Water Dynamics at Protein–Protein Interfaces: A Molecular Dynamics Study of Virus–Host Receptor Complexes.

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · NOVEMBER 2014

Impact Factor: 3.3 · DOI: 10.1021/jp5089096 · Source: PubMed

3 AUTHORS, INCLUDING:

Priyanka Dutta
University of South Florida
3 PUBLICATIONS **1** CITATION

SEE PROFILE

Mohsen Botlani
University of South Florida
11 PUBLICATIONS **17** CITATIONS

SEE PROFILE

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

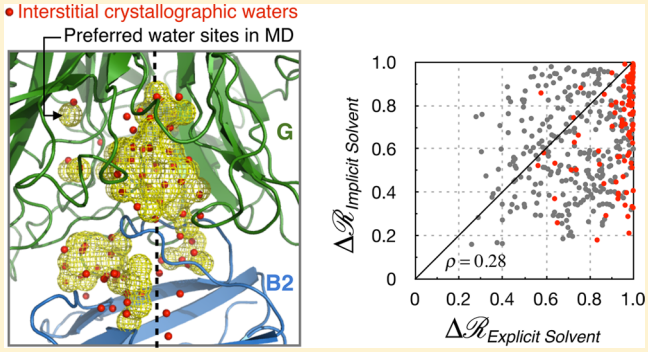
pubs.acs.org/JPCB

Water Dynamics at Protein−Protein Interfaces: Molecular Dynamics Study of Virus−Host Receptor Complexes

Priyanka Dutta, Mohsen Botlani, and Sameer Varma*

Department of Cell Biology, Microbiology and Molecular Biology, University of South Florida, Tampa, Florida 33620, United States

ABSTRACT: The dynamical properties of water at protein−water interfaces are unlike those in the bulk. Here we utilize molecular dynamics simulations to study water dynamics in interstitial regions between two proteins. We consider two natural protein−protein complexes, one in which the Nipah virus G protein binds to cellular ephrin B2 and the other in which the same G protein binds to ephrin B3. While the two complexes are structurally similar, the two ephrins share only a modest sequence identity of ∼50%. X-ray crystallography also suggests that these interfaces are fairly extensive and contain exceptionally large amounts of waters. We find that while the interstitial waters tend to occupy crystallographic sites, almost all waters exhibit residence times of less than hundred picoseconds in the interstitial region. We also find that while the differences in the sequence of the two ephrins result in quantitative differences in the dynamics of interstitial waters, the trends in the shifts with respect to bulk values are similar. Despite the high wetness of the protein−protein interfaces, the dynamics of interstitial waters are considerably slower compared to the bulk—the interstitial waters diffuse an order of magnitude slower and have 2−3 fold longer hydrogen bond lifetimes and 2−1000 fold slower dipole relaxation rates. To understand the role of interstitial waters, we examine how implicit solvent models compare against explicit solvent models in producing ephrin-induced shifts in the G conformational density. Ephrin-induced shifts in the G conformational density are critical to the allosteric activation of another viral protein that mediates fusion. We find that in comparison with the explicit solvent model, the implicit solvent model predicts a more compact G−B2 interface, presumably because of the absence of discrete waters at the G−B2 interface. Simultaneously, we find that the two models yield strikingly different induced changes in the G conformational density, even for those residues whose conformational densities in the apo state are unaffected by the treatment of the bulk solvent. Together, these results show that the explicit treatment of interstitial water molecules is necessary for a proper description of allosteric transitions.



INTRODUCTION

The dynamical properties of water at protein surfaces have been studied extensively using both experimental and computational techniques.[1−14] Experimental techniques typically include nuclear magnetic resonance spectroscopy, magnetic resonance dispersion, dielectric relaxation, neutron scattering as well as time-resolved fluorescence. These techniques measure the time scales of rotational and translation dynamics of interfacial waters, but they detect different physical phenomena and, consequently, are subject to different interpretations. Computational techniques primarily constitute molecular dynamics simulations in which interatomic forces are obtained by treating polarization effects implicitly or explicitly, or at the quantum mechanical level, and the choice of the specific treatment is determined by the system size, observation time and spatial resolution. Together, these techniques allow observation of a wide range of time and length scales. While a complete picture of the complex interactions of hydration water molecules that accounts for all the data is still lacking, there is little doubt that the dynamics of interfacial waters are statistically different from those in the bulk. In fact, there is also now growing consensus that water molecules at the protein−water interface translate and rotate slower compared to the bulk.[15−48] Additionally, the extent of the shift depends on the complex interplay of protein surface chemistry and topology, that is, whether waters are present in cavities or near hydrophobic/hydrophilic moieties, or near backone/side chain groups, or near concave/convex surfaces.

In this work, we extend such studies to systems where water molecules occupy interstitial regions between two proteins. Specifically, we consider two natural protein−protein complexes, both of which are formed independently during the fusion of Nipah viruses with host cells. Nipah are emerging zoonotic pathogens that cause illness and fatality in livestock and humans.[49−54] In one of the two protein−protein complex, the Nipah virus G protein binds to cellular ephrin B2, and in the other, the same G protein binds to ephrin B3. These two ephrins share only a modest sequence identity of ∼50%, even at the protein−protein interface. The interaction of G with these ephrins constitute the first step in Nipah infection.[55−57]

Received: September 3, 2014
Revised: November 7, 2014
Published: November 24, 2014

dx.doi.org/10.1021/jp5089096 | J. Phys. Chem. B 2014, 118, 14795−14807

X-ray crystallography[58,59] suggests that the interfaces of G with ephrins B2 and B3 are extensive, with interface areas greater than 2500 Å^2. In addition, the X-ray structure of the G−B2 complex contains a large number of interstitial waters, which is, in fact, three times greater than the average numbers of waters found in other protein−protein interfaces of comparable sizes.[60,61] Does the resolution of waters in X-ray structures generally imply that their dynamics are sluggish, or is the interface between G and ephrins sufficiently wet for the dynamics of interstitial waters to be similar to the bulk? In general, how do the dynamics of water at protein−protein interfaces compare against the dynamics of water at protein−water interfaces? We address these questions by carrying out and analyzing all-atom molecular dynamics simulations of the G−B2 and G−B3 complexes. We find that while the interstitial waters tend to occupy crystallographic sites, most waters exhibit residence times of less than hundred picoseconds in the interstitial region. In addition, despite the exceptionally high wetness of the two protein−protein interfaces, the dynamics of interstitial waters are slow, and, to some extent, comparable to the dynamics of water at protein−water interfaces.

Crystallographically resolved waters are generally considered to be functionally important.[62−65] But what specific physiological role do the crystallographically resolved interstitial waters serve in the G−ephrin complexes? The binding of ephrin to G causes G to activate another viral protein, F.[55−59] Upon activation, F mediates virus-host membrane fusion. Since ephrins and F bind to G at mutually exclusive sites, the effect of ephrin-binding must transduce to the F-binding site of G to activate F.[55−59,66] According to the thermodynamic view of allosteric signaling,[66−74] this ephrin-binding response, or F-activating signal, is contained within the changes in the conformational density of G brought about by ephrin binding. Mathematically, if $\mathcal{R}_{apo}$ and $\mathcal{R}_{bound}$ represent, respectively, the conformational densities of G in its apo and ephrin-bound states, then the F-activating signal is contained within the ephrin-induced conformational density shift $\Delta\mathcal{R} := \mathcal{R}_{apo} \rightarrow \mathcal{R}_{bound}$. Now if the waters at the G−ephrin interface were to contribute to the allosteric activation of F, then they must contribute to $\Delta\mathcal{R}$.

To examine whether the interstitial waters contribute to $\Delta\mathcal{R}$, we determine $\Delta\mathcal{R}$ from explicit solvent simulations and compare them quantitatively to the $\Delta\mathcal{R}$ obtained from a separate set of implicit solvent simulations. Such comparative studies have been conducted previously, although in the context of protein folding.[75] While several protein−protein complexes have been modeled successfully using implicit solvent models,[76−79] we expect explicit and implicit solvent models to yield different $\Delta\mathcal{R}$ because (a) the two models describe bulk solvent differently, and (b) implicit solvent models do not account for the discrete nature of water molecules at the G−ephrin interface. Given the high numbers of waters at the G−ephrin interface, the lack of their specific volumes in the implicit solvent model can alter the G−ephrin interface. In addition, since several crystallographic waters in the G−B2 interface are located at regions where they can potentially hydrogen bond simultaneously with both proteins,[58,59,61] their absence could directly alter G−B2 binding modes. The absence of discrete waters at the G−B2 interface could, therefore, lead to altered $\Delta\mathcal{R}$. We find that the implicit solvent model indeed predicts a different, more compact G−B2 interface compared to the explicit solvent model. Simultaneously, we find that $\Delta\mathcal{R}$ estimated from explicit solvent simulations are statistically different from those obtained from implicit solvent simulations, even for those amino acids whose $\mathcal{R}_{apo}$ are unaffected by the treatment of the bulk solvent. Together, these results show that the explicit treatment of interstitial water molecules is necessary for a proper description of allosteric transitions, supporting the hypothesis that interstitial waters contribute directly to the allosteric regulation of viral fusion.

METHODS

Molecular Dynamics Simulations. *Explicit Solvent.* We carry out three separate explicit solvent MD simulations of the Nipah G head domain in 150 mM NaCl solution. In one MD simulation, G is ligand free, and in the remaining two MD simulations, G is bound separately to ephrins B2 and B3. Note that while the simulations of the G−B2 and G−B3 complexes are used for analyzing the dynamics of interstitial waters, the purpose of simulating G in its ligand free state is to enable the estimation of ephrin-induced changes in the G conformational density. The starting conformations of the apo and B2 bound states are taken from the X-ray structure of the G−B2 complex (PDB ID: 2VSM).[58] Our choice for taking the starting conformation of the apo state of G from the X-ray structure of the G−B2 complex was motivated by the observation that the differences between the X-ray structures of G in the apo[59] and B2 bound states[58] are small (RMSD = 1.2 Å). More importantly, the lengths of N-termini of the G protein resolved in the two studies are different, and we did not want to bias our analysis of ephrin-induced changes from this difference. The coordinates of the 3 amino acids of G, P208, V209, and V210, that were not resolved by X-ray crystallography of the G−B2 complex were constructed using ModLoop.[80] The starting conformation for the B3 bound state is taken from the X-ray structure of the G−B3 complex (PDB ID: 3D12).[59] The algorithm PDB2PQR[81] is used for optimizing the positions of the protein hydrogen atoms in the gas phase. The *N*-acetyl-D-glucosamine chains of G were removed in both cases, as they are distant from the G−ephrin binding interface and are unlikely to influence the dynamics of interstitial water. The water molecules that were resolved in the X-ray structure of the G−B2 complex were retained. We note that while no water molecules were resolved in the X-ray structure of the G−B3 complex, perhaps due to its lower resolution, the sampling in the G−B3 simulation appears sufficient to allow for water equilibration. The length of the trajectory is over three orders in magnitude longer than the computed residence time of waters in the interstitial region. The MD unit cells corresponding to the apo, B2 bound, and B3 bound states of G contain a total of 30367, 40900, and 40666 water molecules, respectively.

All the three explicit solvent MD simulations are carried out under isobaric−isothermal boundary conditions, and using Gromacs version 4.5.3.[82] Temperature is maintained at 310 K using an extended ensemble approach[83,84] and with a coupling constant of 0.2 ps. An extended ensemble approach[85] is also used for maintaining pressure. Pressure is maintained at 1 bar using a coupling constant of 1 ps and a compressibility of 4.5×10^{-5} bar^{-1}. We also examine the effect of imposing alternative boundary conditions, including the canonical and micro-canonical boundary conditions, as discussed in the Results.

The protein and ions are described using OPLS-AA parameters,[86] and the water molecules are described using

dx.doi.org/10.1021/jp5089096 | J. Phys. Chem. B 2014, 118, 14795−14807

TIP4P parameters.[87] While the TIP4P water model is known to overestimate diffusion,[88] it reproduces bulk water reorientational processes observed in the experimental spectrum.[89] Note that while we do not model induced effects explicitly, which are important to both ionic interactions[90,91] and hydrogen bonding,[92−95] we discuss the implications of this approximation in the Results. Electrostatic interactions are computed using the particle mesh Ewald scheme[96] with a Fourier grid spacing of 0.15 nm, a sixth-order interpolation, and a direct space cutoff of 10 Å. van der Waals interactions are computed explicitly for interatomic distance up to 10 Å. Charge neutrality of the three MD unit cells are maintained by selecting appropriate differences between the numbers of Na^+ and Cl^- ions. The bonds in proteins are constrained using the P-LINCS algorithm,[97] and the geometries of the water molecules are constrained using SETTLE.[98] These constraints permit use of an integration time step of 2 fs. The motion of the center of mass is reset every 100 ps. We generate a 640 ns long trajectory of the unbound state of the G protein. The G−B2 and G−B3 complexes are simulated for 407 and 468 ns, respectively. Our analysis of water dynamics is based on the final 40 ns time frames of these trajectories. For the analysis portions of the simulations, the coordinates are saved at time intervals of 0.25 ps. Velocities and coordinates needed for the calculation of diffusion coefficients and hydrogen bond lifetimes, respectively, are saved at every integration time step of 2 fs, but for a shorter time frame of 256 ps. Because of the large sizes of these systems, each of the these 256 ps long trajectories require over 1 TB of storage space. The degree of sampling in coordinates and velocities appears sufficient, as indicated by the accompanying error analyses.

Implicit Solvent. We also carry out two separate implicit solvent MD simulations, one in which G is ligand free and the other in which G is complexed with ephrin B2. The purpose of these simulations is to estimate ephrin-induced changes in the G conformational density, but without accounting for the explicit contributions from water structure and dynamics. The starting conformations for these simulations are the same as those used in the explicit solvent simulation. The partial charges on the atoms and the connectivity between them are described using the OPLS-AA force field,[86] and their Born radii are computed at every time step using the Onufriev−Bashford−Case (OBC) scheme.[77] Both systems are simulated under isothermal conditions, where the temperature is maintained at 310 K using an extended ensemble approach[83,84] and with a coupling constant of 0.2 ps. In accordance with the OBC scheme, the solvent is described using a dielectric constant of 80, and the electrostatics and van der Waals interactions are truncated beyond interatomic distances of 24 Å. Also in accordance with the OBC scheme, the free energy associated with cavity formation is modeled as a linear function of the atomic surface area using a surface tension of 0.005 kcal/mol/Å^2. The bonds in proteins are constrained using the P-LINCS algorithm.[97] The G protein and the G−B2 complex are simulated for 1.1×10^6 integration time steps, and the first 0.1×10^6 time steps are left out as equilibration.

Quantifying Differences between Ensembles. The difference between two three-dimensional (3D) conformational ensembles is quantified using a method based on support vector machines.[66,99] This method returns a quantitative estimate for the difference between two ensembles that we refer to as the discriminability index (η). It is normalized and bounded, that is, $\eta \in [0,1)$, and it takes up a value closer to unity as the difference between the ensembles increases. This form of comparison is advantageous over a comparison of summary statistics obtained separately from the two ensembles as it avoids artifactual biases arising from information loss.[66,99−101] Another benefit of comparing ensembles over representative structures is that it circumvents the problem of selecting a representative conformation from a rugged potential energy surface. A further advantage of comparing ensembles directly is that the resulting quantification naturally embodies differences in conformational fluctuations.

The details of this method are described elsewhere,[66,99] and here we present an abbreviated version. Traditionally, support vector machines (SVMs) are utilized for binary classification.[102,104,105] The SVM classification function

$$F(\mathbf{r}) = \sum_{i=1}^{2m} \alpha_i y_i K(\mathbf{r}_i, \mathbf{r}) \tag{1}$$

is first trained on a set of instances $\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_{2m}$ whose group identities, $y_1, y_2, ..., y_{2m} = \{-1, +1\}$, are known. This classification function is then used for predicting the group identity y of an unclassified instance $\mathbf{r}$. To quantify the difference between two conformational ensembles, $\{\mathbf{r}^+\}_1^m$ and $\{\mathbf{r}^-\}_1^m$, the SVM is also trained to recognize the difference between them, where $\mathbf{r}$ now refers to coordinates in Euclidean space. However, instead of utilizing the trained function for predictive purposes, its underlying mathematical construct is used for obtaining a physically meaningful quantitative estimate for the difference between the two ensembles.

In eq 1, K is a nonlinear kernel function that satisfies the inner product $K(\mathbf{r}_i, \mathbf{r}_j) = \langle \phi(\mathbf{r}_i), \phi(\mathbf{r}_j) \rangle_R$ in a transformed feature space R. The function ϕ is a transformation from Euclidean space to a high-dimensional kernel Hilbert space R in which the conformational ensembles are separated. Note that a knowledge of the specific form of the function ϕ is not required so long as the kernel is an inner product in the transformed feature space.[103−105] We chose a Gaussian radial distribution function as the kernel, that is, $K(\mathbf{r}_i, \mathbf{r}_j) = \exp(-\gamma \|\mathbf{r}_i - \mathbf{r}_j\|^2)$, which satisfies the aforementioned condition.[106] This kernel is also chosen due to its stationarity and its performance in classification as compared to linear, polynomial, or sigmoidal kernels.[106] The parameter γ in the kernel, which has units of 1/Å^2, controls the width of the kernel, and thereby the smoothness of the underlying nonlinear classifier. It represents essentially the influence a given conformation has on its neighboring conformations. Smaller γ correspond to larger Gaussian widths, which imply a larger contribution of a conformation to the classification. The α_i in eq 1 are Lagrange multipliers that are obtained by maximizing the auxiliary function,

$$\sum_{i=1}^{2m} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(\mathbf{r}_i, \mathbf{r}_j) \tag{2}$$

under the constraints $0 \le \alpha_i \le C$ and $\sum \alpha_i y_i = 0 \; \forall i$, where C is a predefined upper limit on the magnitude of the Lagrange multipliers. Maximization of this auxiliary function[107,108] produces the α_i that maximize the separability between $\{\mathbf{r}^+\}_1^m$ and $\{\mathbf{r}^-\}_1^m$. The discriminability index η is defined as a function of these Lagrange multipliers. We assign $C = 10^2$ and $\gamma = 10^{-1}$ that minimizes the error between the computed and analytical values of η for discrete Gaussian distributions.[66,99] The mean

dx.doi.org/10.1021/jp5089096 | J. Phys. Chem. B 2014, 118, 14795−14807

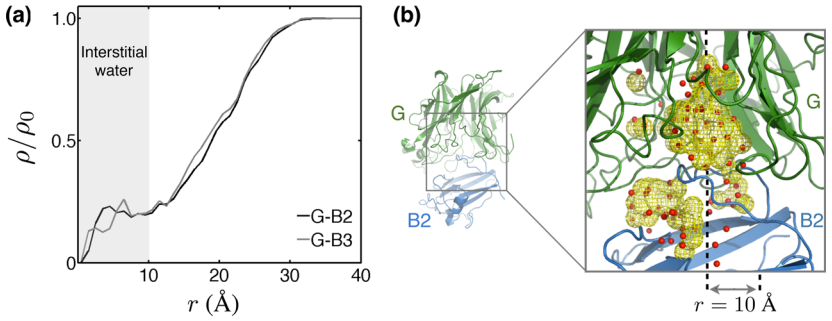


Figure 1. Water distribution in the interstitial regions of the G−ephrin complexes. (a) Normalized density of water (ρ/ρ_0) as a function of the perpendicular distance from the axes joining the geometric centers of G and ephrins. Note that only those waters are considered that lie within the right circular cylindrical region whose bounding disks contain the geometric centers of G and ephrins. (b) Partial view of the G−B2 complex showing a superimposition of the 69 interstitial waters resolved in the X-ray structure (red spheres), and the 65 highest occupancy regions observed in the MD simulation (yellow mesh). The axis of the right circular cylindrical region that connects the geometric centers of the G and B2 is drawn as a dashed black line.

absolute error between computed and analytical values of η is 3.2%.[66]

RESULTS AND DISCUSSION

To characterize the properties of water molecules present in the interfaces of the G−B2 and G−B3 protein complexes, we generate separate $\sim^1/_2$ μs long MD trajectories of the G−B2 and G−B3 complexes, respectively, in explicit solvent. Our analysis is based on the final 40 ns time frames of these trajectories. The initial portions of these trajectories were set aside to equilibrate water densities. This is particularly necessary in the case of the G−B3 complex in which the quality of the X-ray diffraction forbade resolution of water molecules. The extent of equilibration appears sufficient as it is more than three orders in magnitude longer than the computed residence times of waters in the interstitial region. We begin the analysis by defining a scheme to distinguish between interstitial and bulk waters.

Interstitial Waters. The interfaces formed between G and ephrins are uneven. Consequently, this rules out a definitive scheme for discriminating between interstitial and bulk waters. In this work, we utilize a scheme based on density distribution. Figure 1a shows the water density variation as a function of the perpendicular distance from the axes joining the geometric centers of G and ephrins. Note that in this calculation we consider only those waters that lie within the right circular cylindrical regions bounded by disks containing the geometric centers of G and ephrins. The heights of these cylinders are the distances between the geometric centers of the G and ephrins. The average distances between the geometric centers of G and ephrins are 32.8 ± 0.2 and 34.7 ± 0.2 Å, respectively, for the G−B2 and G−B3 complexes. We define the interstitial regions as 20 Å wide cylinders, as radial distances of 10 Å correspond roughly to the first inflection points in the density distribution functions. We note that there are no Na^+ or Cl^- ions present in these interstitial regions, as determined from their respective radial distribution functions. The average numbers of waters in the interstitial regions of the G−B2 and G−B3 complexes are 65.3 ± 4.0 and 67.6 ± 3.9, respectively. These averages are comparable to the 69 waters resolved in this interstitial region in the X-ray structure of the G−B2 complex. Furthermore, the 65 highest occupancy regions in the MD simulation of the G−B2 complex coincide well with the positions of the water molecules resolved in in the X-ray structure (Figure 1b).

Diffusion Coefficients. Figure 2 illustrates the integrated form of the velocity autocorrelation of interstitial waters:

$$D(\tau) = 1/3 \int_0^\tau \langle\langle \mathbf{v}(0)\cdot\mathbf{v}(t)\rangle\rangle \, dt \quad (3)$$

The double angular brackets denote averages computed over the ensemble as well as the number of waters in the interstitial region. The diffusion coefficient of the interstitial waters is obtained from the Green−Kubo relationship[109] as a limiting case $D = D(\tau \to \infty)$. The diffusion coefficient obtained from periodic systems, however, needs to be corrected for finite size effects. According to the hydrodynamic theory of diffusion,[110,111] the leading order correction to the diffusion coefficient obtained from a cubic periodic cell is inversely proportional to the length of the cell,

$$D_{self} = D + k_b T\xi/6\pi\eta L = D + \alpha/L \quad (4)$$

In the equation above, η is the viscosity, ξ = 2.837 is a constant, and L is the length of the cubic cell. The higher order corrections to eq 4 are relatively small. Instead of computing the viscosity needed for estimating the correction, we estimate the correction factor α empirically.[111,112] To accomplish this, we compute the diffusion coefficient of bulk water D from three different cubic cells of lengths 24.83, 31.85, and 40.70 Å (Figure 3). In all of the three simulations, water densities are maintained at 0.987 kg/dm^3 and the temperature is maintained at 310 K. In addition, D is estimated from velocity autocorrelations. While the slope of the line fitted to these data points yields α, the ordinate intercept yields the diffusion

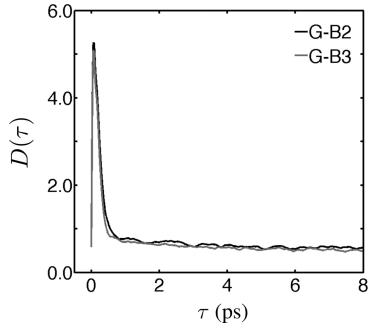


Figure 2. Integrated form of the velocity autocorrelation of waters, $D(\tau)$, occupying the interstitial regions in the G−B2 and G−B3 protein complexes.

dx.doi.org/10.1021/jp5089096 | J. Phys. Chem. B 2014, 118, 14795−14807

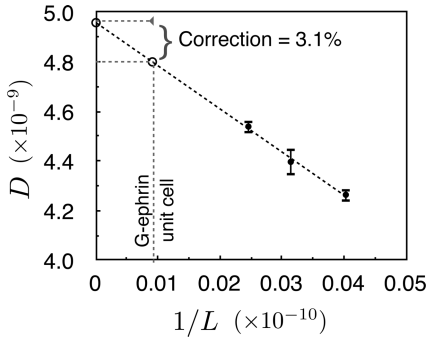


Figure 3. Diffusion coefficient D of water estimated from cubic periodic cells with different lengths L. D has units of m^2/s and $1/L$ has units of m^{-1}. The self-diffusion coefficient of water corrected for finite size effects is the intercept on the ordinate, $D_{self} = 4.96 \times 10^{-9}$ m^2/s.

coefficient of bulk water corrected for finite size effects. We find $D_{self} = 4.96 \times 10^{-9}$ m^2/s. For the system sizes we use for simulating the G−ephrin complexes, we find that the finite size correction is 3.1% of D. After accounting for this correction, we find that the diffusion coefficients of interstitial waters in the G−B2 and G−B3 complexes are $0.62 \pm 0.07 \times 10^{-9}$ and $0.57 \pm 0.16 \times 10^{-9}$ m^2/s, respectively. The standard deviations are obtained by dividing the trajectory into four separate blocks. The estimated values of the diffusion coefficients of interstitial waters are almost an order in magnitude smaller than the computed bulk water diffusion coefficient. This shift is also significantly larger than the mismatch between the computed and experimental estimates for the diffusion coefficient of bulk water. The experimental estimate for the diffusion coefficient of bulk water at 310 K is 3.1×10^{-9} m^2/s.[113,114] Essentially, we find that despite the exceptionally high wetness of the interfaces,[60] the interstitial waters diffuse fairly sluggishly. In addition, the effect of the difference in the chemistries of the two ephrins on water diffusion is small.

Residence times. Do the waters in the interstitial regions exchange with the bulk? Figure 4 shows the residence time correlation,

$$R(\tau) = \frac{\langle s(0)s(\tau)\rangle}{\langle s\rangle} \tag{5}$$

of the water molecules present in the interstitial regions of the G−B2 and G−B3 complexes. In the equation above, the product $s(0)s(\tau)$ takes up a value of unity if a water molecule occupies the interstitial region continuously over a time interval τ. Note that in the estimation of averages, we do not include

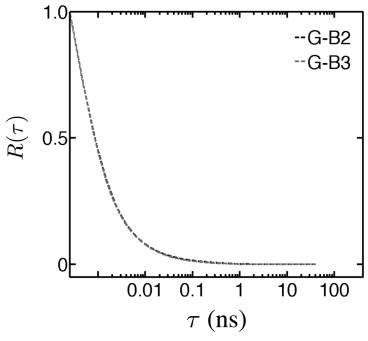


Figure 4. Residence time correlation of the waters, $R(\tau)$, occupying the interstitial regions in the G−B2 and G−B3 protein complexes.

the waters present continuously from the beginning to the end of the analysis portion of the trajectory. We find only one such "trapped" water in the G−B2 complex, and four such waters in the G−B3 complex. These trapped waters amount to less than 6% of the total waters in the interstitial regions. Modeling the residence time correlation as a sum of two exponential functions,[10]

$$R(\tau) = Ae^{-\tau/\tau_1} + (1 - A)e^{-\tau/\tau_2} \tag{6}$$

reveals two distinct subpopulations, A and $(1 - A)$, of fast and slow exchanging waters. We use unweighted least-squares fitting to determine the values of A, τ_1, and τ_2. In the interstitial region of the G−B2 complex, we find that 93% of the waters have residence times of $\tau_1 = 1.4$ ps, and the remaining fraction of waters have longer residence times of $\tau_2 = 64.7$ ps. We find a similar distribution of fast and slow exchanging waters in the interstitial region of the G−B3 complex, with $A = 91\%$, $\tau_1 = 1.3$ ps and $\tau_2 = 44.7$ ps.

We recognize that in these calculations the denotation of a continuous occupancy over time τ is subject to the choice of the observation time interval. Nevertheless, we expect that observation time intervals shorter than the ones we use, $\Delta\tau < 0.25$ ps, will reveal additional discontinuities, resulting in smaller computed residence times.[115] Therefore, the definitive conclusions we draw from this analysis are that more than 94% of the waters in the interstitial regions exchange with the bulk and that they have residence times less than hundred picoseconds. More importantly, while the interstitial waters tend to occupy crystallographic positions (Figure 1b), they exchange with bulk solvent. Note that these residence times should not be compared directly to the reported residence times of waters at protein−water interfaces because $R(\tau)$ is also a function of the shape and size of the observation volume.

Hydrogen Bond Dynamics. To evaluate quantitatively the dynamics of hydrogen bonds made by interstitial waters we use a method based on autocorrelation functions of hydrogen bond populations.[116−119] In this method, the hydrogen bond correlation function is defined as

$$c(\tau) = \frac{\langle h(0)h(\tau)\rangle}{\langle h\rangle} \tag{7}$$

where $h(\tau)$ is a hydrogen bond indicator function. $h(\tau)$ takes up a value of unity if a tagged hydrogen bond at $\tau = 0$ also exists at time τ. Otherwise, $h(\tau) = 0$. Therefore, $c(\tau)$ describes the probability that a tagged hydrogen bond at $\tau = 0$ is also bonded at a time τ, regardless of whether it breaks intermittently during the time interval τ. To account for the intermittent breaking of hydrogen bonds, a second correlation function is defined,

$$n(\tau) = \frac{\langle h(0)[1 - h(\tau)]H(\tau)\rangle}{\langle h\rangle} \tag{8}$$

In this correlation function, $H(\tau)$ takes up a value of unity if a tagged hydrogen bonded pair at $\tau = 0$ is not hydrogen-bonded at time τ, but the donor and acceptor atoms of the tagged pair are within some predefined distance. Otherwise, $H(\tau) = 0$. Essentially, $n(\tau)$ describes the probability that the donor and acceptor atoms of a tagged hydrogen bond at $\tau = 0$ are not hydrogen bonded at τ, but are within some distance that allows the pair to re-engage in hydrogen bonding. In the original Luzar−Chandler model,[118−120] this cutoff distance was chosen as 3.5 Å, which corresponds roughly to the first minimum in the oxygen−oxygen radial distribution function of bulk water.[121,122]

dx.doi.org/10.1021/jp5089096 | J. Phys. Chem. B 2014, 118, 14795−14807

Consequently, this choice of cutoff distance implies that donor−acceptor pairs that are inside each-others first coordination shell are considered to have the ability to re-engage in hydrogen bonding. Associating $c(\tau)$ and $n(\tau)$ with populations of unbroken and broken hydrogen bonds at time τ, respectively, the kinetics of hydrogen bond formation can be modeled as

$$\frac{dc(\tau)}{d\tau} = -kc(\tau) + k'n(\tau) \tag{9}$$

where k and k' are rate constants. The inverse of the rate constant k is considered to be the average hydrogen bond lifetime, that is, $\tau_{HB} = 1/k$. The time derivative of $c(\tau)$ is essentially a reactive flux correlation function,[118,119]

$$\frac{dc(\tau)}{d\tau} = \frac{\langle \dot{h}(0)[1 - h(\tau)]\rangle}{\langle h\rangle} \tag{10}$$

where $\dot{h}(0) = dh/d\tau|_{\tau = 0}$.

To determine the hydrogen bond life times of interstitial waters from eq 9, we first determine the correlation functions $c(\tau)$, $n(\tau)$ and $dc(\tau)/d\tau$, and then use least-squares fitting to solve for the rate constants k and k'.[123] We adopt a geometric definition of hydrogen bond used in the Luzar−Chandler model: a pair of donor (D) and acceptor (A) atoms, with at least one of them being an interstitial water oxygen, are considered hydrogen bonded if they are separated by less than 3.5 Å, and simultaneously the A−D−H angle is less than 30°. The angle of 30° corresponds roughly to the amplitude of librations that break hydrogen bonds, as estimated from Debye−Waller factors. We note that while other definitions for hydrogen bonds are possible, such as those based on energetics or electronic structure, they yield similar results for hydrogen bonding kinetics in bulk water.[124] While we consider all nitrogen and oxygen atoms in proteins to be potential acceptors and the subset of acceptors that are bonded to hydrogens as donors, we do not include in our analysis the weak hydrogen bonds involving protein carbon atoms.[125−128] Our rationale for not considering weak hydrogen bonds is that their thermodynamic stability emerges primarily from induced effects, which are not modeled explicitly in nonpolarizable force field simulations.[91,127,129]

Figure 5 shows the time evolution of the number of hydrogen bonds made by interstitial waters in the G−B2 complex. As noted above, there are on an average 65 waters present in the interstitial region of the G−B2 complex. We find that these interstitial waters make 153 ± 10 distinct hydrogen bonds, out of which 48% involve protein functional groups. An analysis of the protein−water hydrogen bonds shows that in the majority of cases water serves as a donor. We also find that, on an average, each interstitial water molecule is engaged in forming 3.3 hydrogen bonds, which is comparable to the number of hydrogen bonds made by a water molecule in the bulk phase. Separate simulations of bulk water indicate that a water molecule makes on an average 3.4 hydrogen bonds in the bulk phase, which is consistent with previous studies of the TIP4P water model.[92] We also find that about 10% of the interstitial waters (6 ± 2) form hydrogen bonds simultaneously with both proteins in the complex, essentially bridging the interaction between them. A visual inspection of these bridging waters indicates that they are not clustered at any specific site of the protein−protein interface. The interstitial waters in the G−B3 complex exhibit similar overall statistics. The interstitial waters in the G−B3 complex make 165 ± 10 distinct hydrogen bonds, out of which about 50% are with protein functional groups. In this case, 10 ± 3 waters serve to bridge the two proteins.

Figure 6 shows the autocorrelation functions of hydrogen bonds made by interstitial waters in the G−B2 complex. We calculate correlation functions for all hydrogen bonds made by the interstitial waters and also separately for the water−water and water−protein hydrogen bonds. These correlations are estimated using observation time intervals of 2 fs, which is necessary for computing $dc(\tau)/d\tau$.[120] We find that estimates of $dc(\tau)/d\tau$ obtained from a numerical derivative of $c(\tau)$ do not match those obtained from eq 10 for observation time intervals greater than 2 fs. The correlation profiles for interstitial waters in the G−B3 complex are similar to those in the G−B2 complex.

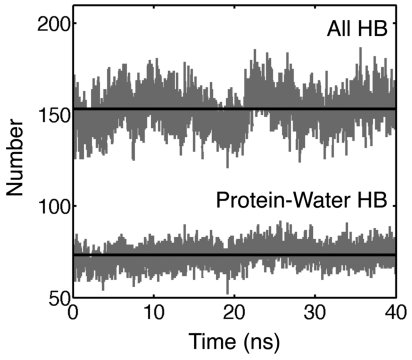


Figure 5. Time evolution of the number of hydrogen bonds (HB) made by interstitial waters in the G−B2 complex. The solid lines overlaying the number evolutions correspond to average values. The total number of hydrogen bonds is a sum of the numbers of protein−water and water−water hydrogen bonds.

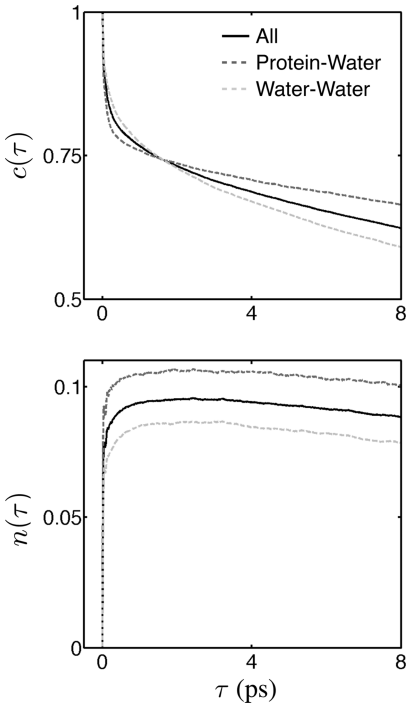


Figure 6. Autocorrelation functions of hydrogen bond populations, $c(\tau)$ and $n(\tau)$, of the interstitial waters in the G−B2 complex.

dx.doi.org/10.1021/jp5089096 | J. Phys. Chem. B 2014, 118, 14795−14807

The rate constants obtained from these correlation functions are listed in Table 1. The data in Table 1 also examine the effect of external boundary conditions on hydrogen bond dynamics. We find that, for bulk water, canonical boundary conditions produce a hydrogen bond lifetime similar to that obtained from a Gibbs ensemble. In addition, choosing a relaxation time constant for the Nose−Hoover chain that weakens temperature coupling by a factor of 5 has a negligible effect on τ_{HB}. Simulating bulk water under microcanonical boundary conditions, however, produces a τ_{HB} longer than that obtained from a canonical ensemble. Nevertheless, this difference is not due *per se* to the altered thermodynamic boundary conditions. It arises primarily due to the changes introduced into the Hamiltonian and integration algorithm for obtaining energy conservation and temperature stability in the microcanonical ensembles.[82] The following changes were introduced—the restraints on all bonds, including those of water molecules, are released and, consequently, a smaller integration time step of 0.5 fs is used; while long-range electrostatics are computed using the particle mesh Ewald scheme, short-range electrostatic interactions are truncated using a switch function in which they are reduced linearly between 9 and 10 Å; and finally, the van der Waals interactions are computed by renormalizing them such that they decay to zero smoothly over the 10 Å of direct space. Implementing these changes into the canonical ensemble produces a τ_{HB} quantitatively comparable to that obtained from the microcanonical ensemble. We, therefore, attribute the differences in interstitial water dynamics between the Gibbs and the microcanonical ensembles to the altered simulation protocol and not to the altered boundary conditions.

We find that irrespective of the external boundary conditions the hydrogen bond lifetimes (τ_{HB}) of interstitial waters are 2−3 times longer than those in the bulk. This increase in τ_{HB} is not due entirely to the longer lifetimes of hydrogen bonds involving protein side chains. The water−water hydrogen bond lifetimes in the G−B2 interface are two times longer than those in bulk water. Nevertheless, these shifts are within the range expected at protein−water interfaces,[8,10,12,14] and so these results suggest that the hydrogen bond dynamics at protein−protein interfaces are statistically similar to those at protein−water interfaces. We also note that we do not account for explicit polarization effects in these calculations, but it has been observed that inclusion of polarization results in longer hydrogen bond lifetimes,[92,94,95] perhaps due to damped librations. In that regard, the hydrogen bond lifetimes computed for interstitial waters in G−B2/B3 complexes can be considered as lower limits.

Table 1. Hydrogen Bond Life Times of Water Molecules τ_{HB} in the Bulk Phase and in the Interstitial Regions of the G−B2 and G−B3 Complexes[a]

system		ensemble	k	k'	τ_{HB} (ps)
bulk	expt				2.6
		NPT	0.56	1.17	1.79
		$NVT_{(\tau_T=1)}$	0.55	1.11	1.82
		$NVT_{(\tau_T=5)}$	0.56	1.17	1.79
		$NVT^*_{(\tau_T=5)}$	0.33	1.01	3.03
		$NVE^*_{(T=313\pm5K)}$	0.32	0.95	3.13
G−B2	all	*NPT*	0.23	1.44	4.44
	Wat−Wat		0.24	1.58	4.22
	Wat−Pro		0.18	1.10	5.53
	all	$NVE^*_{(T=310\pm1K)}$	0.12	0.87	8.13
	Wat−Wat		0.14	0.92	7.35
	Wat−Pro		0.09	0.67	10.99
G−B3	all	*NPT*	0.28	1.71	3.55
	Wat−Wat		0.34	2.01	2.91
	Wat−Pro		0.16	0.96	6.21

[a] τ_T is the relaxation time constant of the Nose−Hoover chain. The asterisk refers to the modifications made to the Hamiltonian and to the integration algorithm to obtain energy conservation and temperature stability in the microcanonical ensembles (see text for details). In the cases of the two protein−protein complexes, the life times are estimated for all hydrogen bonds made by interstitial waters and also separately for the hydrogen bonds made between interstitial waters (Wat−Wat) and between interstitial waters and the protein (Wat−Pro). The experimental value is taken from reference.[130].

Dipole correlations. While we find that the hydrogen bond lifetimes of interstitial waters are longer compared to those in bulk water, how do the overall rotational dynamics of waters in interstitial regions differ from those in bulk water? To examine this we compute the dipole correlation function,

$$M_{\parallel}(\tau) = \frac{\langle \sum \mu_{\parallel}(0) \sum \mu_{\parallel}(\tau)/n_0 n_\tau \rangle}{\langle (\sum \mu_{\parallel}(0)/n_0)^2 \rangle} \tag{11}$$

where $\mu_{\parallel}$ is the component of the water dipole moment parallel to the axis joining the geometric centers of G and ephrins. n_0 and n_τ refer to the numbers of interstitial waters in frames separated by a time interval τ. Under isotropic conditions, such as in bulk water, and for all $n_\tau = n_0$, $M_{\parallel}(\tau) = M(\tau)$. Figure 7*a* illustrates $M_{\parallel}(\tau)$ estimated for interstitial waters in the G−B2 and G−B3 complexes. Note that since our observation time interval is 0.25 ps, this analysis does not capture the subpicosecond (ultrafast) rotational relaxation of waters.[131] We note that for both complexes, the correlation function decays asymptotically to a nonzero value, which indicates an orientational preference for the interstitial waters. We confirm this by evaluating the radial distribution of water dipoles about the axes joining the geometric centers of G and ephrins (Figure 7*b*). We attribute this orientational preference to the asymmetric distribution of charged residues at the G−ephrin interface. We model these dipole correlations as sums of two exponential functions,

$$M_{\parallel}(\tau) = (1 - B)[A e^{-\tau/\tau_1} + (1 - A) e^{-\tau/\tau_2}] + B \tag{12}$$

where A and $(1 - A)$ refer, respectively, to the fraction of waters exhibiting fast and slow relaxation rotational time constants of τ_1 and τ_2. The constant B accounts for the limiting nonzero value of $M(\tau \to \infty)$ and refers to the fraction of waters

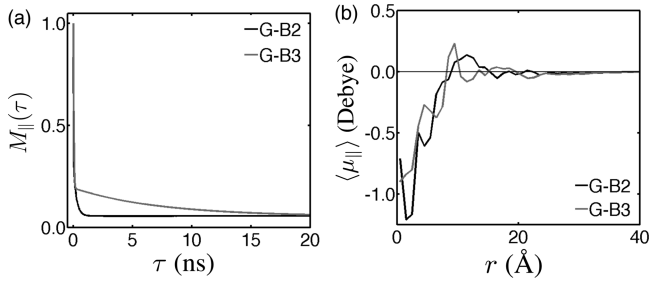


Figure 7. (a) Dipole correlations, $M_{\parallel}(\tau)$, of the interstitial waters in the G−B2 and G−B3 complexes. (b) Distribution of water dipoles, $\mu_{\parallel}$, estimated as a function of the radial distance, r, from the axis joining the geometric centers of G and B2/B3.

dx.doi.org/10.1021/jp5089096 | J. Phys. Chem. B 2014, 118, 14795−14807

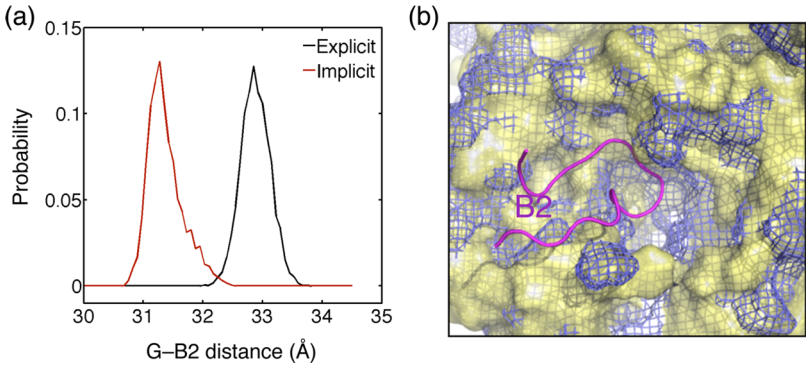


Figure 8. Effect of the treatment of the solvent on the structural properties of the G−B2 interface. (a) Probability distribution of distances between the geometric centers of G and B2 proteins obtained from implicit and explicit solvent simulations of the G−B2 complex. (b) Partial view of the G protein showing its B2 binding site. The structure of the G protein taken from the explicit solvent simulation of the G−B2 complex is rendered as a yellow vdW surface, and the structure of the G protein taken from the implicit solvent simulation of the G−B2 complex is rendered as a blue vdW mesh. The GH loop of B2 that inserts into the G cavity is shown as a magenta ribbon.

that exhibit a permanent orientational preference along the axes joining the geometric centers of G and ephrins. We use weighted least-squares fitting to determine the values of A, B, τ_1, and τ_2. In the interstitial region of the G−B2 complex, we find that 5.6% of the water molecules exhibit a permanent orientational preference along the axis joining the geometric centers of G and B2. We also find that 69% of the waters have reorientational times of τ_1 = 6 ps and 25% have longer reorientational times of τ_2 = 249 ps. While the fast relaxation time is of the order to the relaxation time in the bulk phase,[131,132] more than a quarter of interstitial waters relax at rates 2 orders in magnitude slower than bulk water, which is reminiscent of an anisotropic polar environment. An alternative interpretation of this result is that the set of interstitial waters have two different relaxations time scales, one corresponding to that of bulk water, and another that is a hundred-fold slower. In the interstitial region of the G−B3 complex, the rotational relaxation of water is even slower, with 81% of the waters relaxing with a τ_1 = 38 ps and 14% relaxing with a τ_2 = 7317 ps. We find a similar residual fraction of waters permanently oriented along the axis joining the geometric centers of G and B3. The difference in relaxation rates of interstitial waters in the G−B2 and G−B3 complexes emerges primarily from differences between the sequences of ephrins B2 and B3. This is because the G−B2 and G−B3 complexes are structurally similar (RMSD < 2 Å). In general, these results show that chemistry can affect dipole relaxation rates significantly, and highlight that the dielectric response of interstitial solvent is unlike that of bulk water.

Role of Interstitial Water in Signal Transduction. Consistent with the X-ray structure of the G−B2 complex,[58] our explicit solvent MD simulations indicate that the G−ephrin complexes sandwich exceptionally high numbers of waters in their interstitial regions. Additionally, while the interstitial waters tend to occupy crystallographic sites, most interstitial waters exchange with the bulk solvent every hundred picoseconds. But what specific physiological role do these interstitial waters serve?

The binding of ephrin B2/B3 to G causes G to activate another viral protein, F.[55−59] Upon activation, F mediates virus-host membrane fusion. Since ephrins and the F protein bind to G at mutually exclusive sites, the effect of ephrin binding must transduce to the F-binding site of G to activate F.[55−59,66] This F-activating allosteric signal, is contained within the changes in the conformational density of G brought about by ephrin binding.[66] If $\mathcal{R}_{apo}$ and $\mathcal{R}_{bound}$ represent, respectively, the conformational densities sampled by the G protein in its apo and ephrin-bound states, then the F-activating signal is contained within the ephrin-induced conformational density shift $\Delta\mathcal{R} := \mathcal{R}_{apo} \rightarrow \mathcal{R}_{bound}$. Now if the waters at the G−ephrin interface were to contribute to the allosteric activation of F, then they must contribute to $\Delta\mathcal{R}$.

To examine whether the interstitial waters contribute to $\Delta\mathcal{R}$, we determine $\Delta\mathcal{R}$ from explicit solvent simulations and compare them quantitatively to the $\Delta\mathcal{R}$ obtained from a separate set of implicit solvent simulations. We expect explicit and implicit solvent models to yield different $\Delta\mathcal{R}$ because (a) the two models describe bulk solvent differently, and (b) implicit solvent models do not account for the discrete nature of water molecules at the G−ephrin interface. Given the high numbers of waters at the G−ephrin interface, the lack of their specific volumes in the implicit solvent model can alter the G−ephrin interface. In addition, since about 10% of the interstitial waters hydrogen bond simultaneously with both proteins, bridging the interaction between the two proteins, their absence could directly alter G−B2 binding modes. The absence of discrete waters at the G−B2 interface could, therefore, lead to altered $\Delta\mathcal{R}$.

Figure 8 shows two primary differences in the G−B2 interface predicted from explicit and implicit solvent simulations. Compared to the explicit solvent model, (a) the implicit solvent model predicts a smaller distance between G and B2, which presumably results from the absence of discrete waters at the interface, and (b) the implicit solvent model predicts a smaller width of the G cavity into which the GH loop of B2 inserts, which can also be attributed to the absence of discrete waters at the interface. These two structural differences signify that in the absence of explicit waters in the implicit solvent model, the G−B2 interface is more compact.

Figure 9a compares the $\Delta\mathcal{R}$ obtained from explicit and implicit solvent simulations. The $\Delta\mathcal{R}$ are estimated separately for each amino acid of G in terms of a quantity that we refer to as discriminability (η).[66,99] This quantity is normalized and bounded, that is, $\eta \in [0,1)$, and it takes up a value closer to unity as the difference between the conformational densities increases. We determine η separately for each residue in the G head domain between its representative ensembles in the apo

dx.doi.org/10.1021/jp5089096 | J. Phys. Chem. B 2014, 118, 14795−14807

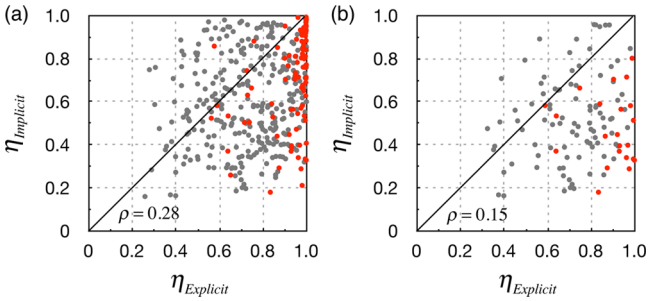


Figure 9. Comparison of B2-induced conformational density shifts in G, as estimated from implicit and explicit solvent simulations. (a) The 416 dots represent the conformational density shifts for the 416 residues in G. The dots colored red correspond to residues that we found to be part of the allosteric signaling pathway.[66] We find that η estimated from explicit and implicit solvent simulations are statistically different with a Pearson correlation of 0.28. (b) The 114 dots represent the conformational density shifts of a subset of the residues of G that satisfy the condition given by eq 13. Even for these residues, the η estimated from explicit and implicit solvent simulations are statistically different with a Pearson correlation of 0.15.

and B2 bound states. Since we are simulating a 416 residue segment of the G head domain, a comparison between the two G ensembles yields 416 η values, one η for each residue. Each of the two ensembles is represented by 2001 conformations extracted at regular intervals from their respective simulations. Doubling or reducing the ensemble size by a factor of 2 affects the quantification minimally.[99] Note that prior to extracting the ensemble of a residue from the G head domain, all the simulated configurations of G head domain are least-squares fitted on to the X-ray coordinates of the G head domain. Structure fitting is necessary to remove the bias of η against whole molecule rotation and translation, as that is not the goal of this comparison. We expect the least-squares algorithm to be adequate for structure fitting because the structural differences between the apo and bound states are small.[133,134] Note also that during ensemble comparison we consider only heavy atoms.

We find that the $\Delta\mathcal{R}$ estimated from explicit and implicit solvent simulations are statistically different with a Pearson correlation of 0.28 (Figure 9a). This difference is, in fact, even more pronounced for residues that are known to participate in allosteric signaling.[66] This difference, however, reflects the overall effect of treating the solvent using a mean field approximation. While the absence of discrete interstitial waters in the implicit solvent simulation also contribute to this difference, this analysis does not delineate their specific role.

To gain further insight into the specific contribution of interstitial waters to $\Delta\mathcal{R}$, we carry out an additional analysis. We identify the subset of residues in the G protein whose conformational densities in the apo state are unaffected by the treatment of the bulk solvent. To accomplish this, we estimate for each residue in the apo state of the G protein the difference between its conformational density obtained from explicit and implicit solvent simulations. We denote this difference as $\eta_{imp\leftrightarrow exp}$. We then filter out the residues whose $\eta_{imp\leftrightarrow exp}$ are smaller than a certain tolerance. We choose $d^2 = BT/8\pi^2 T_{xray}$ as the tolerance, which is the mean square deviation of a residue obtained from crystallographic B factors.[59] The ratio T/T_{xray} rescales the B factors from X-ray diffraction temperature T_{xray} = 100 K to physiological temperature T = 310 K.[135,136] Consequently, if for a given residue, we find that

$$\eta_{imp\leftrightarrow \exp} < \mathrm{erf}(d/\sqrt{2}) \tag{13}$$

then the difference between its conformational density obtained from explicit and implicit solvent simulations is smaller than the spread in the residue's electron density observed in X-ray diffraction. Note that the error function in eq 13 represents the transformation of the tolerance to the appropriate Hilbert space where η is estimated.[66] We find that 114 out of the 416 residues of G meet this criterion, and even for these residues we find that explicit and implicit solvent simulations produce statistically different B2-induced shifts in conformational density (Figure 9b). Since the conformational densities of these residues in the apo state are not affected by the treatment of solvent, this difference reflects the specific effect of treating the G−B2 interaction using a mean field approximation.

Taken together with the results from Figure 8, this analysis suggests that the differences in $\Delta\mathcal{R}$ estimated from implicit and explicit solvent models are at least in part due to the absence of explicit interstitial waters in the implicit solvent model, supporting the hypothesis that the interstitial waters also contribute to $\Delta\mathcal{R}$. This finding, however, does not generalize the idea that implicit solvent models should not be used for investigating protein−protein complexes. In fact, there are numerous examples in literature where protein−protein interactions have been modeled successfully using implicit solvent models.[76−79] The G−ephrin complex is unique in the sense that it sandwiches an exceptionally large amount of water at its interface, and so perhaps, in such special cases, the treatment of explicit solvent becomes critical.

■ CONCLUSIONS

The dynamical properties of water at protein surfaces have been studied extensively using both experimental and computational techniques. There is also now growing consensus that water molecules at the protein−water interface translate and rotate slower compared to the bulk. Here we extend these studies to systems where water molecules occupy interstitial regions between two proteins. Specifically, we consider two natural protein−protein complexes, both of which are formed independently during the fusion of Nipah viruses with host cells. In one complex, the Nipah virus G protein binds to cellular ephrin B2, and in the other the same G protein binds to ephrin B3. While the two complexes are structurally similar, the two ephrins share only a modest sequence identity of ∼50%, even in the portions that form the interface. X-ray crystallography suggests that these interfaces are fairly extensive, with interface areas greater than 2500 Å^2, and contain three times more interstitial waters compared to the average numbers of waters in other protein−protein interfaces of comparable sizes.

Analysis of our atomistic molecular dynamics simulations shows that while the interstitial waters tend to occupy crystallographic sites, most waters exhibit residence times of less than hundred picoseconds in the interstitial region. We, therefore, argue that the crystallographic sites for water should not be viewed as sites for "bound" waters, but rather as preferred sites for water occupancy. We also find that while the dynamical properties of the interstitial waters in the two complexes are quantitatively different, the trends in the shifts with respect to bulk values are similar. Since the two ephrins are topologically similar, the quantitative differences in water dynamics emerge primarily from the differences in the sequences of the two ephrins. The effect of chemical difference

dx.doi.org/10.1021/jp5089096 | J. Phys. Chem. B 2014, 118, 14795−14807

is seen predominantly in dipole relaxation rates and not as much in diffusion rates or residence times or hydrogen-bond lifetimes.

We also find that despite the exceptional wetness of the protein−protein interfaces, the dynamics of interstitial waters are considerably slower compared to the bulk. Since the degree of wetness of this interface is among the limiting cases, we expect that water dynamics in all other protein−protein interfaces will also be sluggish, however, that remains to be determined. We find that the interstitial waters diffuse at rates 10 times slower compared to bulk water. Additionally, the interstitial waters exhibit hydrogen bond lifetimes 2−3 times longer than bulk water. This increase is not entirely due to the presence of protein−water hydrogen bonds. The water−water hydrogen bond lifetimes also increase in the interstitial regions, indicating that the rattling events that break hydrogen bonds are, in general, slower. Nevertheless, these shifts in hydrogen bond lifetimes are within the range expected at protein−water interfaces, and so these results suggest that the hydrogen bond dynamics at protein−protein interfaces are statistically similar to those in protein−water interfaces. We also find that while the majority of interstitial waters exhibit dipole relaxation times similar to those in the bulk, there is an appreciable fraction whose relaxation times are 100−1000 fold longer than bulk water.

To understand the role of interstitial waters, we examine quantitatively how implicit solvent models compare against explicit solvent models in producing ephrin-induced shifts in the G conformational density. Ephrin-induced shifts in the G conformational density are critical to the allosteric activation of the viral fusion protein, F. We find that the implicit solvent model predicts a more compact G−B2 interface compared to the explicit solvent model, with G and B2 being physically closer to each other presumably because of the absence of the discrete interstitial waters at the G−B2 interface. Simultaneously, we find that the two models yield strikingly different induced changes in the G conformational density, even for those amino acids whose conformational densities in the apo state are unaffected by the treatment of the bulk solvent. Together, these results show that the explicit treatment of interstitial water molecules is necessary for a proper description of allosteric transitions.

AUTHOR INFORMATION

Corresponding Author

*E-mail: svarma@usf.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge the use of services provided by Research Computing at USF. This research was supported by a startup grant to S.V.

REFERENCES

(1) Saenger, W. Structure and Dynamics of Water Surrounding Biomolecules. *Annu. Rev. Biophys. Biol.* **1987**, *16*, 93−114.
(2) Meyer, E. Internal Water-Molecules and H-bonding in Biological Macromolecules-A Review of Structural Features with Functional Implications. *Protein Sci.* **1992**, *1*, 1543−1562.
(3) Otting, G. NMR Studies of Water Bound to Biological Molecules. *Prog. Nucl. Magn. Reson. Spectrosc.* **1997**, *31*, 259−285.
(4) Bellissent-Funel, M. C. Structure and Dynamics of Water near Hydrophilic Surfaces. *J. Mol. Liq.* **1998**, *78*, 19−28.
(5) Bhattacharyya, K.; Bagchi, B. Slow Dynamics of Constrained Water in Complex Geometries. *J. Phys. Chem. A* **2000**, *104*, 10603−10613.
(6) Mattos, C. Protein-Water Interactions in a Dynamic World. *Trends. Biochem. Sci.* **2002**, *27*, 203−208.
(7) Guillot, B. A Reappraisal of What We Have Learnt During Three Decades of Computer Simulations on Water. *J. Mol. Liq.* **2002**, *101*, 219−260.
(8) Bizzarri, A. R.; Cannistraro, S. Molecular Dynamics of Water at the Protein-Solvent Interface. *J. Phys. Chem. B* **2002**, *106*, 6617−6633.
(9) Buch, V.; Paul, D. J. *Water in Confining Geometries*; Springer Series in Cluster Physics; Springer: Berlin, 2003, *XII*, pp: 472.
(10) Bagchi, B. Water Dynamics in the Hydration Layer Around Proteins and Micelles. *Chem. Rev.* **2005**, *105*, 3197−3219.
(11) Chandler, D. Interfaces and the Driving Force of Hydrophobic Assembly. *Nature* **2005**, *437*, 640−647.
(12) Raschke, T. M. Water Structure and Interactions With Protein Surfaces. *Curr. Opin. Struct. Biol.* **2006**, *16*, 152−159.
(13) Chaplin, M. Do We Underestimate the Importance of Water in Cell Biology? *Nat. Rev. Mol. Cell Bio.* **2006**, *7*, 861−866.
(14) Geissler, P. L. Water Interfaces, Solvation, and Spectroscopy. *Annu. Rev. Phys. Chem.* **2013**, *64*, 317−337.
(15) Brunne, R. M.; Otting, E. L. G.; Wuthrich, K.; van Gunsteren, W. F. Hydration of Proteins- A Comparison of Experimental Residence Time of Water-Molecules Solvating the Bovine Pancreatic Trypsin-Inhibitor with Theoretical-Model Calculations. *J. Mol. Biol.* **1993**, *231*, 1040−1048.
(16) Denisov, V. P.; Halle, B. Dynamics of the Internal and External Hydration of Globular-Proteins. *J. Am. Chem. Soc.* **1994**, *116*, 10324−10325.
(17) Lee, S. H.; Rossky, P. J. A Comparison of the Structure and Dynamics of Liquid Water at Hydrophobic and Hydrophilic Surfaces-A Molecular Dynamics Simulation Study. *J. Chem. Phys.* **1994**, *100*, 3334−3345.
(18) Muegge, I.; Knapp, E. W. Residence Times and Lateral Diffusion of Water at Protein Surfaces- Application of BPTI. *J. Phys. Chem.* **1995**, *99*, 1371−1374.
(19) Denisov, V. P.; Halle, B. Protein Hydration Dynamics in Aqueous Solution. *Faraday Discuss.* **1996**, *103*, 227−244.
(20) Bellissent-Funel, M. C.; Zanotti, J. M.; Chen, S. H. Slow Dynamics of Water Molecules on the Surface of a Globular Protein. *Faraday Discuss.* **1996**, *103*, 281−294.
(21) Settles, M.; Doster, W. Anomalous Diffusion of Adsorbed Water: A Neutron Scattering Study of Hydrated Myoglobin. *Faraday Discuss.* **1996**, *103*, 269−279.
(22) Abseher, R.; Schreiber, H.; Steinhauser, O. The Influence of a Protein on Water Dynamics in its Vicinity Investigated by Molecular Dynamics Simulation. *Proteins* **1996**, *25*, 366−378.
(23) Rocchi, C.; Bizzarri, A. R.; Cannistraro, S. Water Residence Times around Copper Plastocyanin: A Molecular Dynamics Simulation Approach. *Chem. Phys.* **1997**, *214*, 261−276.
(24) Cheng, Y. K.; Rossky, P. J. Surface Topography Dependence of Biomolecular Hydrophobic Hydration. *Nature* **1998**, *392*, 696−699.
(25) Denisov, V. P.; Jonsson, B. H.; Halle, B. Hydration of Denatured and Molten Globule Proteins. *Nat. Struct. Biol.* **1999**, *6*, 253−260.
(26) Jordanides, X. J.; Lang, M. J.; Song, X. Y.; Fleming, G. R. Solvation Dynamics in Protein Environments Studied by Photon Echo Spectroscopy. *J. Phys. Chem. B* **1999**, *103*, 7995−8005.
(27) Dellerue, S.; Bellissent-Funel, M. C. Relaxational Dynamics of Water Molecules at Protein Surface. *Chem. Phys.* **2000**, *258*, 315−325.
(28) Makarov, V. A.; Andrews, B. K.; Smith, P. E.; Montgomery Pettitt, B. Residence Times of Water Molecules in the Hydration Sites of Myoglobin. *Biophys. J.* **2000**, *79*, 2966−2974.
(29) Sterpone, F.; Ceccarelli, M.; Marchi, M. Dynamics of Hydration in Hen Egg White Lysozyme. *J. Mol. Biol.* **2001**, *311*, 409−419.
(30) Zhong, D. P.; Pal, S. K.; Zhang, D. Q.; Chan, S. I.; Zewail, A. H. Femtosecond Dynamics of Rubredoxin: Tryptophan Solvation and

Resonance Energy Transfer in the Protein. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 13−18.
(31) Bizzarri, A. R.; Cannistraro, S. Molecular Dynamics of Water at the Protein-Solvent Interface. *J. Phys. Chem. B* **2002**, *106*, 6617−6633.
(32) Merzel, F.; Smith, J. C. Is the First Hydration Shell of Lysozyme of Higher Density than Bulk Water? *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5378−5383.
(33) Marchi, M.; Sterpone, F.; Ceccarelli, M. Water Rotational, Relaxation and Diffusion in Hydrated Lysozyme. *J. Am. Chem. Soc.* **2002**, *124*, 6787−6791.
(34) Russo, D.; Baglioni, P.; Peroni, E.; Teixeira, J. Hydration Water Dynamics of a Completely Hydrophobic Oligopeptide. *Chem. Phys.* **2003**, *292*, 235−245.
(35) Modig, K.; Liepinsh, E.; Otting, G.; Halle, B. Dynamics of Protein and Peptide Hydration. *J. Am. Chem. Soc.* **2004**, *126*, 102−114.
(36) Russo, D.; Hura, G.; Head-Gordon, T. Hydration Dynamics near a Model Protein Surface. *Biophys. J.* **2004**, *86*, 1852−1862.
(37) Russo, D.; Murarka, R. K.; Copley, J. R. D.; Head-Gordon, T. Molecular View of Water Dynamics near Model Peptides. *J. Phys. Chem. B* **2005**, *109*, 12966−12975.
(38) Yang, C.; Sharp, K. A. Hydrophobic Tendency of Polar Group Hydration as a Major Force in Type I Antifreeze Protein Recognition. *Proteins* **2005**, *59*, 266−274.
(39) Dokter, A. M.; Woutersen, S.; Bakker, H. J. Inhomogeneous Dynamics in Confined Water Nanodroplets. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15355−15358.
(40) Ebbinghaus, S.; Kim, S. J.; Heyden, M.; Yu, X.; Heugen, U.; Gruebele, M.; Leitner, D. M.; Havenith, M. 2007. An Extended Dynamical Hydration Shell around Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 20749−20752.
(41) Pizzitutti, F.; Marchi, M.; Sterpone, F.; Rossky, P. J. How Protein Surfaces Induce Anomalous Dynamics of Hydration Water. *J. Phys. Chem. B* **2007**, *111*, 7584−7590.
(42) Sinha, S. K.; Chakraborty, S.; Bandyopadhyay, S. Thickness of the Hydration Layer of a Protein from Molecular Dynamics Simulation. *J. Phys. Chem. B* **2008**, *112*, 8203−8209.
(43) Patel, A. J.; Varilly, P.; Chandler, D. Fluctuations of Water near Extended Hydrophobic and Hydrophilic Surfaces. *J. Phys. Chem. B* **2010**, *114*, 1632−1637.
(44) Sterpone, F.; Stirnemann, G.; Hynes, J. T.; Laage, D. Water Hydrogen-Bond Dynamics around Amino Acids: The Key Role of Hydrophilic Hydrogen-Bond Acceptor Groups. *J. Phys. Chem. B* **2010**, *114*, 2083−2089.
(45) Khodadadi, S.; Curtis, J. E.; Sokolov, A. P. Nanosecond Relaxation Dynamics of Hydrated Proteins: Water Versus Protein Contributions. *J. Phys. Chem. B* **2011**, *115*, 6222−6226.
(46) Gupta, S.; D'Mello, R.; Mark, R. C. Structure and Dynamics of Protein Waters Revealed by Radiolysis and Mass Spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 14882−14887.
(47) King, J. T.; Kubarych, K. J. Site-Specific Coupling of Hydration Water and Protein Flexibility Studied in Solution with Ultrafast 2D-IR Spectroscopy. *J. Am. Chem. Soc.* **2012**, *134*, 18705−18712.
(48) Jana, B.; Pal, S.; Bagchi, B. Hydration Dynamics of Protein Molecules in Aqueous Solution: Anity Among Diversity. *J. Chem. Sci.* **2012**, *124*, 317−325.
(49) Smith, E. C.; Popa, A.; Chang, A.; Masante, C.; Dutch, R. E. Viral Entry Mechanisms: The Increasing d|Diversity of Paramyxovirus Entry. *FEBS J.* **2009**, *276*, 7217−7227.
(50) Ksiazek, T. G.; Rota, P. A.; Rollin, P. E. A Review of Nipah and Hendra viruses with an Historical Aside. *Virus Res.* **2011**, *162*, 173−183.
(51) Bowden, T. A.; Jones, E. Y.; Stuart, D. I. Cells under Siege: Viral Glycoprotein Interactions at the Cell Surface. *J. Struct. Biol.* **2011**, *175*, 120−126.
(52) Lee, B.; Ataman, Z. A. Modes of Paramyxovirus Fusion: A Henipavirus Perspective. *Trends Microbiol.* **2011**, *19*, 389−399.
(53) Steffen, D. L.; Xu, K.; Nikolov, D. B.; Broder, C. C. Henipavirus Mediated Membrane Fusion Virus Entry and Targeted Therapeutics. *Viruses* **2012**, *4*, 280−309.
(54) Talekar, A.; Pessi, A.; Glickman, F.; Sengupta, U.; Briese, T.; Whitt, M. A.; Mathieu, C.; Horvat, B.; Moscona, A.; Porotto, M. Rapid Screening for Entry Inhibitors of Highly Pathogenic Viruses under Low-Level Biocontainment. *PLoS One* **2012**, *7*, e30538.
(55) Negrete, O. A.; Levroney, E. L.; Aguilar, H. C.; Bertolotti-Ciarlet, A.; Nazarian, R.; Tajyar, S.; Lee, B. Ephrin-B2 is the Entry Receptor for Nipah Virus an Emergent Deadly Paramyxovirus. *Nature* **2005**, *436*, 401−405.
(56) Bonaparte, M. I.; Dimitrov, A. S.; Bossart, K. N.; Crameri, G.; Mungall, B. A.; Bishop, K. A.; Choudhury, A.; Dimitrov, D. S.; Wang, L. F.; Eaton, B. T.; Broder, C. C. Ephrin-B2 Ligand is a Functional Receptor for Hendra Virus and Nipah Virus. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 10652−10657.
(57) Negrete, O. A.; Wolf, M. C.; Aguilar, H. C.; Enterlein, S.; Wang, W.; Muhlberger, E.; Su, S. V.; Bertolotti-Ciarlet, A.; Flick, R.; Lee, B. Two Key Residues in Ephrin-B3 are Critical for its Use as an Alternative Receptor for Nipah Virus. *PLoS Pathog.* **2006**, *2*, e7.
(58) Bowden, T. A.; Ariscescu, A. R.; Gilbert, R. J. C.; Grimes, J. M.; Jones, E. Y.; Stuart, D. I. Structural Basis of Nipah and Hendra Virus Attachment to their Cell-Surface Receptor Ephrin-B2. *Nat. Struct. Mol. Biol.* **2008**, *15*, 567−572.
(59) Xu, K.; Rajashankar, K. R.; Chan, Y. P.; Himanen, J. P.; Broder, C. C.; Nikolov, D. B. Host Cell Recognition by the Henipaviruses: Crystal Structures of the Nipah G Attachment Glycoprotein and its Complex with Ephrin-B3. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *22*, 9953−9958.
(60) Rodier, F.; Bahadur, R. P.; Chakrabarti, P.; Janin, J. Hydration of Protein-Protein Interfaces. *Proteins* **2005**, *60*, 36−45.
(61) Ahmed, M. H.; Spyrakis, F.; Cozzini, P.; Tripathi, P. K.; Mozzarelli, A.; Scarsdale, J. N.; Safo, M. A.; Kellogg, G. E. Bound Water at Protein-Protein Interfaces: Partners Roles and Hydrophobic Bubbles as a Conserved Motif. *PLoS One* **2011**, *6*, e24712.
(62) Otwinowski, Z.; Schevitz, R. W.; Zhang, R. G.; Lawson, C. L.; Joachimiak, A.; Marmorstein, R. Q.; Luisi, B. F.; Sigler, P. B. Crystal Structure of Trp Repressor/Operator Complex at Atomic Resolution. *Nature* **1988**, *335*, 321−329.
(63) Sundaralingam, M.; Sekharudu, Y. C. Water-Inserted Alpha-Helical Segments Implicate Reverse Turns as Folding Intermediates. *Science* **1989**, *244*, 1333−1337.
(64) Petrone, P. M.; Garcia, A. E. MHC Peptide Binding is Assisted by Bound Water Molecules. *J. Mol. Biol.* **2004**, *338*, 419−435.
(65) Jiang, L.; Kuhlman, B.; Kortemme, T.; Baker, D. A Solvated Rotamer Approach to Modeling Water-Mediated Hydrogen Bonds at Protein-Protein Interfaces. *Proteins* **2005**, *58*, 893−904.
(66) Varma, S.; Botlani, M.; Leighty, R. E. Discerning Intersecting Fusion-Activation Pathways in the Nipah Virus Using Machine Learning. *Proteins* **2014**, *82*, 3241−3254.
(67) Kern, D.; Zuiderweg, E. R. P. The Role of Dynamics in Allosteric Regulation. *Curr. Opin. Struct. Biol.* **2003**, *13*, 748−757.
(68) Gunasekaran, K.; Ma, B.; Nussinov, R. Is Allostery an Intrinsic Property of all Dynamic Proteins. *Proteins* **2004**, *57*, 433−443.
(69) Swain, J. F.; Gierasch, L. M. The Changing Landscape of Protein Allostery. *Curr. Opin. Struct. Biol.* **2006**, *16*, 102−108.
(70) Bahar, I.; Chenncebhotia, C.; Tobi, D. Intrinsic Dynamics of Enzymes in the Unbound State and Relation to Allosteric Regulation. *Curr. Opin. Struct. Biol.* **2007**, *17*, 633−640.
(71) Cui, Q.; Karplus, M. Allostery and Cooperativity Revisited. *Protein Sci.* **2008**, *17*, 1295−1307.
(72) Zhuravlev, P. I.; Papoian, G. A. Protein Functional Landscapes Dynamics Allostery: A Tortuous Path towards a Universal Theoretical Framework. *Q. Rev. Biophys.* **2010**, *43*, 295−332.
(73) Bu, Z.; Callaway, D. J. Proteins Move! Protein Dynamics and Long-Range Allostery in Cell Signaling. *Adv. Protein Chem. Struct. Biol.* **2011**, *83*, 163−221.
(74) Hilser, V. J.; Wrabl, J. O.; Motlagh, H. N. Structural and Energetic Basis of Allostery. *Annu. Rev. Biophys.* **2012**, *41*, 585−609.
(75) Rhee, Y. M.; Sorin, E. J.; Jayachandran, G.; Lindahl, E.; Pande, V. S. Simulations of the role of water in the protein- folding mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6456−6461.

(76) Wang, T.; Wade, R. C. Implicit Solvent Models for Flexible Protein-Protein Docking by Molecular Dynamics Simulation. *Proteins* **2003**, *50*, 158−169.
(77) Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins* **2004**, *55*, 383−394.
(78) Bonvin, A. M. J. J. Flexible Protein-Protein Docking. *Curr. Opin Struct Biol.* **2006**, *16*, 194−200.
(79) Chen, J.; Brooks, C. L., III; Khandogin, J. Recent Advances in Implicit Solvent Based Methods for Biomolecular Simulations. *Curr. Opin Struct Biol.* **2008**, *18*, 140−148.
(80) Fiser, A.; Do, R. K.; Sali, A. Modeling of Loops in Protein Structures. *Protein Sci.* **2000**, *9*, 1753−1773.
(81) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. PDB2PQR: Expanding and Upgrading Automated Preparation of Biomolecular Structures for Molecular Simulations. *Nucleic Acids Res.* **2007**, *35*, W522−W525.
(82) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient Load-Balanced and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(83) Nose, S. A Molecular-Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52*, 255−268.
(84) Hoover, W. G. Canonical Dynamics - Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31*, 1695−169.
(85) Parrinello, M.; Rahman, A. Polymorphic Transition in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7128−7190.
(86) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparameterization of the OPLS-AA Force Field for Proteins Via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.
(87) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(88) Guevara-Carrion, G.; Vrabec, J.; Hasse, H. Prediction of Self-Diffusion Coefficient and Shear Viscosity of Water and Its Binary Mixtures with Methanol and Ethanol by Molecular Simulation. *J. Chem. Phys.* **2011**, *134*, 074508−074514.
(89) Neumann, M. Dielectric Relaxation in Water: Computer Simulations with the TIP4P Potential. *J. Chem. Phys.* **1986**, *85*, 1567−1580.
(90) Varma, S.; Rempe, S. B. Multi-Body Effects in Ion Binding and Selectivity. *Biophys. J.* **2010**, *99*, 3394−3401.
(91) Rossi, M.; Tkatchenko, A.; Rempe, S. B.; Varma, S. Role of Methyl-Induced Polarization in Ion Binding. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 12978−12983.
(92) Xu, H.; Stern, H. A.; Berne, B. J. Can Water Polarizability be Ignored in Hydrogen Bond Kinetics? *J. Phys. Chem. B* **2002**, *106*, 2054−2060.
(93) Kim, B.; Young, T.; Harder, E.; Friesner, R. A.; Berne, B. J. Structure and Dynamics of the Solvation of Bovine Pancreatic Trypsin Inhibitor in Explicit Water: A Comparative Study of the Effects of Solvent and Protein Polarizability. *J. Phys. Chem. B* **2005**, *109*, 16529−16538.
(94) Lee, H. S.; Tuckerman, M. E. Dynamical Properties of Liquid Water from *ab initio* Molecular dynamics performed in the complete basis set limit. *J. Chem. Phys.* **2007**, *126*, 164501−164516.
(95) Johnson, M. E.; Malardier-Jugroot, C.; Murarka, R. K.; Head-Gordon, T. Hydration Water Dynamics near Biological Interfaces. *J. Phys. Chem. B* **2009**, *113*, 4082−4092.
(96) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An NLog(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(97) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory. Comput.* **2008**, *4*, 116−122.
(98) Miyamoto, S.; Kollman, P. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithms for Rigid Water Molecules. *J. Comput. Chem.* **1992**, *13*, 952−962.
(99) Leighty, R. E.; Varma, S. Quantifying Changes in Intrinsic Molecular Motion Using Support Vector Machines. *J. Chem. Theory. Comput.* **2013**, *9*, 868−875.
(100) Lindorff-Larsen, K.; Ferkinghoff-Borg, J. Similarity Measures for Protein Ensembles. *PLoS One* **2009**, *4*, e4203.
(101) McClendon, C. L.; Hua, L.; Barreiro, G.; Jacobson, M. P. Comparing Conformational Ensembles Using the Kullback-Leibler Divergence Expansion. *J. Chem. Theory. Comput.* **2012**, *8*, 2115−2126.
(102) Cortes, C.; Vapnik, V. Support Vector Networks. *Mach. Learn* **1995**, *20*, 273−297.
(103) Aizerman, M.; Braverman, E.; Rozonoer, L. Theoretical Foundations of the Potential Function Method in Pattern Recognition Learning. *Autom. Rem. Contr.* **1964**, *25*, 821−837.
(104) Scholkopf, B.; Burges, C.; Smola, A. *Advances in Kernel Methods: Support Vector Learning*. MIT Press: Cambridge, MA, 1999; pp 69−88.
(105) Cristianini, N.; Shawe-Taylor, J. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*; Cambridge University Press: Cambridge, U.K., 2000; pp 93−124.
(106) Smola, A. J.; Scholkopf, B.; Muller, K. R. The Connection between Regularization Operators and Support Vector Kernels. *Neural Netw.* **1998**, *11*, 637−649.
(107) Platt, J. C. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. *Technical Report for Microsoft Research*: Microsoft: Redmond, WA, **1998**; *MSR-TR-98−14*; pp 1−21.
(108) Joachims, T. Making Large-Scale SVM Learning Practical. *Technical Report for Computer Science Department*; University of Dortmund: Dortmund, Germany, 1998; *LS-8, report 24*; pp 11−55.
(109) Chandler, D. *Introduction to Modern Statistical Mechanics*. Oxford University Press: Oxford, U.K., 1987 pp 288.
(110) Dunweg, B.; Kremer, K. Molecular Dynamics Simulation of a Polymer Chain in Solution. *J. Chem. Phys.* **1993**, *99*, 6983−6997.
(111) Yeh, I. C.; Hummer, G. System-Size Dependence of Diffusion Coefficients and Viscosities from Molecular Dynamics Simulations with Periodic Boundary Conditions. *J. Phys. Chem. B* **2004**, *108*, 15873−15879.
(112) Tazi, S.; Botan, A.; Salanne, M.; Marry, V.; Turq, P.; Rotenberg, B. Diffusion coefficient and Shear Viscosity of Rigid Water Models. *J. Phys.: Condens. Matter* **2012**, *24*, 284117−284121.
(113) Mills, R. Self-Diffusion in Normal and Heavy Water in the Range 1−45°. *J. Phys. Chem.* **1973**, *77*, 685−688.
(114) Hardy, E. H.; Zygar, A.; Zeidler, M. D.; Holz, M.; Sacher, F. D. Isotope Effect on the Translational and Rotational Motion in Liquid Water and Ammonia. *J. Chem. Phys.* **2001**, *114*, 3174−3181.
(115) van der Spoel, D.; van Maaren, P. J.; Larsson, P.; Tumneanu, N. Thermodynamics of Hydrogen Bonding in Hydrophilic and Hydrophobic Media. *J. Phys. Chem. B* **2006**, *110*, 4393−4398.
(116) Stillinger, F. H. Theory and Molecular Models for Water. *Adv. Chem. Phys.: Non-Simple Liquids*; Prigogine, I., Rice, S., Eds.; J. Wiley and Sons, Inc.: New York, 1975; Vol. *31*, pp 1−101.
(117) Rapaport, D. C. Hydrogen Bonds in Water. *Mol. Phys.* **1983**, *50*, 1151−1162.
(118) Luzar, A.; Chandler, D. Hydrogen-Bond Kinetics in Liquid Water. *Nature* **1996**, *379*, 55−57.
(119) Luzar, A.; Chandler, D. Effect of Environment on Hydrogen Bond Dynamics in Liquid Water. *Phys. Rev. Lett.* **1996**, *76*, 928−931.
(120) Luzar, A. Resolving the Hydrogen Bond Dynamics Conundrum. *J. Chem. Phys.* **2000**, *113*, 10663−10675.
(121) Soper, A. K.; Phillips, M. G. A New Determination of the Structure of Water at 25 °C. *Chem. Phys.* **1986**, *107*, 47−60.
(122) Hura, G.; Sorenson, J. M.; Glaeser, R. M.; Head-Gordon, T. A High-Quality X-ray Scattering Experiment on Liquid Water at Ambient Conditions. *J. Chem. Phys.* **2000**, *113*, 9140−9148.
(123) Varma, S.; Teng, M.; Scott, H. L. Non-Intercalating Contact Points Create Asymmetry between Bilayer Leaflets. *Langmuir* **2012**, *28*, 2842−2848.
(124) Kumar, R.; Schmidt, J. R.; Skinner, J. L. Hydrogen Bonding Definitions and Dynamics in Liquid Water. *J. Chem. Phys.* **2007**, *126*, 204107−12.

(125) Krimm, S. Hydrogen Bonding of C−H···O═C in Proteins. *Science* **1967**, *158*, 530−531.
(126) Desiraju, G. R. The C−H···O Hydrogen Bond: Structural Implications and Supramolecular Design. *Acc. Chem. Res.* **1996**, *29*, 441−449.
(127) Gu, Y. L.; Kar, T.; Scheiner, S. Fundamental Properties of the CH··· O Interaction: is it a True Hydrogen Bond? *J. Am. Chem. Soc.* **1999**, *121*, 9411−9422.
(128) Horowitz, S.; Yesselman, J. D.; Al-Hashimi, H. M.; Trievel, R. C. Direct Evidence for Methyl Group Coordination by Carbon-Oxygen Hydrogen Bonds in the Lysine Methyltransferase SET7/9. *J. Biol. Chem.* **2011**, *286*, 18658−18663.
(129) Scheiner, S.; Kar, T.; Gu, Y. L. Strength of the (CH)-H-α··O Hydrogen Bond of Amino Acid Residues. *J. Biol. Chem.* **2001**, *276*, 9832−9837.
(130) Fenn, E. E.; Wong, D. B.; Fayer, M. D. Water Dynamics at Neutral and Ionic Interfaces. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 15243−15248.
(131) Zasetsky, A. Y. Dielectric Relaxation in Liquid Water: Two Fractions or Two Dynamics? *Phys. Rev. Lett.* **2011**, *107*, 117601−117605.
(132) van der Spoel, D.; van Maaren, P. J.; Berendsen, H. J.C. A Systematic Study of Water Models for Molecular Simulation: Derivative of Water Models Optimized for Use with A Reaction Field. *J. Chem. Phys.* **1998**, *108*, 10220−10230.
(133) Damm, K. L.; Carlson, H. A. Gaussian-Weighted RMSD Superposition of Proteins: A Structural Comparison for Flexible Proteins and Predicted Protein Structures. *Biophys. J.* **2006**, *90*, 4558−4573.
(134) Wolfe, K. C.; Chirikjian, G. S. Quantitative Comparison of Conformational Ensembles. *Entropy* **2012**, *14*, 213−232.
(135) Varma, S.; Chiu, S. W.; Jakobsson, E. The Influence of Amino Acid Protonation States on Molecular Dynamics Simulations of a Bacterial Porin OmpF. *Biophys. J.* **2006**, *90*, 112−123.
(136) Varma, S.; Jakobsson, E. The $cPLA_2$ C2α Domain in Solution: Structure and Dynamics of Its Ca^{2+}-Activated and Cation-Free States. *Biophys. J.* **2007**, *92*, 966−976.