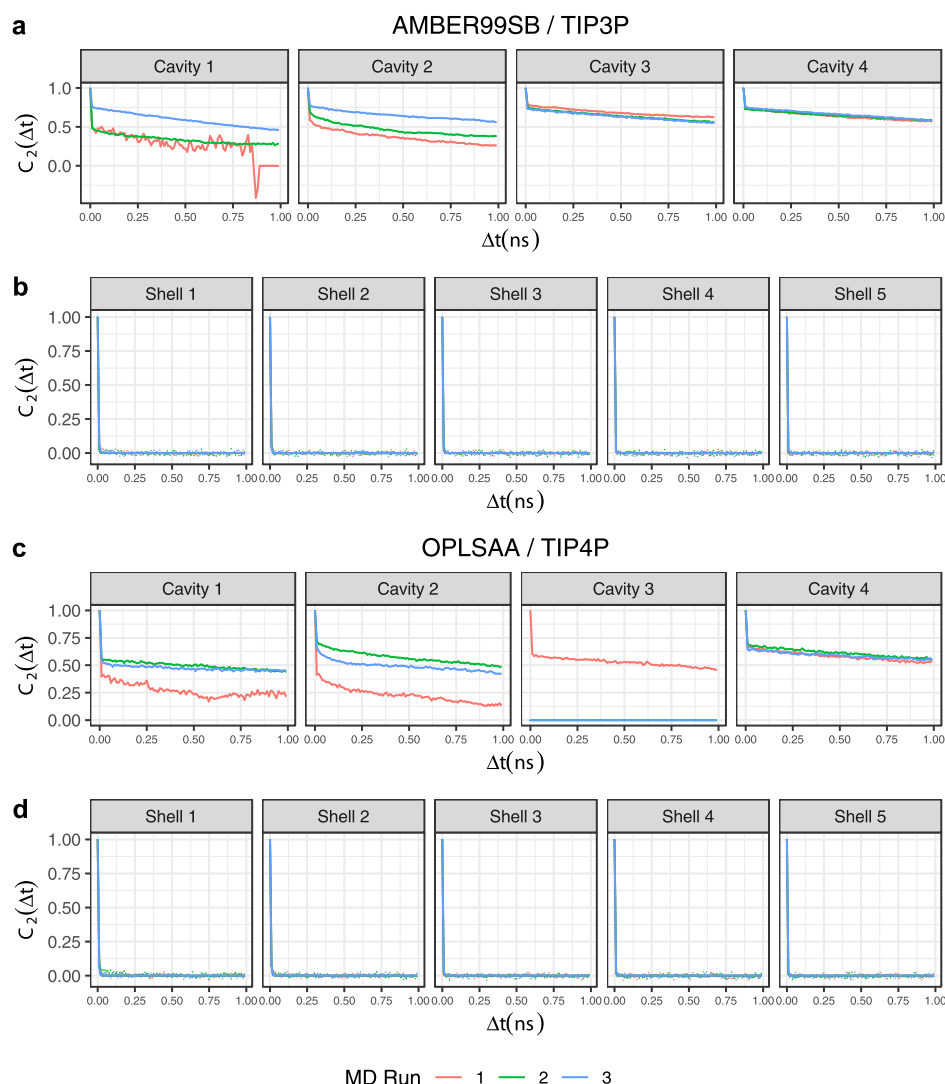


Figure 6. RACF profiles of the water in the cavities and the bulk hydration shells from independent runs using different force-field/water model combinations. (a,b) RACF of the four cavities and the hydration shells, respectively, obtained from the trajectories of simulation performed using the AMBER99SB force field and TIP3P model of water. (c,d) RACF of the cavities and hydration shells from trajectories of simulations performed using the OPLSAA force field and TIP4P model of water. The profiles from the independent runs are colored differently, that is, run 1—orange, run 2—green, and run 3—blue.

does not describe the observed IACF with an appreciable R^2 , then the optimal fit parameters are meaningless.

The values of the fit parameters are listed in Tables 1 and 2. For all four hydrated cavities and the five shells, these values are listed for the two force-field/water model combinations. Because these fits are computed numerically and the results can be machine precision-dependent, we chose to denote absurdly large numbers with a “>10,000” label which is only applicable to the residence time constants. For cases where the fit failed because of numerical errors/exceptions, the values are simply marked with a dash (–), conveying that the biexponential fit was not able to describe the observed IACF profile.

Trends featured by the internal cavity water, which are evident from the data in Table 1, show striking differences from those observed in the bulk (see Table 2), which is concomitant with the visual appearance of the IACF profiles. For cases where the biexponential model fit did not fail, the weight factor w is very high (>0.8). This primarily indicates that internal cavities tend to be hydrated by long-term residing water molecules and their

contribution to the overall trend of the correlation function is >80%, compared to <30% in the bulk. Cases where the optimal fit values could not be computed suggest that the biexponential model for describing time correlation properties for internal cavity waters is insufficient. Because of higher values of w for the cavity waters, the short-term residence times (τ_s) are not meaningful. Instead, the residence time is better judged using the long-term residence time (τ_l) which is typically >10,000 ps. On the contrary, the smaller values of w in the hydration shells located outside the protein’s molecular surface tell us that the lifetime of water there is short. The residence times in the bulk can, therefore, be quantified using the short-term residence time and the value is typically 7–14 ps.

Overall, these results are not unexpected because internal cavities are bound to impose steric constraints on the motion of water, and if polar interaction sites are present, these water molecules will favorably interact with these sites and feature slower diffusion rates. The more permanent nature of water residence at these sites is complemented by the presence of water seen in the crystal structure of this protein at these

cavities.³⁶ However, it is worth noting that the residence times may differ by orders of magnitude between the cavity and bulk waters and they reflect the significant variation in their translational mobilities in these environments. Moreover, large residence times in the cavities also signal that the rate of exchange of water between the cavity and the bulk is low. Learning from the correlation between the exchange rate and dielectric relaxation times,⁷ it is deducible that the typical cavity dielectric should be much smaller than that of the bulk.

For better visualization of the mobility of water in the cavities and in the vicinity of the protein's surface, a collection of movies generated from the MD trajectories are provided as the [Supporting Information](#). Two separate movies show long-term residing water in a low-volume cavity (cavity 4; [Video S1](#)) and relatively short-term residing waters in a larger cavity (cavity 2; [Video S2](#)). The former illustrates the restricted translational diffusion of cavity-bound water and the practically fixed orientation of its dipole moment. The latter shows a relaxed setting in a larger cavity and also provides an example of exchange of water between the cavity site and the bulk. A third movie ([Video S3](#)) shows the mobility of water molecules in the vicinity of LYS138 (a solvent exposed charged residue) and illustrates their motion in the bulk at varying distances from it.

3.3. Dipole Orientation Relaxation Time. The dipole orientation relaxation time was computed for the water in the cavities and the hydration shells. The RACF was computed at these sites, and the relaxation time, τ_R , was determined using the protocol discussed in [Section 2.2](#).

The RACF profiles of the four hydrated cavities and the external shells are shown in [Figure 6](#). A qualitative difference is immediately apparent in the profiles. The profiles observed for the internal cavities feature a very slow decay rate, while the profiles for the hydration shell feature a quick decay rate. The relaxation times obtained after fitting these RACFs to the monoexponential model ([eq 6](#)) are listed in [Tables 3 and 4](#). It is evident that τ_R is significantly larger at the cavity sites than at the hydration shells—typically 2–3 orders of magnitude higher. As is commonly represented, this difference would yield a “slowdown” factor, which is the ratio $\tau_R(\text{cavity})/\langle\tau_R(\text{bulk})\rangle$, of

Table 3. Dipole Orientation Relaxation Times of the Water Molecules in the Internal Cavities^a

cavity	dipole orientation relaxation time τ_R (ps)		prefactor a_0		goodness of fit R^2	
	TIP3P	TIP4P	TIP3P	TIP4P	TIP3P	TIP4P
1	616.08	1227.39	0.56	0.40	0.561	0.429
	1500.12	817.98	0.47	0.40	0.567	0.599
	1874.99	3632.12	0.77	0.61	0.935	0.514
2	1190.83	3689.17	0.55	0.67	0.770	0.633
	1614.73	3658.18	0.63	0.57	0.782	0.441
	3191.32	2720.40	0.76	0.69	0.834	0.761
3	3996.95	—	0.78	—	0.814	—
	3282.76	4206.76	0.75	0.69	0.826	0.636
	3212.84	5032.20	0.74	0.53	0.818	0.232
4	3504.76	2925.81	0.74	0.59	0.788	0.526
	4019.83	—	0.74	—	0.743	—
	3660.03	5144.20	0.76	0.65	0.817	0.456

^aFor the four internal cavities, values of the relaxation times (τ_R) obtained after optimally fitting their respective RACF profiles using the single exponential model are listed. “—” conveys that RACF profiles could not be fitted using the exponential model for these cases. ^bData from MD run 1. ^cData from MD run 2. ^dData from MD run 3.

Table 4. Dipole Orientation Relaxation Times of the Water Molecules in the Bulk Hydration Shells^a

shell	dipole orientation relaxation time τ_R (ps)		prefactor a_0		goodness of fit R^2	
	TIP3P	TIP4P	TIP3P	TIP4P	TIP3P	TIP4P
1	3.65 ^b	4.07	1.00	1.00	0.989	0.982
	2.82 ^c	3.85	1.00	1.00	0.988	0.990
	3.77 ^d	2.08	1.00	1.00	0.988	0.989
2	3.75	2.32	1.00	1.00	0.991	0.987
	3.50	2.95	1.00	1.00	0.990	0.989
	3.24	4.55	1.00	1.00	0.990	0.968
3	2.43	4.04	1.00	1.00	0.991	0.988
	3.35	2.71	1.00	1.00	0.988	0.992
	2.58	2.83	1.00	1.00	0.991	0.989
4	2.10	2.24	1.00	1.00	0.992	0.992
	2.51	3.54	1.00	1.00	0.987	0.986
	2.46	3.92	1.00	1.00	0.991	0.985
5	0.25	2.89	1.00	1.00	0.989	0.991
	2.21	2.86	1.00	1.00	0.990	0.989
	2.81	2.81	1.00	1.00	0.990	0.988

^aFor the bulk shells, values of the relaxation times (τ_R) obtained after optimally fitting their respective RACF profiles using the single exponential model are listed. “—” conveys that RACF profiles could not be fitted using the exponential model for these cases. ^bData from MD run 1. ^cData from MD run 2. ^dData from MD run 3.

more than 10^3 . This slowdown is significantly larger than the slowdown observed in the first hydration shell, a value generally <10 .

For cases where the fit failed because of numerical errors/exceptions while computing, the values are simply marked with “—” and they exemplify that the monoexponential fit is not universally able to describe RACF profiles. Accordingly, the goodness of fit (R^2) of the RACF of the internal cavities to the monoexponential model ([eq 6](#)) suggests that the fit does not describe the time correlation for the cavities and it does for the external hydration shells. For the cavities, R^2 ranges between 0.2 and 0.9 with a median value of 0.8, whereas the average value is 0.99 for the hydration shells. These values of R^2 are not meant to describe the absolute nature of the time correlation of rotational mobility and therefore must only be interpreted as a relative measure of the difference in the rotational mobilities of the bulk water and the internal cavity water.

3.4. Dipole Moment Fluctuations. Dipole moment fluctuation, defined as $(\delta\mu)^2 = \mu^2 - \bar{\mu}^2$, has a direct relationship with the dielectric value (ϵ), by virtue of the linear response expression.

$$\epsilon - 1 = \frac{4\pi}{3Vk_B T} (\delta\mu)^2 \quad (9)$$

Although widely prescribed (alongside its variants) for computing the total dielectric of systems solvated in polar solvents from MD simulations,^{6,46–48} certain cautions must be taken when biomolecules (with their complex geometries and dynamics) are examined for obtaining the localized distribution of dielectric. First, the fluctuation (and therefore the dielectric) is derived from the difference in the running average of μ^2 (2-norm of the total dipole moment in a region) and square of the running average of the dipole moment vector $\bar{\mu}$. If the running averages do not converge, the dielectric computed from [eq 9](#) will be spurious and not reflective of the equilibrium screening

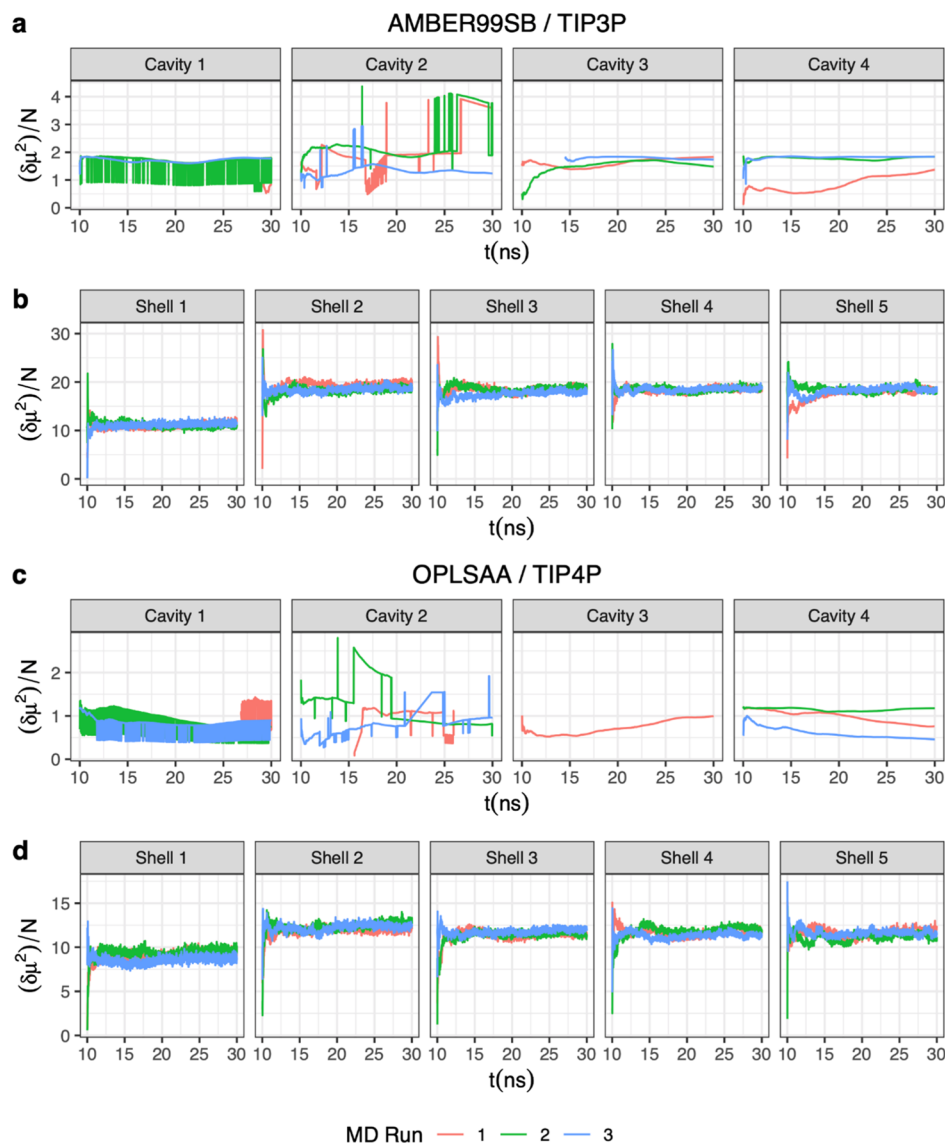


Figure 7. Dipole moment fluctuation per water molecule in the cavities and the bulk water shells as a function of simulation time. (a,b) Average dipole moment fluctuation of the four cavities and the hydration shells, respectively, are shown as a function of time when these simulations were performed using the AMBER99SB force field and TIP3P model of water. (c,d) Same is shown for simulations performed using the OPLSAA force field and TIP4P model of water. The profiles from the independent runs are colored differently, that is, run 1—orange, run 2—green, and run 3—blue.

capacities. Second, the volume (“ V ” in eq 9) cannot be correctly determined for complicated geometries that are prevalent in the current study. We have therefore restricted our analyses to computing regional or local $(\delta\mu)^2$ only. In fact, we have computed and compared $(\delta\mu)^2/N$: fluctuation per molecule, where N is the average number of dipole moment vectors present in a region. This factor indirectly accounts for the effect of the volume term in eq 9 and provides a more sensible insight into the polarizability in a region.

In Figure 7, $(\delta\mu_w)^2/N$ (subscript “w” denotes water) is plotted as a function of time for the four hydrated cavities and the five external hydration shells for both force-field/water model combinations. The lack of convergence of the fluctuation per molecule in the cavity regions is clearly evident from Figure 7a,c. Regardless of the nonconvergence, the range of the fluctuations suggests that fluctuation per molecule is significantly smaller in the cavity region compared to the hydration shells (Figure 7b,d) for which the fluctuations appear to be well-converged. In the hydration shells, the converged value of the fluctuation is ~ 10

debye² for the nearest shell and 17–20 debye² (TIP3P model) and 11–13 debye² (TIP4P model). In the cavities, however, water tends to exhibit a fluctuation of only ~ 1 –2 debye² per molecule. Because of the fewer number of water molecules in the cavities, the fluctuation at such a per-molecule basis has a chaotic appearance [specifically for cavity 2 which also has the largest range of water occupancy (Figure 4)]. A higher fluctuation is indicative of the ability of the dipole moment vectors to quickly respond/acclimate to the changing electric field in its vicinity and, hence, a higher polarizability. Qualitatively speaking, this implies that the dielectric in the exterior shells must be around an order of magnitude larger than that in the solvent-accessible cavity regions. These differences not only highlight the remarkable ratio of the average fluctuation but also hint at the characteristic features of the water model used in the explicit solvent simulations. The lower dipole moment fluctuation of the TIP4P model of water in the bulk features a fluctuation that is consistent with its relatively lower dielectric constant (~ 53).^{49,50} The same applies to the TIP3P model of water

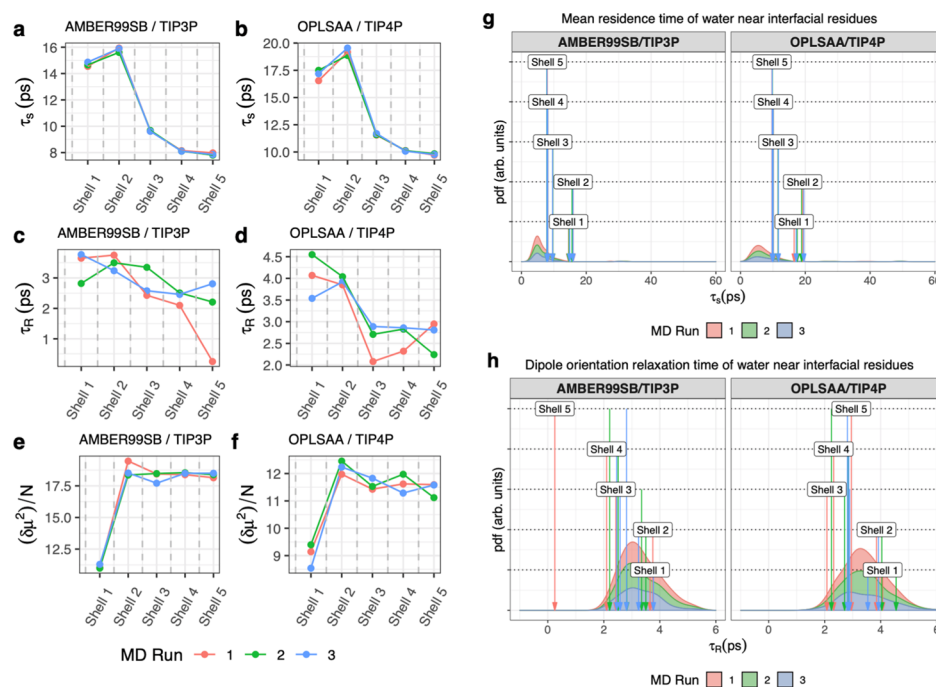


Figure 8. Heterogeneity of tempo-spatial properties of water molecules in the bulk solvent region. (a,b) Variation in the mean residence time (τ_s) of water across different shells as seen in simulations performed using AMBER99SB/TIP3P and OPLSAA/TIP4P, respectively. (c,d) Variation in the dipole orientation relaxation time (τ_R) of water dipole moment across different shells for the respective force-field/water model combinations. (e,f) Variation observed in the average dipole moment fluctuation of water in these shells for the respective force-field/water model combinations. (g) Distribution of the mean residence time values of water in the first hydration shell when the effects of the interfacial protein residues were resolved. The collective value of the mean residence time for the shells [also shown in “(a)” and “(b)” for each shell] is marked using an arrow to emphasize the contrast between distributed and combined effects. (h) Same as “(g)” but with the dipole orientation relaxation time values. Data pertaining to each independent run is colored differently but consistently in all plots, that is, run 1—orange, run 2—green, and run 3—blue.

which features a larger dipole moment fluctuation in the bulk and is shown to have a larger dielectric (>90).⁵⁰

3.5. Distance-Dependent Properties of Exterior Bulk Hydration Shells. Investigations into the variation in the water residence time, the dipole orientation relaxation time, and the dipole fluctuation per molecule across the five exterior hydration shells revealed further heterogeneity. The plots in Figure 8 show that closer to the molecular surface, the hydration shells have longer residence times and orientation relaxation times than those that are farther away. These resonate with the findings that biological water exhibits restricted translational and rotational mobilities.^{6,7,12} Specifically, the first and second hydration shells exhibit residence times of >14 ps (Figure 8a,b) and an orientation relaxation time of >3 ps (Figure 8c,d) which is almost twice as much as the farthest shell (the bulk region).

A further analysis of the residue-dependent effects on these time scales was also conducted. This was performed while keeping in mind that the protein has a complex molecular surface and that such topological variations can incur local effects in the nearby biological water. For the water around each of the interfacial residues of the protein (identified as those with relative SASA $>50\%$), the mean residence times and the dipole orientation relaxation times were computed using the same approach. Figure 8 illustrates the distribution of these time scales and also shows the corresponding value for each of the shells in order to emphasize the distinction between the distributed and the combined effect. Figure 8g shows that a major portion of the distribution of mean residence times of water in the first hydration shell can be likened to the bulk (or Shell “5”). However, the range is larger (2–50 ps) which signifies that there are regions on the molecular surface of the protein that retain

interactions with a water molecule for much longer than the others. This explains why the combined effect is a longer residence time. A similar trend is also apparent in the distribution of the dipole orientation relaxation time (Figure 8h). Although the range is not as large as the mean residence time, the distribution still illustrates that local effects are present. These local effects in the protein–water interface are not unexpected. As pointed out in ref 51, the presence of deep clefts and re-entrant patches that affect solvent accessibility are some of the geometry-based factors that can lead to local variations in the nearby water structure. The local chemical composition also plays a role. However, the Gaussian-based models in DelPhi do not account for distinct chemical properties of the atoms, and hence, the focus was only laid on obtaining the distribution of the time scales and not on their association with the chemical properties of the nearby residues/atoms.

Besides the time scales, variation in spatial properties is also evident from the degree of dipole moment fluctuation per molecule across the hydration shells. It is also lower in the first hydration shell (reduced by 50% in the case of this protein). The fluctuations per molecule plotted here were obtained after averaging the corresponding $(\delta\mu)^2/N$ values from the last 500 snapshots—a regime where convergence is vivid. Together with the time scale differences, the reduced fluctuations suggest restricted mobility and change in the structural order of water. In the hydration layers farther away, the gradual approach to bulk-like properties is also evident. Such a transition has been reported previously by theoretical and experimental studies (see ref 52 for relevant studies).

Another consideration to note is that the reduced mobility of the water in the first hydration shell (or the interfacial water) is

not at par with what is featured by the cavity-bound waters. The residence time and the dipole orientation relaxation time of the interfacial water are still shorter than those of the cavity water by more than one order of magnitude (tbl1Tables 1 and 2tbl12). This is reinforced by the comparison of the “*w*” factor (eq 4) that denotes the weight of the long-term residing water at a hydration site (see tables mentioned above). For the interfacial water (and the hydration shells in general), the value is <0.3, while in the cavity, the value is >0.8.

These general trends are independent of the force field and the water model used, although a meticulous look at the data might indicate some microscopic differences. Collectively, these variations reflect a distance-dependent distribution of tempo-spatial properties of water brought upon by the presence of a dipolar solute. Other studies using different metrics, such as tetrahedrality of the water⁶ and its alignment order parameters with respect to the protein,⁵ have also demonstrated the change that occurs in the structure of the bulk solvent because of the immersion of a dipolar molecule. Our observations resonate with these.

3.6. Dielectric Heterogeneity from Implicit Solvent Models: Two-Dielectric Versus Gaussian-Based Dielectric Distribution Models. Up to this point, we have demonstrated via explicit water simulation, the differences in the properties of cavity-bound, interfacial, and bulk water. We now shift to evaluating the use of implicit solvent models in order to capture these details through their approach of dielectric permittivity assignment. We used three different dielectric distribution models—the traditional two-dielectric model, the Gaussian model, and the super-Gaussian model, all with and without the use of a separate cavity-dielectric (ϵ_{gap}) term. The primary goal is to conceptually validate the Gaussian-based approach of smooth dielectric distribution, which is expected to assign a dielectric value to the cavities and regions near the molecular surface, which is larger than the reference internal dielectric but smaller than that of the bulk water.

The abovementioned MD results provided an assessment of the dynamics of water molecules inside cavities and at the protein–water interface. Assuming that the simulation time is long enough, the corresponding snapshot structures can be considered representative of the conformational space of the protein. Thus, as a first step to assess the ability of implicit models to reproduce the explicit solvent MD results and effects, we investigated if the implicit models could reproduce the ensemble-averaged polar solvation energy. The following discussion will elucidate what is meant by “ensemble-averaged polar solvation energy”. In our previous work,³⁵ we demonstrated that the two-dielectric model implemented in DelPhi delivers almost identical polar solvation energy as the thermodynamics integration (TI) performed in explicit water, provided the structures were kept rigid. Along these lines, we subjected each snapshot from MD simulation runs to the traditional two-dielectric DelPhi calculations and obtained the polar solvation energy. The average of these energies results in the ensemble-averaged polar solvation energy or $\langle \Delta G_{\text{polar}}^{\text{sol}} \rangle$.

Next, a particular conformation of the protein was used with these implicit solvent models to compute its polar solvation energy ($\Delta G_{\text{polar}}^{\text{sol}}$). This was compared with the ensemble-averaged value. In this work, we chose the vacuum-minimized structures with the Gaussian-based models and solvent-minimized structure for the traditional two-dielectric model based on our previous findings that these combinations work best in reproducing ensemble average using a single structure.³⁵

We expanded our search space for optimal parameters compared to the previous work³⁵ by systematically varying the value of the internal dielectric constant (ϵ_{ref} or $\epsilon_{\text{protein}}$) for both dielectric models but only σ and ϵ_{gap} for the Gaussian-based models. The results for the Gaussian- and super-Gaussian-based models are summarized in Tables 5 and 6. Table 7 reports the results from

Table 5. Optimal Combination of ϵ_{ref} and σ for Different Values of “*m*” Used in the Gaussian/Super-Gaussian Models without an Explicit ϵ_{gap} Term^a

<i>m</i>	ϵ_{ref}	σ	signed error (kT) ($\Delta G_{\text{polar}}^{\text{sol}} - \Delta G_{\text{polar}}^{\text{sol}}$)	% error
AMBER99SB/TIP3P				
1 ^b	4	0.95	+11.58	0.32
2 ^c	2	1.05	−29.31	0.80
3	2	1.06	−55.43	1.51
OPLSAA/TIP4P				
1	1	0.94	−73.31	2.00
2	2	1.05	+79.30	2.12
3	2	1.05	−5.37	0.14

^aThese combinations were obtained after surveying all the combinations in order to match the ensemble-averaged polar solvation energy with an in vacuo minimized configuration of the lowest signed error. The relative error is indicated by the % error. ^bGaussian model. ^cSuper-Gaussian model.

Table 6. Optimal Combination of ϵ_{ref} and σ for Different Values of “*m*” Used in the Gaussian/Super-Gaussian Models in the Presence of an Explicit Value for ϵ_{gap} Term^a

<i>m</i>	ϵ_{ref}	ϵ_{gap}	σ	signed error (kT) ($\Delta G_{\text{polar}}^{\text{sol}} - \Delta G_{\text{polar}}^{\text{sol}}$)	% error
AMBER99SB/TIP3P					
1 ^b	1	40	0.97	+26.92	0.73%
2 ^c	4	70	1.00	+15.19	0.41%
3	2	80	1.06	+0.39	0.01%
OPLSAA/TIP4P					
1	4	80	0.93	+3.53	0.09%
2	2	40	1.04	+2.21	0.05%
3	2	30	1.04	+21.11	0.56%

^aThese combinations were obtained after surveying all of the combinations in order to match the ensemble-averaged polar solvation energy with an in vacuo minimized configuration of the lowest signed error. ^bGaussian model. ^cSuper-Gaussian model.

the traditional two-dielectric model in the same purview. The performance of different combinations of parameters is illustrated in Figures S2 and S3 in the Supporting Information. Although the results indicate that Gaussian-/super-Gaussian-based models could reproduce ensemble-averaged polar solvation energy better than the traditional two-dielectric model, we do not emphasize this observation as comparison of data from a single protein can be statistically meaningless. Instead, we use the optimal parameter values (ϵ_{ref} , σ , and ϵ_{gap}) to assess the dielectric permittivity distribution inside cavities and around the protein.

Figure 9 shows the dielectric value assigned to the cavity and the exterior hydration shell regions. For each cavity, its dielectric permittivity was derived by taking the average of the dielectric assigned to all the grid points within 0.1 Å of its centroid. For the exterior shells, the probe points were determined using the DelPhi’s SURFPOT module⁴⁰ as detailed in the methods (Figure 2a). It is clear that in the cavity regions (Figure 9a,b), the Gaussian/super-Gaussian models assign dielectric values that are intermediate between the solute dielectric (ϵ_{ref}) and solvent

Table 7. Assessment of the Traditional Two-Dielectric Model's Ability to Reproduce Ensemble-Averaged Polar Solvation Energy^a

$\epsilon_{\text{protein}}$	signed error (kT) ($\Delta G_{\text{polar}}^{\text{solv}} - \Delta G_{\text{polar}}^{\text{solvent}}$)		% error	
	AMBER99SB/TIP3P	OPLSAA/TIP4P	AMBER99SB/TIP3P	OPLSAA/TIP4P
1	-87.23	181.05	2.38%	4.84%

^aThe best results are obtained with the internal protein dielectric value ($\epsilon_{\text{protein}} = 1.0$). The results from both force-field/water model combinations are shown.

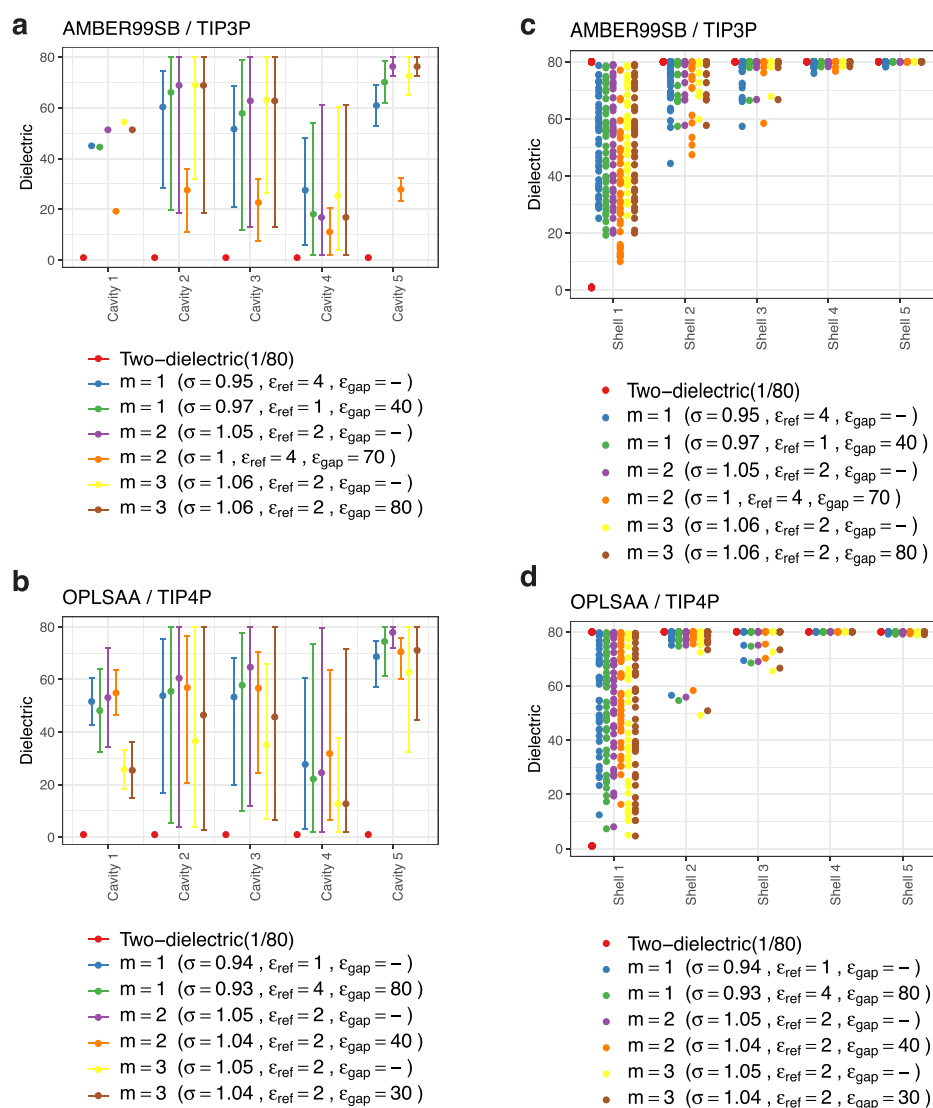


Figure 9. Dielectric values assigned in the cavities and the hydration shells using different dielectric models provided with optimal input parameters. The input parameters were assessed on their capability to best reproduce the ensemble-averaged polar solvation energy using a single structure with these dielectric models. For the traditional two-dielectric model, $\epsilon_{\text{ref}} = 1$ was used. For Gaussian/super-Gaussian models, the use of the ϵ_{gap} term is indicated by a finite value for it, otherwise its absence is marked by “-”. (a,b) Dielectric values in the cavity region using AMBER99SB and OPLSAA charges for protein atoms. The error bars denote the range of the dielectric values observed in the vicinity of the cavity centers. (c,d) Dielectric values in the bulk hydration shells using AMBER99SB and OPLSAA charges. The legend to these plots mentions the optimal values of the input parameters to these models (also see [tblS Tables 5](#) and [6tbl6](#)).

dielectric ($\epsilon_{\text{solvent}}$). In contrast, the traditional model assigns a value of 1 ($=\epsilon_{\text{protein}}$). These observations indicate that Gaussian-based models reflect the capacity of internal cavities to hydrate. Simultaneously, they also capture the restricted mobility of the water molecules visiting these cavities as pointed out in our analyses of the explicit solvent simulations. Furthermore, they reflect the heterogeneity of the tempo-spatial properties observed in the bulk of the solvent by assigning gradually

increasing dielectric values to the solvent regions as a function of distance from the surface of the protein (Figure 9c,d). This gradual transition has a visible contrast from the abrupt change in dielectric ensued by the traditional model at the solute–solvent interface, the difference being noticeable from looking at the distribution of values in shells 1 and 2.

It is deducible from these observations that the Gaussian-based models can capture the effects of the increased residence

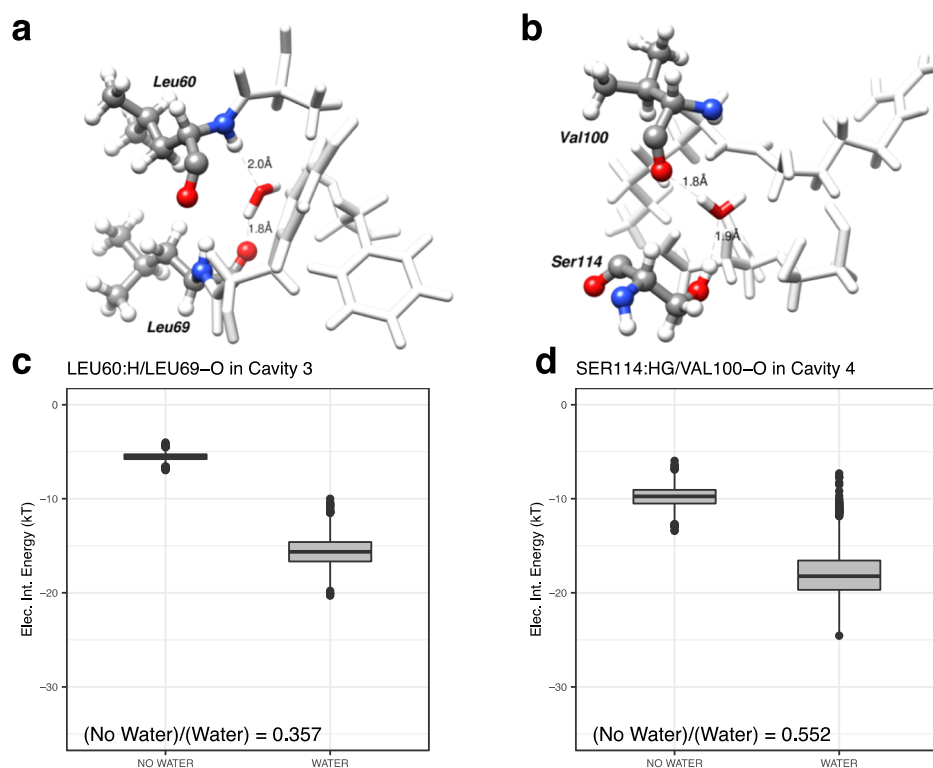


Figure 10. Effect of cavity water on the pairwise electrostatic interaction energy. (a,b) Illustration showing the water present in cavity 3 and cavity 4 from which the two atom pairs were chosen. In cavity 3, the interaction between Leu60:H and Leu69:O was computed, and in cavity 4, that between Ser114:HG and Val100:O was computed. In each case, the water, shown in licorice representation, forms separate hydrogen bonds with these atoms. For both cavities, the other cavity lining residues are shown in white and the residues L60, L69, V100, and S114 are displayed in “ball and stick” representation and colored differently to add emphasis. (c,d) Box plots showing the distribution of the pairwise electrostatic energies in kT of the atom pairs in the presence (“WATER”) and absence (“NO WATER”) of water.

and rotational relaxation times, the reduced dipolar fluctuations of the water molecules close to the interface, and the transition to the bulk properties farther away from it. These alterations in the properties of the bulk water, purely due to the proximity to protein atoms, result in reduced polarizability in that locality and a restricted ability to screen electrostatic forces, which manifest in terms of reduced permittivity. Therefore, through the use of these models, a qualitative match of the explicit model phenomena can be established using an implicit solvent model without losing accuracy in the prediction of polar solvation free energies. Our analyses of the dielectric distribution rendered by these models highlight the advantages of using a heterogeneous dielectric distribution model over the traditional model, in addition to being more realistic in its depictions of the solvated system.^{29,34}

3.7. Limitations of Applicability of the Macroscopic Dielectric Effects on the Microscopic Scale. Dielectric permittivity is a macroscopic quantity. It presumes that the medium dipoles (permanent or induced) are much smaller than the volume of the system where the permittivity is assessed. When one deals with microscopic objects such as biological macromolecules and the activity of dipoles (water, side chain, and backbone dipoles) in Angstrom length scales, the concept of permittivity arising from dipoles may diverge from the statistical mechanical treatment. First, the number of dipoles is too few to provide a statistically meaningful rendition of permittivity. Second, other short-range forces are at par with the electrostatic forces. To illustrate this point, the effect of cavity water on the pairwise electrostatic interaction energy of residue pairs was examined. From a macroscopic point of view, it is expected that

the presence of water molecules (water dipole) will reduce pairwise interactions resulting in the so-termed screening effect. However, calculations of energies at these length scales reveal something contrary but accurate.

Two pairs of atoms from two different cavities were chosen based on our observations regarding their participation in forming hydrogen bonds with the cavity waters. The goal was to derive their pairwise interaction energy and examine the effects of cavity water on it. Each pair consisted of an acceptor and a hydrogen covalently bonded to a donor with a priori knowledge that water was very likely mediating their interaction. As a result of this choice, the atoms were oppositely charged if not identical in magnitude. The atom pairs used were (a) the amide hydrogen of Leu60 and the backbone carboxyl oxygen of Leu69 from “cavity 3” and (b) the hydroxyl group hydrogen of Ser114 and the backbone carboxyl oxygen of Val100 from “cavity 4”. All of these atoms were found to form hydrogen bonds with water (Figure S1) and placed diametrically across their corresponding cavity (Figure 10). The average distance between these pairs, as seen in the simulations, was 4.55 and 3.94 Å, respectively.

Pairwise energy calculations of these residues were performed for each of the snapshots obtained from a single MD trajectory using the FRC module of DelPhi (Section 2.6). As stated, the objective was to assess the effect of cavity water on these energies and the focus was on elucidating their screening capacities of the protein atom interactions. Because dielectric effects are only sensible in the context of an ensemble, the interaction energies were computed for every snapshot of an MD trajectory with and without the cavity-bound water. At each snapshot, the

orientation of the cavity-bound water molecules was kept as seen in the trajectory.

The results of these analyses are shown in Figure 10. The stabilization of the pairwise interaction in the presence of water molecules is evident as the interaction energy is consistently more negative than that obtained in the absence of water. Contrary to what one might expect of water in the implicit solvent model and also to what we have stated so far, these results indicate that the mere presence of water does not result in screening of the strength of any electrostatic interaction. Instead, it appears to enhance the interaction of the cavity atoms through hydrogen bonds. As the water occupies a cavity and wiggles in that space under the effect of the steric factors imposed by the cavity atoms at the same time, it tries to engage in favorable hydrogen bond interactions. This happening in our case is evident from the nonzero number of hydrogen bonds formed between cavity waters and residues lining it (Figure 4). The restricted rotational motion is also evident from the long dipole orientation relaxation times of the cavity-bound water (Table 3). These favorable interactions are necessary as they compensate for the desolvation penalty paid by the cavity water, as it has less mobility (translational and orientational) than if it was in the bulk solvent.

The abovementioned two examples should not be considered as proof that macroscopic approaches cannot be applied to protein electrostatics. One should be careful in applying macroscopic approaches, depending on the question being asked. Our investigation suggests that if one wants to model pairwise interaction between two charges mediated by a dipole (a water dipole in this case) and the distance between charges is comparable with the length of the dipole, the macroscopic approach will not work and the water must be explicitly taken into account. This has significant implications for modeling electron transfer between closely situated donor/acceptor sites. Additionally, the effect of the same water in the same cavity on the rest of the protein electrostatics can be modeled via a macroscopic approach, as indicated by the results regarding polar solvation energy and the value of the dielectric permittivity assigned by the Gaussian-based method.

4. CONCLUSIONS

The primary focus of this work is to provide further support of the concept of modeling dielectric permittivity in molecular biology via a Gaussian-based smooth dielectric function. From a modeling standpoint, our Gaussian-based approach is centered on two key factors: (a) inhomogeneity in the distribution of charged, polar, and nonpolar residues in the protein's volume and (b) the difference in the ability of bulk water versus cavity and interfacial water to polarize (re-orient dipoles under an applied electric field). Although the first factor was discussed and results were presented in our previous work,^{1,29,31,34} the second one has not been directly discussed so far.

This work focused on a particular biomolecule: the IL-1 β protein (PDB: 2NVH), whose cavity hydration properties have attracted research interest in the past.^{36,37} Using two different force-field and water model combinations (AMBER99SB/TIP3P and OPLSAA/TIP4P), six independent MD simulations (each 30 ns long) were conducted and the trajectory was used to observe the effect of localization of a water molecules on its temporal and spatial properties. The temporal properties investigated were the mean residence times and the dipole orientation relaxation times, and the spatial property investigated was the average fluctuation of the dipole moment. We

have demonstrated that water molecules in the cavity have much longer residence and dipole orientation relaxation times than the bulk water, a difference of at least an order of magnitude. We have also shown that the fluctuation in the dipole moment is fast-paced in the bulk region and very restricted (and unpredictable) in the cavities. Our results also show that the protein charges perturb bulk water behavior noticeably in the nearest two hydration shells. Overall, data from explicit water simulations suggest that heterogeneity of water is microscopically found and that waters in cavities and near protein surfaces have much less ability to re-orient compared with the bulk water. This is considered to be a further justification of the applicability of the Gaussian-based approach to model protein electrostatics.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.0c00151>.

Volume and identifiers of the water molecules and the protein residues present in the five cavities in the crystal structure of IL-1 β protein; hydrogen-bonding sites in the protein cavities; optimal input parameters to a dielectric model based on their signed error; and optimal combination of ϵ_{ref} and σ for the Gaussian/super-Gaussian dielectric models (PDF)

Example of long-term residence of a water molecule in a cavity. The video shows a water molecule (in "Licorice" representation) in "cavity 4" as seen in an MD simulation trajectory with the AMBER99SB force field and TIP3P model of water. Cavity 4 was found to be occupied by only one water molecule for the majority of the trajectory. The residues lining this cavity are shown in "CPK" representation and are colored according to their type—"green" for polar, "white" for nonpolar, "blue" for basic, and "red" for acidic (MPG)

Example of short-term residence of water molecules in a cavity. This video shows multiple water molecules (in "Licorice" representation) that visit cavity 2 (a larger cavity than cavity 4) during the course of an MD simulation performed using the AMBER99SB force field and TIP3P model of water. It illustrates, in addition to a higher occupancy of water in this cavity, the process of the exchange of water molecules between the cavity and the bulk. The residues lining this cavity are shown in "CPK" representation and are colored according to their type—"green" for polar, "white" for nonpolar, "blue" for basic, and "red" for acidic (MPG)

Water in the vicinity of a hydrophilic amino acid. The video shows the movement of all the water molecules (in "Licorice" representation) that, at some point in the course of 20 ns long MD simulation, interacted with LYS138—a solvent exposed residue. In the trajectory, the frequently varying position and orientation of these water molecules reflect their high mobility and high exchange rate with regions farther away from the surface of the protein. This is in contrast to the mobility and exchange rate seen in the cavities (MP4)

■ AUTHOR INFORMATION

Corresponding Authors

Arghya Chakravorty – Computational Biophysics and Bioinformatics, Department of Physics and Astronomy, Clemson

University, Clemson, South Carolina 29634, United States;
✉ orcid.org/0000-0002-4467-1135; Email: arghyac@g.clemson.edu

Emil Alexov – Computational Biophysics and Bioinformatics,
Department of Physics and Astronomy, Clemson University,
Clemson, South Carolina 29634, United States; ✉ orcid.org/0000-0001-5346-0156; Email: ealexov@clemson.edu

Authors

Shailesh Panday – Computational Biophysics and Bioinformatics,
Department of Physics and Astronomy, Clemson University,
Clemson, South Carolina 29634, United States

Swagata Pahari – Computational Biophysics and Bioinformatics,
Department of Physics and Astronomy, Clemson University,
Clemson, South Carolina 29634, United States

Shan Zhao – Department of Mathematics, University of Alabama,
Tuscaloosa, Alabama 35487, United States

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jcim.0c00151>

Author Contributions

A.C. and S.P. prepared the system and ran the molecular dynamics simulations. A.C. and S.P. wrote modules in DelPhi. A.C., S.Z., and E.A. contributed to the development of the theory. A.C. and E.A. prepared the manuscript. All authors have given approval for the final version of the manuscript.

Funding

The work was supported by a grant from NSF, DMS 1812597.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge Clemson University for the generous allotment of computer time on its Palmetto cluster.

REFERENCES

- (1) Li, L.; Li, C.; Zhang, Z.; Alexov, E. On the Dielectric “Constant” of Proteins: Smooth Dielectric Function for Macromolecular Modeling and Its Implementation in DelPhi. *J. Chem. Theory Comput.* **2013**, *9*, 2126–2136.
- (2) Grant, J. A.; Pickup, B. T.; Nicholls, A. A Smooth Permittivity Function for Poisson-Boltzmann Solvation Methods. *J. Comput. Chem.* **2001**, *22*, 608–640.
- (3) Kuntz, I. D.; Kauzmann, W. Hydration of Proteins and Polypeptides. *Adv. Protein Chem.* **1974**, *28*, 239–345.
- (4) Chopra, G.; Summa, C. M.; Levitt, M. Solvent Dramatically Affects Protein Structure Refinement. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 20239–20244.
- (5) Persson, F.; Söderhjelm, P.; Halle, B. The Spatial Range of Protein Hydration. *J. Chem. Phys.* **2018**, *148*, 215104.
- (6) Ghosh, R.; Banerjee, S.; Hazra, M.; Roy, S.; Bagchi, B. Sensitivity of Polarization Fluctuations to the Nature of Protein-Water Interactions: Study of Biological Water in Four Different Protein-Water Systems. *J. Chem. Phys.* **2014**, *141*, 22D531.
- (7) Nandi, N.; Bagchi, B. Dielectric Relaxation of Biological Water. *J. Phys. Chem. B* **1997**, *101*, 10954–10961.
- (8) Hua, L.; Huang, X.; Zhou, R.; Berne, B. J. Dynamics of Water Confined in the Interdomain Region of a Multidomain Protein. *J. Phys. Chem. B* **2006**, *110*, 3704–3711.
- (9) Rick, S. W.; Stuart, S. J.; Berne, B. J. Dynamical Fluctuating Charge Force Fields: Application to Liquid Water. *J. Chem. Phys.* **1994**, *101*, 6141–6156.
- (10) Persson, F.; Söderhjelm, P.; Halle, B. How Proteins Modify Water Dynamics. *J. Chem. Phys.* **2018**, *148*, 215103.

- (11) Qin, Y.; Wang, L.; Zhong, D. Dynamics and Mechanism of Ultrafast Water-Protein Interactions. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 8424–8429.
- (12) Levitt, M.; Sharon, R. Accurate Simulation of Protein Dynamics in Solution. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 7557–7561.
- (13) Levy, Y.; Onuchic, J. N. Water Mediation in Protein Folding and Molecular Recognition. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 389–415.
- (14) Meyer, T.; Kieseritzky, G.; Knapp, E.-W. Electrostatic PKa Computations in Proteins: Role of Internal Cavities. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 3320–3332.
- (15) Kandori, H. Role of Internal Water Molecules in Bacteriorhodopsin. *Biochim. Biophys. Acta, Bioenerg.* **2000**, *1460*, 177–191.
- (16) Alexov, E. G.; Gunner, M. R. Calculated Protein and Proton Motions Coupled to Electron Transfer: Electron Transfer from QA - to QB in Bacterial Photosynthetic Reaction Centers †. *Biochemistry* **1999**, *38*, 8253–8270.
- (17) Alexov, E.; Miksovska, J.; Baciou, L.; Schiffer, M.; Hanson, D. K.; Sebban, P.; Gunner, M. R. Modeling the Effects of Mutations on the Free Energy of the First Electron Transfer from QA - to QB in Photosynthetic Reaction Centers. *Biochemistry* **2000**, *39*, 5940–5952.
- (18) Gunner, M. R.; Alexov, E. A Pragmatic Approach to Structure Based Calculation of Coupled Proton and Electron Transfer in Proteins. *Biochim. Biophys. Acta, Bioenerg.* **2000**, *1458*, 63–87.
- (19) Ikura, T.; Urakubo, Y.; Ito, N. Water-Mediated Interaction at a Protein-Protein Interface. *Chem. Phys.* **2004**, *307*, 111–119.
- (20) Baron, R.; Setny, P.; McCammon, J. A. Water in Cavity-Ligand Recognition. *J. Am. Chem. Soc.* **2010**, *132*, 12091–12097.
- (21) Zhou, Y.; Morais-Cabral, J. H.; Kaufman, A.; MacKinnon, R. Chemistry of Ion Coordination and Hydration Revealed by a K⁺ Channel-Fab Complex at 2.0 Å Resolution. *Nature* **2001**, *414*, 43–48.
- (22) Bhattacharyya, K. Nature of Biological Water: A Femtosecond Study. *Chem. Commun.* **2008**, 2848–2857.
- (23) Qiu, W.; Zhang, L.; Okobiah, O.; Yang, Y.; Wang, L.; Zhong, D.; Zewail, A. H. Ultrafast Solvation Dynamics of Human Serum Albumin: Correlations with Conformational Transitions and Site-Selected Recognition. *J. Phys. Chem. B* **2006**, *110*, 10540–10549.
- (24) Decherchi, S.; Colmenares, J.; Catalano, C. E.; Spagnuolo, M.; Alexov, E.; Rocchia, W. Between Algorithm and Model: Different Molecular Surface Definitions for the Poisson-Boltzmann Based Electrostatic Characterization of Biomolecules in Solution. *Commun. Comput. Phys.* **2013**, *13*, 61–89.
- (25) Song, X. An Inhomogeneous Model of Protein Dielectric Properties: Intrinsic Polarizabilities of Amino Acids. *J. Chem. Phys.* **2002**, *116*, 9359–9363.
- (26) Castner, E. W.; Fleming, G. R.; Bagchi, B.; Maroncelli, M. The Dynamics of Polar Solvation: Inhomogeneous Dielectric Continuum Models. *J. Chem. Phys.* **1988**, *89*, 3519–3534.
- (27) Hazra, T.; Ahmed Ullah, S.; Wang, S.; Alexov, E.; Zhao, S. A Super-Gaussian Poisson-Boltzmann Model for Electrostatic Free Energy Calculation: Smooth Dielectric Distribution for Protein Cavities and in Both Water and Vacuum States. *J. Math. Biol.* **2019**, *79*, 631.
- (28) Goh, G. B.; García-Moreno E, B.; Brooks, C. L. The High Dielectric Constant of Staphylococcal Nuclease Is Encoded in Its Structural Architecture. *J. Am. Chem. Soc.* **2011**, *133*, 20072–20075.
- (29) Chakravorty, A.; Jia, Z.; Peng, Y.; Tajiri, N.; Wang, L.; Alexov, E. Gaussian-Based Smooth Dielectric Function: A Surface-Free Approach for Modeling Macromolecular Binding in Solvents. *Front. Mol. Biosci.* **2018**, *5* (). <https://doi.org/10.3389/fmolb.2018.00025>.
- (30) Li, C.; Jia, Z.; Chakravorty, A.; Pahari, S.; Peng, Y.; Basu, S.; Koirala, M.; Panday, S. K.; Petukh, M.; Li, L.; Alexov, E. DelPhi Suite: New Developments and Review of Functionalities. *J. Comput. Chem.* **2019**, *40*, 2502–2508.
- (31) Wang, L.; Li, L.; Alexov, E. PKa Predictions for Proteins, RNAs, and DNAs with the Gaussian Dielectric Function Using DelPhi PKa. *Proteins: Struct., Funct., Bioinf.* **2015**, *83*, 2186–2197.

- (32) Peng, Y.; Sun, L.; Jia, Z.; Li, L.; Alexov, E. Predicting Protein-DNA Binding Free Energy Change upon Missense Mutations Using Modified MM/PBSA Approach: SAMPDI Webserver. *Bioinformatics* **2018**, *34*, 779.
- (33) Peng, Y.; Alexov, E. Computational Investigation of Proton Transfer, PKa Shifts and PH-Optimum of Protein-DNA and Protein-RNA Complexes. *Proteins: Struct., Funct., Bioinf.* **2017**, *85*, 282.
- (34) Li, L.; Li, C.; Alexov, E. On the Modeling of Polar Component of Solvation Energy Using Smooth Gaussian-Based Dielectric Function. *J. Theor. Comput. Chem.* **2014**, *13*, 1440002.
- (35) Chakravorty, A.; Jia, Z.; Li, L.; Zhao, S.; Alexov, E. Reproducing the Ensemble Average Polar Solvation Energy of a Protein from a Single Structure: Gaussian-Based Smooth Dielectric Function for Macromolecular Modeling. *J. Chem. Theory Comput.* **2018**, *14*, 1020–1032.
- (36) Quillin, M. L.; Wingfield, P. T.; Matthews, B. W. Determination of Solvent Content in Cavities in IL-1 β Using Experimentally Phased Electron Density. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19749–19753.
- (37) Ernst, J.; Clubb, R.; Zhou, H.; Gronenborn, A.; Clore, G. Demonstration of Positionally Disordered Water within a Protein Hydrophobic Cavity by NMR. *Science* **1995**, *267*, 1813–1817.
- (38) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701.
- (39) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- (40) Chakravorty, A.; Jia, Z.; Li, L.; Alexov, E. A New DelPhi Feature for Modeling Electrostatic Potential around Proteins: Role of Bound Ions and Implications for Zeta-Potential. *Langmuir* **2017**, *33*, 2283–2295.
- (41) Makarov, V. A.; Andrews, B. K.; Smith, P. E.; Pettitt, B. M. Residence Times of Water Molecules in the Hydration Sites of Myoglobin. *Biophys. J.* **2000**, *79*, 2966–2974.
- (42) Schoenborn, B.; Garcia, A.; Knott, R. Hydration in Protein Crystallography. *Prog. Biophys. Mol. Biol.* **1995**, *64*, 105–119.
- (43) Yeh, Y.-I.; Mou, C.-Y. Orientational Relaxation Dynamics of Liquid Water Studied by Molecular Dynamics Simulation. *J. Phys. Chem. B* **1999**, *103*, 3699–3705.
- (44) Bizzarri, A. R.; Cannistraro, S. Molecular Dynamics of Water at the Protein-Solvent Interface. *J. Phys. Chem. B* **2002**, *106*, 6617–6633.
- (45) Panday, S. K.; Shashikala, M. H. B.; Koirala, M.; Pahari, S.; Chakravorty, A.; Peng, Y.; Li, L.; Jia, Z.; Li, C.; Alexov, E. Modeling Electrostatics in Molecular Biology: A Tutorial of DelPhi and Associated Resources [Article v1.0]. *Living J. Comput. Mol. Sci.* **2019**, *1*, 10841.
- (46) Neumann, M. Dipole Moment Fluctuation Formulas in Computer Simulations of Polar Systems. *Mol. Phys.* **1983**, *50*, 841–858.
- (47) Gereben, O.; Pusztai, L. On the Accurate Calculation of the Dielectric Constant from Molecular Dynamics Simulations: The Case of SPC/E and SWM4-DP Water. *Chem. Phys. Lett.* **2011**, *507*, 80–83.
- (48) Simonson, T.; Perahia, D. Internal and Interfacial Dielectric Properties of Cytochrome c from Molecular Dynamics in Aqueous Solution. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1082–1086.
- (49) Kusalik, P. G.; Svishchev, I. M. The Spatial Structure in Liquid Water. *Science* **1994**, *265*, 1219.
- (50) van der Spoel, D.; van Maaren, P. J. The Origin of Layer Structure Artifacts in Simulations of Liquid Water. *J. Chem. Theory Comput.* **2006**, *2*, 1.
- (51) Luise, A.; Falconi, M.; Desideri, A. Molecular Dynamics Simulation of Solvated Azurin: Correlation between Surface Solvent Accessibility and Water Residence Times. *Proteins: Struct., Funct., Genet.* **2000**, *39*, 56–67.
- (52) Makarov, V.; Pettitt, B. M.; Feig, M. Solvation and Hydration of Proteins and Nucleic Acids: A Theoretical View of Simulation and Experiment. *Acc. Chem. Res.* **2002**, *35*, 376–384.