



Published in final edited form as:

*J Phys Chem B*. 2005 September 1; 109(34): 16529–16538. doi:10.1021/jp051569v.

## 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

Byungchan Kim, Tom Young, Edward Harder, Richard A. Friesner, and B. J. Berne\*

Department of Chemistry and Center for Bimolecular Simulation, Columbia University, 3000 Broadway, New York, New York 10027

### Abstract

To isolate the effects of the inclusion of polarizability in the force field model on the structure and dynamics of the solvating water in differing electrostatic environments of proteins, we present the results of molecular dynamics simulations of the bovine pancreatic trypsin inhibitor (BPTI) in water with force fields that explicitly include polarization for both the protein and the water. We use three model potentials for water and two model potentials for the protein. Two of the water models and one of the protein models are polarizable. A total of six systems were simulated representing all combinations of these polarizable and nonpolarizable protein and water force fields. We find that all six systems behave in a similar manner in regions of the protein that are weakly electrostatic (either hydrophobic or weakly hydrophilic). However, in the vicinity of regions of the protein with relatively strong electrostatic fields (near positively or negatively charged residues), we observe that the water structure and dynamics are dependent on both the model of the protein and the model of the water. We find that a large part of the dynamical dependence can be described by small changes in the local environments of each region that limit the local density of non-hydrogen-bonded waters, precisely the water molecules that facilitate the dynamical relaxation of the water–water hydrogen bonds. We introduce a simple method for rescaling for this effect. When this is done, we are able to effectively isolate the influence of polarizability on the dynamics. We find that the solvating water's relaxation is most affected when both the protein and the water models are polarizable. However, when only one model (or neither) is polarizable, the relaxation is similar regardless of the models used.

### I. Introduction

The importance of water–protein interactions has been cited in the dynamic<sup>1,2</sup> and structural properties of proteins.<sup>3,4</sup> Understanding how biological solutes such as proteins perturb the structural and dynamic properties of proximal water molecules and how in turn this effects the properties of the protein is a necessary step toward characterizing the role of water in solvated biological systems.

Using simple force fields to simulate biomolecular systems at the atomic level provides insight into structural and dynamic details not yet available through experimental techniques.

Commonly the electrostatic part of such force fields employs a system of fixed point charges interacting via Coulomb's law.<sup>5–7</sup> However, the electronic structures of molecules depend on their environment. This is clearly manifested in water by the magnitude of the average dipole moment, which is approximately 40% larger<sup>8,9</sup> in the liquid phase compared to that in the gas phase. For homogeneous systems, such as neat fluids, ignoring polarization in modeling the

\*Author to whom correspondence should be addressed. E-mail: berne@chem.columbia.edu.

electrostatic energy is often deemed sufficient. However, recent work has shown that the variation in the field of neat fluids appears to be significant enough that fixed charge models are not able to quantitatively capture dynamic properties.<sup>10</sup> In solvated proteins, the electrostatic environment ranges from nonpolar near hydrophobic residues to highly polar near hydrophilic and charged residues to a bulk-water-like environment far from the protein. It is very likely that inhomogeneous systems with spatially varying electrostatic fields require the explicit inclusion of polarization to properly treat the electrostatic potential.

Intuitively, one would expect the importance of polarization to vary depending on the specific residue–water interactions involved. How the inclusion of polarization into an explicit atom, solvation model affects the structure and dynamics of the protein and water is not yet well understood.

The primary focus of this work is twofold: first, to investigate the structure and dynamics of the solvating water in regions of the protein that have differing electrostatic environments and second to gain a better understanding of how the inclusion of polarizability in the force field models affects these properties in the different regions. The solvent-exposed surface of bovine pancreatic trypsin inhibitor (BPTI) provides a diversity in the electrostatic field environments that ranges from hydrophobic (e.g., alanine) to the strong field environs of charged residues such as lysine and aspartic acid. It is this diversity that makes the BPTI in explicit water system an ideal candidate for our study.

We simulated systems interacting through different polarizable and nonpolarizable model force fields. The water models employed are the polarizable TIP4P-FQ<sup>11</sup> and RPOL<sup>12</sup> models and the nonpolarizable TIP4P model.<sup>13</sup> We also used two model force fields for the protein, one of which includes explicit polarization (PFF)<sup>14</sup> while the other does not (OPLS-AA).<sup>6,15</sup> All combinations of the water and protein force field models were simulated for a total of six systems. We refer the reader to the above cited articles for details of the water and protein models.

We investigated the structure and dynamics of the solvating water with respect to the local radial distribution, dipole moment distribution, and hydrogen-bond dynamics. Each of these measures is system-dependent and varies due to a number of model characteristics such as the Lennard-Jones parameters, the nonpolarizable aspects of the electrostatic parameters, as well as the polarizable contributions to the interactions. When possible, we comment on whether the differences noted in each of the above measures is due to polarization.

In sections 2 and 3, we give a brief outline of the model force fields and simulation details. Section 4 compares the tertiary structures of the protein obtained from simulation to experimental results. Section 5 outlines the structural and dynamical quantities that we calculated from the simulation trajectories. Section 6 gives the definitions of the electrostatic regions of the protein. Sections 7 and 8 outline the results for the structural and dynamical quantities. Sections 9 and 10 discuss these results.

## II. Protein and Water Models

The polarizable model used for the polypeptide (PFF) comes from the work of Kaminski et al.<sup>14</sup> The model places fixed partial charges on all atomic positions and on massless virtual sites representing the lone pairs of the oxygen and sulfur atoms. The electrostatic parameters are fit from gas-phase electronic structure calculations<sup>16</sup> using density-functional theory (DFT) with the B3LYP method<sup>17,18</sup> and the cc-pVTZ(-f) basis set. The dipole polarizabilities of the atomic sites are parametrized by a series of electrostatic perturbations, using dipolar probes applied to the target molecule. The resulting change in the electrostatic potential is measured at a set of grid points outside the van der Waals surface of the molecule. Polarizabilities are chosen to

minimize deviations from the DFT calculation. The fixed charges and the other remaining parameters are chosen to best approximate the electrostatic potential from the unperturbed DFT calculation. Stretching and bending energies for PFF are retained from the OPLS-AA force field<sup>6</sup> while the torsional energy is reparametrized.<sup>14</sup> Further details can be found in the respective references.<sup>6,14</sup> The electrostatic energy consists of a system of fixed point charges and point polarizable dipoles interacting according to the Coulomb potential. The 1–2 and 1–3 interactions are omitted owing to the breakdown of the bare Coulomb potential at such short intersite distances. The Coulomb potential is screened for specific residue–water interactions as described in ref<sup>19</sup>. Short-range repulsion and dispersion is represented by a Lennard-Jones function, where we apply the geometric sum rule ( $\sigma_{ij} = (\sigma_i\sigma_j)^{1/2}$  and  $\epsilon_{ij} = (\epsilon_i\epsilon_j)^{1/2}$ ) for the interaction between particle  $i$  and  $j$ . A potential scaling factor for the Lennard Jones part of the potential is set to zero for particles connected by a valence bond or angle, set to 0.5 for intramolecular 1–4 interactions, and is 1.0 for all other pairs. The Lennard-Jones parameters are derived from *ab initio* dimer energies of organic compound analogues of the residues and from the OPLS-AA force field.

We employ three commonly used water models to solvate the polypeptide, A fixed charge TIP4P<sup>13</sup> model and two polarizable water models, a TIP4P-FQ<sup>11</sup> fluctuating charge model and an inducible point dipole model RPOL.<sup>12</sup> All three models employ an interacting Lennard-Jones site placed on the oxygen atom. Intermolecular interactions between electrostatic sites are described by the bare Coulomb potential. The TIP4P-FQ model includes an intramolecular interaction between the charges within the molecule that is parametrized along with the other electrostatic parameters empirically. The RPOL model places point polarizable dipoles on the oxygen and hydrogen atoms and omits intramolecular electrostatic interactions.

Each of the two protein force fields (the fixed charge OPLS-AA and the polarizable PFF) are combined with the three water force fields (fixed charge TIP4P and polarizable TIP4P-FQ and RPOL) to give a total of six model systems.

### III. Simulation Details

The simulation procedure is as follows. The starting structure was obtained from the BPTI (PDB code 4PTI) structure in the Brookhaven Protein Data Bank.<sup>20</sup> This crystal structure included a protein monomer and 60 water molecules. The protocol for preparation and equilibration of the solvated protein with each of the six interaction models followed the procedure outlined in by Harder et al.<sup>19</sup> To ensure an accurate evaluation of the particle dynamics, the simulations are carried out in the microcanonical ensemble ( $NVE$ ), free of artificial perturbations necessary for simulations in the isothermal/isobaric ensemble.<sup>21,22</sup> The production simulations were run using the velocity Verlet algorithm for 2 ns with a 1 fs time step (0.75 fs for PFF/TIP4P-FQ). The initial configurations for the simulations were taken from equilibrated simulations at a constant temperature of 298 K and a constant pressure of 1 atm. A further 500 ps of equilibration in  $NVE$  was found to be necessary, leaving the final 1.5 ns for analysis. The RATTLE<sup>23</sup> constraint algorithm was used to keep the water molecular geometry rigid, and the bonds between the protein heavy atoms and hydrogens were held fixed. A generalized P3M Ewald method to include point dipoles in addition to point charges is used to resolve the electrostatic energy.<sup>19</sup> The Ewald parameters for the simulations were  $\eta = 0.37 \text{ \AA}^{-1}$ , a spherical truncation of the real space potential at  $R_{\text{cut}} = 10 \text{ \AA}$ , a grid spacing of 0.75  $\text{\AA}$ , and an assignment order  $P = 6$ . The minimum energy electrostatic configuration is solved iteratively at the outset of each simulation segment. The extended Lagrangian method was used to propagate these variables during the simulation.<sup>11,24</sup> We refer the reader to a previous work for the exact details of the simulation protocol.<sup>19</sup>

## IV. Structure of the Protein

In this section, we provide a brief analysis of the tertiary structure of the protein to assure the reader that different combinations of force fields yield reasonable tertiary structures for the protein. A more in depth analysis of the system dependence of the protein structure will be presented in a future work.

Figure 1 shows the root-mean-square deviation (RMSD) of the backbone  $\alpha$ -carbons for the three of the simulated systems (OPLS-AA/TIP4P-FQ, PFF/TIP4P, and OPLS-AA/RPOL) with respect to the crystal structure.<sup>20</sup> Results for the other three systems (OPLS-AA/TIP4P, PFF/TIP4P-FQ, and PFF/RPOL) were presented in a previous work.<sup>19</sup> Figure 2 shows the RMSD of the protein's heavy atoms for the same systems with respect to the crystal structure. The insets in each figure are the RMSD values of 20 different NMR structures<sup>25</sup> compared to the experimental crystal structure.<sup>20</sup> The variations from the crystal structure for all six systems are comparable to the variations seen between the NMR and the crystallographical experiments.

Table 1 shows the time-averaged values of the RMSD. For convenience, the values for all six systems are shown here.

## V. Measured Quantities

### A. Local Radial Distribution Function

The solvent accessible to a given protein atom is highly dependent on the local protein structure and hence is dependent on the location of the atom in the protein as a whole. This presents a well-known difficulty in calculating a meaningful parallel to the standard radial distribution function. In this work, we follow the prescription of Brooks et al.<sup>26</sup> where the local density of water around a specific protein atom is defined as

$$\rho_{\text{local}} = \frac{\langle N_w \rangle}{V_R}$$

where  $\langle N_w \rangle$  is the average number of solvent atoms (in this case the number of water oxygen atoms) within a certain radius of the protein atom and  $V_R$  is the volume of the sphere of a chosen local radius  $R$ , which in our case was chosen to be 6 Å. The local radial distribution-function with respect to particle  $i$ ,  $g_i^{\text{local}}$ , is computed as follows

$$g_i^{\text{local}}(r) = \frac{\langle \delta N(r) \rangle}{4\pi r^2 \delta r \rho_{\text{local}}}$$

where  $\langle \delta N(r) \rangle$  is the number of water oxygen atoms between  $r$  and  $r + \delta r$ .

### B. Hydrogen Bond and Hydrogen-Bond Relaxation

We employ a commonly used definition of a hydrogen bond. A hydrogen bond between two waters is defined by the following two criteria:<sup>27</sup> First, the oxygen–oxygen distance between the two waters must be less than or equal to 3.5 Å. Second, the angle between the oxygen1–oxygen2 vector and the oxygen–hydrogen bond vector must be less than or equal to 30°. The hydrogen-bond correlation function is defined as<sup>27</sup>

$$c(t) = \frac{\langle h(0)h(t) \rangle}{\langle h \rangle}$$

where  $h(t) = 1$  if a given hydrogen bond exists at time  $t$  and  $h(t) = 0$  if it does not.

We limit our investigation to hydrogen bonds that exist between two water molecules, one of which is in the solvation shell of the protein and the other is not.

### C. Number of Adjacent Waters

In the work of Xu and Berne,<sup>28</sup> a correlation between the number of adjacent non-hydrogen-bonded waters and the dynamical relaxation of hydrogen bonds was noted. A similar analysis is presented here.

For each region, the number of adjacent waters was calculated in a standard way. An imaginary spherical shell of radius 3.5 Å is drawn around the oxygen of a tagged water, and the number of water oxygens that lie within this sphere (excluding the tagged water's oxygen) is considered to be the total number of neighboring waters  $n_t$ . This number is then split into the sum of two quantities. The number of adjacent waters that are involved in hydrogen bonds with the tagged water is  $n_{hb}$ , and the number that are not is  $n_{adj}$ .

## VI. Definition of Electrostatic Regions

Part of this work is dedicated to understanding the effect that differing electrostatic environments have on the structure, energetics, and dynamics of the solvating water. To investigate this, we have defined five regions that we will refer to as the bulk, hydrophobic, hydrophilic, positively charged (lysine), and negatively charged (aspartic acid) regions. These regions are roughly ordered with regards to the increasing strength of their respective electrostatic fields.

We define a water molecule to be in the bulk region if the point position of its oxygen atom lies outside an imaginary shell drawn around the protein 6 Å from the protein atom. The structure and dynamics of the water molecules in this region are similar to that of the neat system. We also determined the average dipole moment, average number of hydrogen bonds, and hydrogen-bond relaxation dynamics in the region outside a shell of 8 Å and found no variation in these properties from the 6 Å shell within the statistical error.

A water molecule is considered to be in a given electrostatic region if it is proximal to region-specific protein atoms. These protein atoms for the hydrophobic region are defined as the side-chain carbon atoms of the hydrophobic residues alanine, isoleucine, leucine, phenylalanine, proline, and valine. The protein atoms of interest for the hydrophilic region are the backbone oxygens of any uncharged residue. The protein atoms of interest for the positively charged and negatively charged regions are the side-chain nitrogens of lysine and the side-chain oxygens of aspartic acid, respectively. In addition, we employed the following restrictions to each of the protein surface regions:

1. Each protein atom of interest has a proximal water in two-thirds of the configurations generated by the simulations. Here, we take proximal to mean within 3.5 Å of the atom of interest for the hydrophilic and charged regions and 4.0 Å for the hydrophobic region. Additionally, the protein atom of interest is the closest protein heavy atom to the given water.<sup>29</sup>

2. No side-chain nitrogens or oxygens of charged residues lie within 6 Å of the atom of interest. This is determined from the initial structure.
3. The protein atom is not involved in an internal hydrogen bond in more than 10% of the configurations generated by the simulations.
4. Conditions 1 and 3 are met for all combinations of force fields.

## VII. Structural Results

### A. Local Radial Distribution Function

We calculated the local radial distribution function of the protein and proximal water for each of the six systems in each of the four electrostatic regions of the protein.

In this study, we wish to isolate the effects of polarizability on the structure and dynamics of the system. This is nontrivial since differences in the radial distribution function can reflect both the influence of polarizability as well as other aspects of the interaction potential such as the Lennard-Jones and nonpolarizable electrostatic parameters. To aid in differentiating these contributions, we present two sets of graphs for the local radial distribution function in each region. One set in each figure contains three graphs labeled a, b, and c. Each graph in this set has one model of water grouped with both models of proteins. The other set in each figure has two graphs labeled d and e. Each graph in this set has the local radial distribution function (RDF) results for one model of protein with all three models of water. When the differences between curves are similar in each graph of a given set, it implies that the differences noted are likely a result of influences other than the polarizability. When the differences between curves change from graph to graph within a set, the differences can often be related to the influence of polarizability.

**1. RDF in the Hydrophobic Region**—For all systems studied, the local radial distribution functions in the hydrophobic region are very similar. The most noticeable difference is for the TIP4P-FQ model of water, which has a slightly more pronounced first peak with both protein potentials compared to those of the other two water models. We do not attribute this to polarizability for two reasons. First, such a trait is not apparent in the other polarizable model of water (RPOL). Second, we would expect this difference to be increased with the polarizable model of the protein due to the mutual enhancement of the polarization in both the protein and the water. Figure 3 shows that this is not the case. The curves for the systems of the TIP4P-FQ model show little dependence on the protein model and hence little dependence on whether the protein is polarizable in this weakly electrostatic region.

**2. RDF in the Hydrophilic Region**—Figure 4 shows the local radial distribution functions of the protein's carbonyl oxygen and neighboring water oxygens in the hydrophilic region. With the polarizable protein model and each model of water, the distance at which the local radial distribution functions become nonzero and the positions of the first peak are slightly shifted inward when compared to the nonpolarizable OPLS protein model. The distance of this shift is approximately 0.15 Å for each model of water. The waters in the hydrophilic region are slightly more closely packed to the PFF proteins than those of the OPLS protein. This trend is the same for all of the models of water, which implies that the differences are a result of influences other than the polarizable interactions.

For parts d and e in Figure 4, the trends for different models of water are again very similar. In both graphs, the curve for the TIP4P model of water is very slightly shifted inward and the first peak is somewhat more pronounced. The fact that these characteristics are similar in each graph indicates that they are most likely a result of differences in the nonpolarizable contributions of the various protein and water force fields.

**3. RDF in the Lysine Region**—Figure 5 shows the local radial distribution function in the vicinity of the positively charged nitrogen on the above-defined lysine residues. In Figures 5a–c, the curves for the PFF protein have first peaks that are narrower and higher and have peak positions that are shifted to lower values of  $r$ . These trends are consistent in all three graphs. No conclusive observations can be drawn about the effect of polarizability from this set of curves.

In the top-right graph (Figure 5d, OPLS protein), the three curves have peaks that are close to each other in height with the RPOL being slightly higher. However, in the bottom-right graph (Figure 5e, PFF protein), the first peak heights are higher and the peaks for the polarizable models of water are significantly higher than that for the TIP4P (nonpolarizable) water. All of the systems with the PFF protein have peak positions located at shorter values of  $r$ . This effect is greater for the polarizable models of water (TIP4P-FQ and RPOL) than it is for the TIP4P system when compared to the relative positions for the OPLS-AA systems. (This is most apparent when looking at the short side of the first peak.) Such behavior is consistent with what we observe for the water molecule's dipole moments in this region, which is that the combination of both polarizable protein and water induces a relatively large water dipole moment that affects the interaction energy profile and the minimum of the potential of mean force (radial distribution function).

It is interesting to note that in both Figures 5d and 5e the curves for RPOL become nonzero and peak at lower values of  $r$  than for the other models of water. Since this trend is similar in both graphs of this figure, we attribute this to the parametrization of the two models. However, these noted differences are enhanced in the polarizable protein graph, which suggests that the additional shift is due to the inclusion of polarizability in both the protein and the water models.

**4. RDF in the Aspartic Acid Region**—Figure 6 shows the local radial distribution function in the vicinity of the negatively charged oxygens of aspartic acid Asp3. Each of the three parts labeled a, b, and c in this figure shows a different water model with both protein models. In the graph for the nonpolarizable TIP4P water, the two curves are very similar. However, in Figures 6b and 6c, the curves for the systems with both polarizable water and protein and polarizable water have peak positions shifted to lower values of  $r$ , and the first peaks are somewhat less pronounced.

Parts d and e in Figure 6 show the curves for systems for all three models of water with each protein model, respectively. In Figure 6d, the differences between the three curves are slight. The graph labeled Figure 6e in this figure shows the results for systems with the PFF protein. In this graph, the curves for systems with polarizable water models have first peaks that are shifted to shorter values of  $r$  and are lower than the corresponding peak for the nonpolarizable TIP4P system.

The trends in Figure 6 show that the combination of both polarizable protein and polarizable water leads to significant changes in the local structure of the water in the vicinity of the negatively charged aspartic acid.

## B. Dipole Moments

The dipole moment of polarizable water models is a measure of how the differing electrostatic environments affect the molecular electronic structure of the water.

For the TIP4P model of water, the dipole moment is fixed and the distribution of dipole moments is trivially a delta function. However, the fluctuating charge (TIP4P-FQ) model and the point dipole polarizable (RPOL) model have dipole moments that can fluctuate. The analysis in this section will be limited to the four systems that have polarizable models of water.

Table 2 shows the mean dipole moment of water molecules in each electrostatic region. Figure 7 shows the dipole moment distribution of the water molecules in each of the electrostatic regions. In this figure, curves of the same color are for systems with the same polarizable water model. The dashed and solid lines correspond to systems with the polarizable (PFF) and nonpolarizable (OPLS-AA) protein models, respectively. In the bulk, hydrophobic, and hydrophilic regions, the dashed and solid curves are identical within the noise, indicating that the polarizability of the protein has little influence. In the stronger field regions (proximal to lysine and aspartic acid), the dashed and solid curves have clear differences. In the lysine region, for systems in which both the water and the protein are polarizable, the mean dipole moment and the position of the peaks in the dipole moment distributions are significantly increased from what is seen when the protein is not polarizable. For these same systems, in the aspartic acid region, the dipole moment distributions are noticeably broader and the mean dipole moments are slightly higher.

## VIII. Hydrogen-Bond Dynamics on the Protein Surface

In this section, we will describe the results of our analysis of the hydrogen-bond dynamics in the five regions defined above. We will also illustrate how the behavior in each of these electrostatic regions is dependent upon the model of protein, the model of water, or the combination of the two.

The hydrogen-bond correlation function is defined as above<sup>27</sup>

$$c(t) = \frac{\langle h(0)h(t) \rangle}{\langle h \rangle}$$

In general, this correlation function decays more slowly for polarizable models of water.<sup>30</sup> The decay of this correlation function is nonexponential though we can still define a relaxation time  $\tau_{\text{rlx}}$  that is the time such that  $c(\tau_{\text{rlx}}) = e^{-1} c(0)$ .

We present two sets of scaled units in our analysis of this relaxation time. The first takes into account that the different water models have different relaxation times in the neat systems. The second attempts to account for both this fact and for variations in the local structure that affect the number of waters that can facilitate hydrogen-bond relaxation. We call the two sets of times retarded and scaled retarded, respectively. Details follow in the next two subsections.

### A. Retarded Units

The hydrogen-bond relaxation is highly dependent on the model of water where polarizable models tend to decay more slowly than nonpolarizable models. This is a well-known result from computer simulations.<sup>30</sup> However, we wish to make meaningful comparisons between relaxation times for systems with different models of water. For this reason, we define the bulk hydrogen-bond relaxation time in each system as a system-dependent unit of time.

For each of the six simulations, the scaled relaxation in a region is defined as

$$\tilde{\tau}_{\text{ret}}^{\text{region}} = \frac{\tau_{\text{rlx}}^{\text{region}}}{\tau_{\text{rlx}}^{\text{bulk}}}$$



Properties presented in these units will be denoted with a tilde, and the times will be referred to as retarded times.

When the relaxation times of the water in the bulk region are represented in these retarded units, the times are identically equal to unity for all systems. When the relaxation times in other regions are represented in terms of these scaled units, we are able to make meaningful comparisons between the relative effect that the electrostatic regions have on the dynamics in different systems. The form of this scaling is similar to the calculation of retardation factors used to compare dynamical results from different experimental techniques.

## B. Scaled Retarded Units

The making and breaking of hydrogen bonds is a highly cooperative process. Since hydrogen-bond interactions are quite strong, it is somewhat reasonable to assume that the reactive pathway for the long-term breaking of hydrogen bonds involves an intermediary facilitating water. Such an assumption implies that thermal fluctuations alone do not lead to the breaking of hydrogen bonds in the system of interest.

In the course of our analysis, we have calculated the average number of waters proximal to each water molecule of interest. We will call this quantity, which is dependent on both the protein region and the system,  $n_t$ . We have further divided  $n_t$  into two categories, those proximal waters that are involved in a hydrogen bond with the water of interest ( $n_{hb}$ ) and those that are not ( $n_{adj}$ ). We have found it useful to consider that the rate of relaxation is linearly proportional to  $n_{adj}$ . This involves two major assumptions: first, that the relaxation of hydrogen bonds occurs with the facilitation of adjacent waters that are not hydrogen-bonded to the tagged water of interest<sup>28,31</sup> and second that the rate of such facilitation is directly proportional to the mean local density of adjacent waters that are not involved in hydrogen-bonding with the water of interest. The first assumption is a well-accepted principle in hydrogen-bond dynamics. The second assumption is a standard first-order approximation. Though this is a somewhat simple approximation, it has proven to be quite useful in our comparisons of dynamics in the different electrostatic regions.

We define the relaxational unit of time scaled by the number of adjacent non-hydrogen-bonded waters and the bulk relaxation in the following way

$$\tilde{\tau}_{adj} = \frac{\tau_{rlx}^{region} n_{adj}^{region}}{\tau_{rlx}^{bulk} n_{adj}^{bulk}}$$

We refer to times represented in these units as scaled retarded times. When the relaxation times are represented in these units, it accounts for a large number of factors that can influence the number of adjacent non-hydrogen-bonded neighbors ( $n_{adj}$ ). These include system-dependent variations in the local radial distribution functions, local structural variations within the residue itself, as well as tertiary structural variations that may occur in the different systems due to the flexibility of the protein.

Work by Luzar<sup>31</sup> proposes that to reach a transition state for diffusion (a long-term breaking of the hydrogen bond) a fluctuation in the structure that makes a hydrogen-bonding site available must occur. Such fluctuations occur when a water that is not hydrogen-bonded to the water of interest makes such a site available. The use of scaled retarded units assumes that the rate at which such a transition state is available is linearly proportional to the local density of non-hydrogen-bonded waters. If this assumption is valid, then deviations from unity in the

scaled retarded times would indicate changes induced in the free energy profile of the transition state compared to the free energy profile of the bulk.

### C. Bulk Water Dynamics

We define a water to be in the bulk region if its oxygen atom lies outside an imaginary shell drawn around the protein 6 Å from the closest protein atom. When the relaxation times are calculated, only pairs of waters that both lie in the bulk region are included.

The hydrogen-bond relaxation times in the bulk for each of the six simulations are shown in Table 3. For comparison, published results from Xu et al.<sup>30</sup> for the relaxation in neat water are given in Table 4. The relaxation times in the bulk region are close to those reported for neat water at the temperature of interest. For the TIP4P and RPOL models of water, the bulk relaxation appears to be somewhat slower around the polarizable protein than that around the nonpolarizable protein. However, these differences are consistent with variations noted in the mean kinetic energies of the systems (also in Table 4). Variance in the mean kinetic energies between the systems is observed since only a single configuration was sampled from the *NPT* ensemble of each system, respectively.

In this study, there is no real indication that the hydrogen-bond relaxation times of waters in the bulk are influenced by the model of the protein. This is consistent with our analysis of the dipole moment. In this study, we found no evidence that the dynamical or electrostatic properties of water in the bulk region differed from those in the corresponding neat systems. This is also consistent with experimental results.<sup>1,32</sup>

### D. Relaxation in the Hydrophobic Region

Table 5 shows the hydrogen-bond relaxation times in the hydrophobic region represented in unscaled, retarded, and scaled retarded units.

Though the unscaled times are significantly different for each of the six systems, the retarded times are very similar. All six systems have relaxations that are between 1.4 and 1.49 times slower than those in the corresponding bulk with a slightly stronger retardation being noted for the RPOL models of water.

When the relaxation rates are represented in the scaled retarded units, the relaxation for all six systems is close to unity. This implies that a good deal of the retardation in the hydrophobic region can be explained solely on the basis of the lowered local density of waters capable of facilitating the relaxation.

The retarded and scaled retarded times are similar for all six systems studied in this region.

### E. Relaxation in the Hydrophilic Region

Table 6 shows the retarded relaxation times ( $\tilde{\tau}_{\text{ret}}$ ) in the hydrophilic region for the six systems studied. These relaxation times are roughly grouped by the model of water with the TIP4P-FQ systems being somewhat slower. However, the variation in the retarded relaxation times is not large. The same table shows the relaxation in scaled retarded units. When the times are scaled by  $n_{\text{adj}}$ , the dynamics are similar in all six systems.

For this weakly electrostatic environment, no significant variation in the hydrogen-bond relaxation dynamics is noted between the six systems studied.

## F. Relaxation in the Positively Charged Lysine Region

Table 7 shows the hydrogen-bond relaxation times in the positively charged lysine region represented in unscaled, retarded, and scaled retarded units. The unscaled relaxation times vary considerably from system to system. However, the systems with polarizable water relax more slowly than the systems with nonpolarizable water. Also, each system with polarizable water relaxes more slowly with the polarizable protein than with the nonpolarizable protein.

When the relaxation times are scaled by the bulk relaxation times, a clear trend becomes apparent. The two systems in which both the protein and the water models are polarizable relax more slowly than the other four systems. However, the other four systems have both retarded and scaled retarded relaxation times that are similar. The scaled retarded relaxation times are very similar for these four systems.

## G. Relaxation in the Negatively Charged Aspartic Acid Region

Table 8 shows the hydrogen-bond relaxation times in the aspartic acid region represented in unscaled, retarded, and scaled retarded units.

In unscaled units, the relaxation times are somewhat similar for systems with the same models of water and there is a seemingly small dependence on the model of the protein. When the relaxation times are expressed in retarded units, this weak trend is still apparent. However, when the relaxation times are expressed in scaled retarded units, a different and clear trend becomes apparent. The relaxation times for the systems in which both the protein and the water are polarizable are slower than those for the other four systems. However, the other four systems have both retarded and scaled retarded relaxation times that are similar to each other. This trend for scaled retarded relaxation is the same as what was observed in the positively charged lysine region.

## IX. Comments

Here, we give a short comment on the results presented for each of the properties investigated.

### A. Hydrogen-Bond Relaxation

We have presented an analysis of the hydrogen-bond dynamics that uses the relaxation in the bulk as a system-dependent reference state. The method of analysis also scales for variations in the local environment as reflected in  $n_{\text{adj}}$ . This method of analysis provides a means by which the influence of polarizability in the water and the protein models can be isolated. It also provides a means for accounting for variations in the local water environment of the residues. For the four systems in which both the protein and the water models did not include explicit polarizability, the scaled retarded dynamics were similar in all regions. This similarity does not imply that each of the simulated systems gives comparable results for the dynamics in the solvation shell. They do not. However, the similarity strongly suggests that the separation of the contributions does correctly describe the dynamics.

The results of our analysis of the dynamics have several interesting implications. First, the bulk water system can be used as a reference state for understanding the dynamics on the surface of a protein. This is important since each water model has bulk dynamics that vary. Without such a reference state, comparisons between systems with differing models of water are relatively meaningless. Second, the dynamics of solvation shell water are very sensitive to small local changes in the structural environment (as reflected in the local density of non-hydrogen-bonded water molecules). Furthermore, this dependence on the local environment is on a length scale shorter than common simulational structural measures of proteins; i.e., differences in  $n_{\text{adj}}$  that have a large effect on the dynamics are not apparent in standard RMSD

analyses. Third, mutual enhancement has the effect of slowing the hydrogen-bond dynamics in the regions of the protein surface with a strong electrostatic field.

Our results imply that to accurately simulate the dynamics of water in the solvation shell of proteins all three of the above would need to be accurate, the bulk water dynamics, the fine local structure, and the influence of mutually enhancing polarizability.

## B. Dipole Moment Distribution

The results for the dipole moment distribution for the four systems simulated with polarizable models of water were mostly straightforward. In the regions with a strong electrostatic field (lysine and aspartic acid), the mean dipole moment for a given model of water was larger when the protein model was polarizable (PFF) than when it was not (OPLS). This is consistent with a mutual enhancement of the polarizabilities.

For all four polarizable water systems, the dipole moments for the solvating waters in the hydrophobic region of the protein were only slightly reduced from those in the bulk. Also, the mean dipole moments of the solvating waters in the hydrophobic region are higher than those found in the hydrophilic region. We attribute both of the above to the fact that the average number of hydrogen bonds between water molecules in the hydrophobic region is only slightly lower than that found in the bulk and in the hydrophilic region the average number is lower by close to one.

We also observed that the TIP4P-FQ model of water has a much broader distribution of dipole moments particularly in the negatively charged aspartic acid region. We attribute this to an enhanced polarizability in the H–O–H plane of this model.

## C. Radial Distribution Function

The local radial distribution function of solvating waters is very sensitive to the parameters of the model system. This sensitivity is due to factors both related and unrelated to the inclusion of polarizability. However, we were able to qualitatively isolate the influences of polarizability on the local radial distribution functions in the various regions. The results were consistent with the other properties investigated. In regions of the protein where the electrostatic field is weak (hydrophobic and hydrophilic regions), the effect is small and indiscernible. In regions where the field is strong (lysine and aspartic acid), the effect is small or indiscernible except when both the protein and the water models include explicit polarizability. In this case, the effect on the local radial distribution function is consistent with enhancing the attractive interactions between the protein and water. This scenario is consistent with mutual enhancement of the polarizabilities.

## X. Conclusions

We have investigated the structure and dynamics of water in the solvation shell of BPTI with six simulated systems that have a combination of polarizable and nonpolarizable models of both protein and water. We have calculated these properties in regions of the protein surface distinguished by the local electrostatic environment. In each of the electrostatic regions, we have attempted, with some degree of success, to isolate what effects on the structure and dynamics result from the inclusion (or exclusion) of polarizability in the model systems.

From model combinations where either the protein or water model do not include electrostatic polarization, we find the effect of polarization on the studied properties to be minimal. In contrast, we see a qualitative change in behavior when both the protein and the water model can polarize. The impact is most important in regions of the protein that have a strong electrostatic field (i.e., near to charged residues). In these regions, the systems capable of

mutually enhancing interactions have stronger protein–water bonds (as reflected in the shift in the local radial distribution function), the dipole moment distribution of the solvating waters becomes broader with the peak shifted to higher values, and the hydrogen-bond dynamics are slower.

Water and protein molecules have electrostatic properties that are capable of responding to their environments. When they interact, their electrostatic fields are able to mutually adjust. In our simulations, the ability of the water and the protein to have mutually enhancing electrostatic interaction had a significant effect on both the structural and the dynamical properties of the solvating water in the vicinity of charged residues. Appropriately modeling the solvation of charged residues, in turn, affects a number of crucial properties of a protein. For example, the formation of salt bridges is dependent on the relative free energy of the solvated charged residues. Similarly, the rate of protein folding is likely sensitive to the dynamics of the solvating water.

The effects of mutual enhancement on the structure and the dynamics of water strongly suggest that the inclusion of polarizability in both the protein and the water models should not be ignored. This is particularly true for systems in which the electrostatic environment varies significantly from neat water.

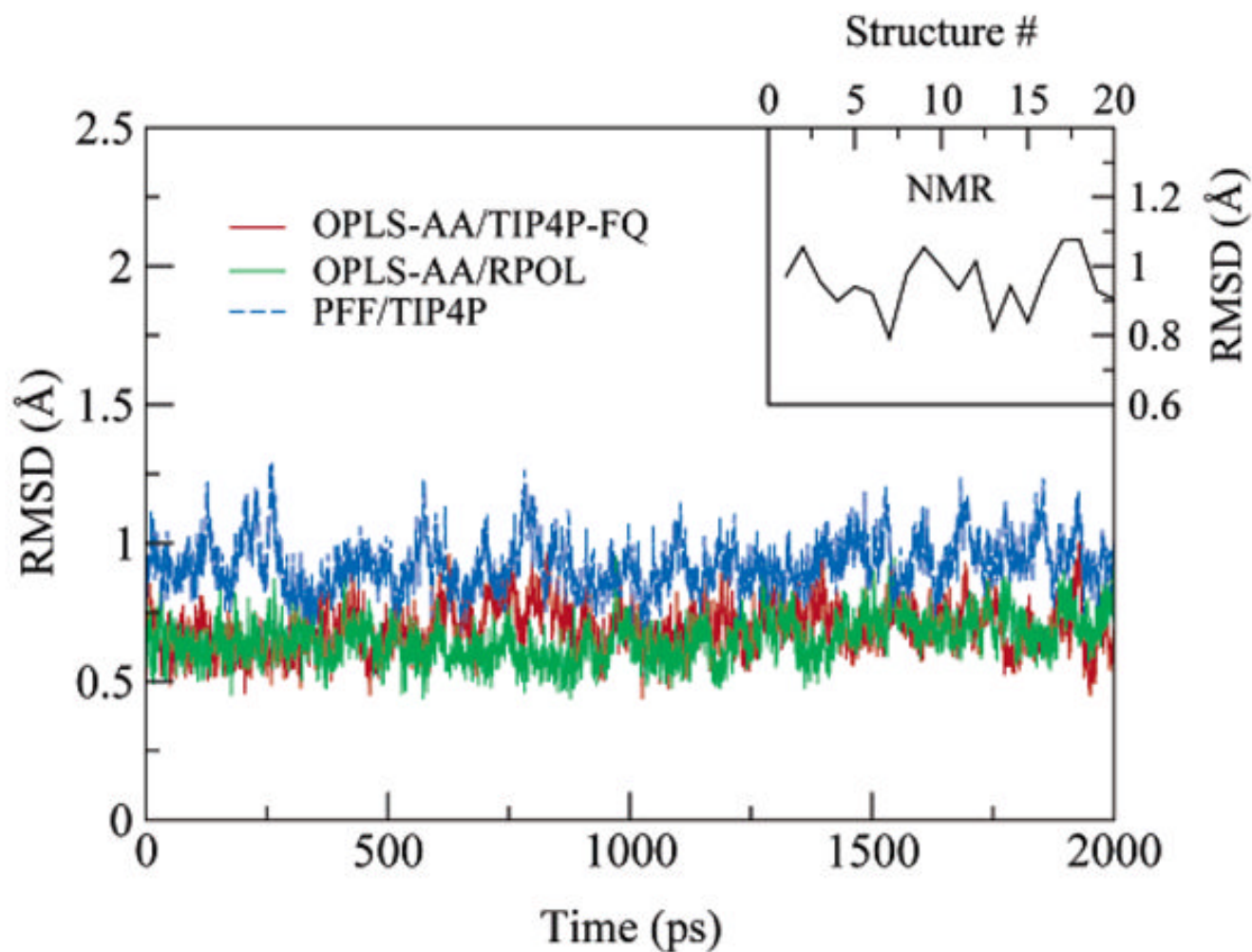
## Acknowledgments

This work was supported by grants to B. J. Berne from the National Institutes of Health (GM4330) and Richard A. Friesner from the National Institutes of Health (GM52018) and by a grant of computer time from the Environmental Molecular Sciences Laboratory at the Pacific Northwest National Laboratory. B.K. and T.Y. contributed equally to this work.

## References and Notes

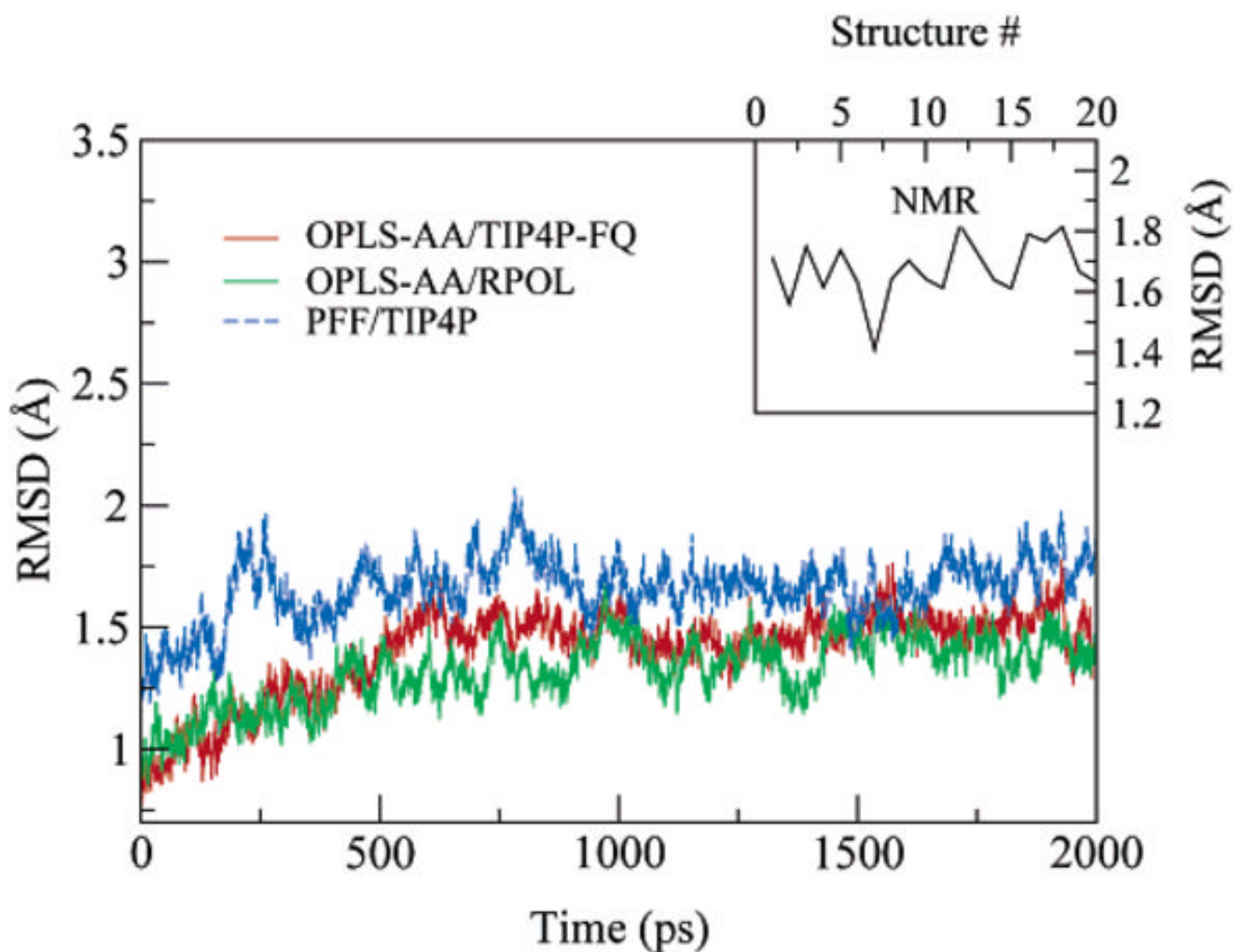
1. Halle B, Davidovic M. *Proc Natl Acad Sci USA* 2003;100:12135. [PubMed: 14528004]
2. Barron LD, Hecht L, Wilson G. *Biochemistry* 1997;36:13143. [PubMed: 9376374]
3. Papoian GA, Ulander J, Eastwood MP, Luthey-Schulten Z, Wolynes PG. *Proc Natl Acad Sci USA* 2004;101:3352. [PubMed: 14988499]
4. Zhou RH, Berne B. *Proc Natl Acad Sci USA* 2002;99:12777. [PubMed: 12242327]
5. Mackerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck J, Field MJ, Fisher S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau F, Mattos C, Michnick S, Ngo T, Nguyen D, Prodhorn B, Reiher WE III, Roux B, Schlenkrich M, Smith J, Stote R, Straub J, Watanabe M, Wirkiewicz-Kuczera D, Yin D, Karplus M. *J Phys Chem B* 1998;102:3586.
6. Jorgensen WL, Maxwell DS, Tirado-Rives J. *J Am Chem Soc* 1996;118:11225.
7. Cornell W, Cieplak P, Bayly C, Gould I, Merz KM Jr, Ferguson D, Spellmeyer D, Fox T, Caldwell J, Kollman P. *J Am Chem Soc* 1995;117:5179.
8. Gubskaya AV, Kusalik PG. *J Chem Phys* 2002;117:5290.
9. Kuo IFW, Mundy CJ. *Science* 2004;303:658. [PubMed: 14752157]
10. Fecko C, Eaves J, Loparo J, Tokmakoff A, Geissler P. *Science* 2003;301:1698. [PubMed: 14500975]
11. Rick SW, Stuart SJ, Berne BJ. *J Chem Phys* 1994;101:6141.
12. Dang LX. *J Chem Phys* 1992;97:2659.
13. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein M. *J Chem Phys* 1983;79:926.
14. Kaminski GA, Stern HA, Berne BJ, Friesner RA, Cao YX, Murphy RB, Zhou R, Halgren TA. *J Comput Chem* 2002;23:1515. [PubMed: 12395421]
15. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. *J Phys Chem B* 2001;105:6474.
16. Stern HA, Kaminski GA, Banks JL, Zhou R, Berne BJ, Friesner RA. *J Phys Chem B* 1999;103:4730.
17. Becke AD. *Phys Rev A* 1988;38:3098. [PubMed: 9900728]
18. Lee C, Yang W, Parr RG. *Phys Rev B* 1988;37:785.

19. Harder E, Kim B, Friesner RA, Berne BJ. *J Chem Theory Comput* 2005;1:169.
20. Deisenhofer J, Steigemann W. *Acta Crystallogr, Sect B: Struct Sci* 1975;31:238.
21. Tuckerman ME, Martyna GJ, Klein ML. *J Chem Phys* 1992;97:2635.
22. Martyna GJ, Tobias DJ, Klein ML. *J Chem Phys* 1994;101:4177.
23. Andersen HC. *J Comput Phys* 1983;52:24.
24. Sprik M, Klein ML. *J Chem Phys* 1988;89:7556.
25. Berndt KD, Guntert P, Orbons LPM, Wuthrich K. *J Mol Biol* 1992;227:757. [PubMed: 1383552]
26. Brooks CL, Karplus M. *J Mol Biol* 1989;208:159. [PubMed: 2769750]
27. Luzar A, Chandler D. *Nature* 1996;379:55.
28. Xu HF, Berne B. *J Phys Chem B* 2001;105:11929.
29. Cheng YK, Rossky PJ. *Nature* 1998;392:696. [PubMed: 9565030]
30. Xu HF, Stern HA, Berne BJ. *J Phys Chem B* 2002;106:2054.
31. Luzar A. *Faraday Discuss* 1996;103:29.
32. Pal SK, Peon J, Zewail AH. *Proc Natl Acad Sci USA* 2002;99:1763. [PubMed: 11842218]



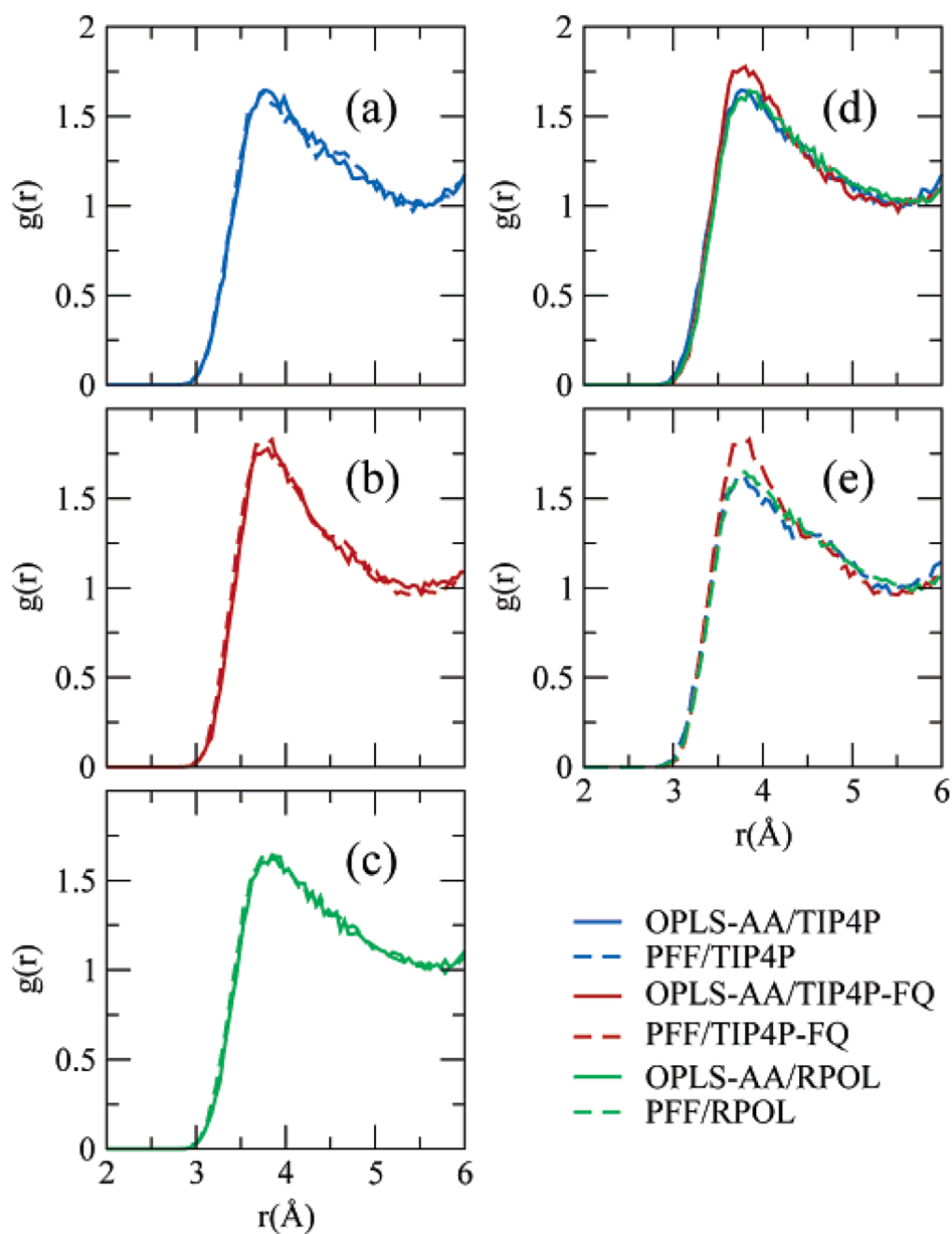
**Figure 1.**

Root-mean-square deviation of C $\alpha$  atoms between the simulation structures and the experimental crystal structure as a function of simulation time. Terminal residues, which show large fluctuations from NMR experiments as well as simulation, are not included in this analysis. The inset shows the RMSD between the 20 NMR structures and the crystal structure. All models do a reasonable job representing the protein native state for the 2 ns duration of the simulation.



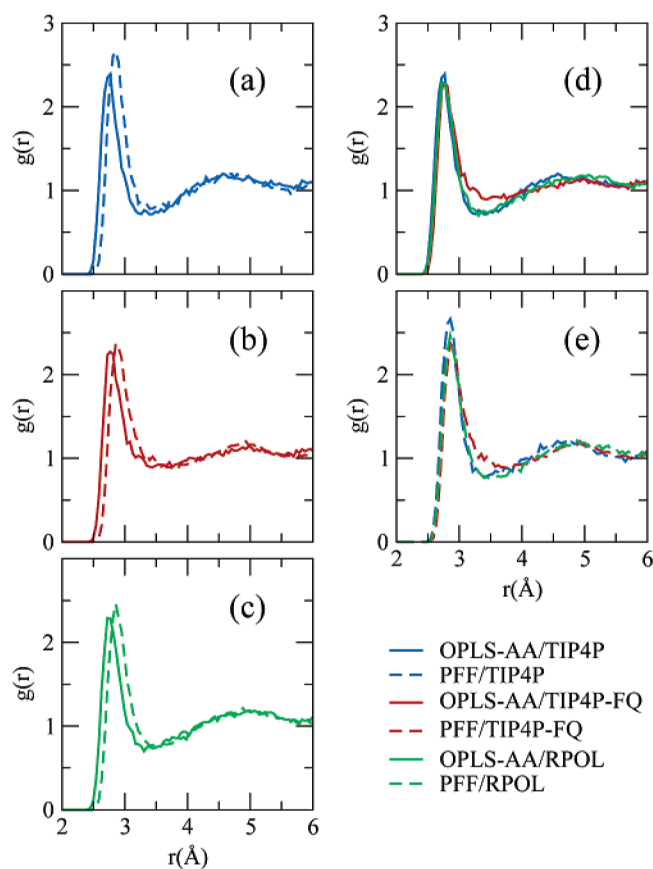
**Figure 2.** Root-mean-square deviation of heavy atoms between the simulation structures and the experimental crystal structure as a function of simulation time. Terminal residues, which show large fluctuations from NMR experiments as well as simulation, are not included in this analysis. The inset shows the RMSD between the 20 NMR structures and the crystal structure.



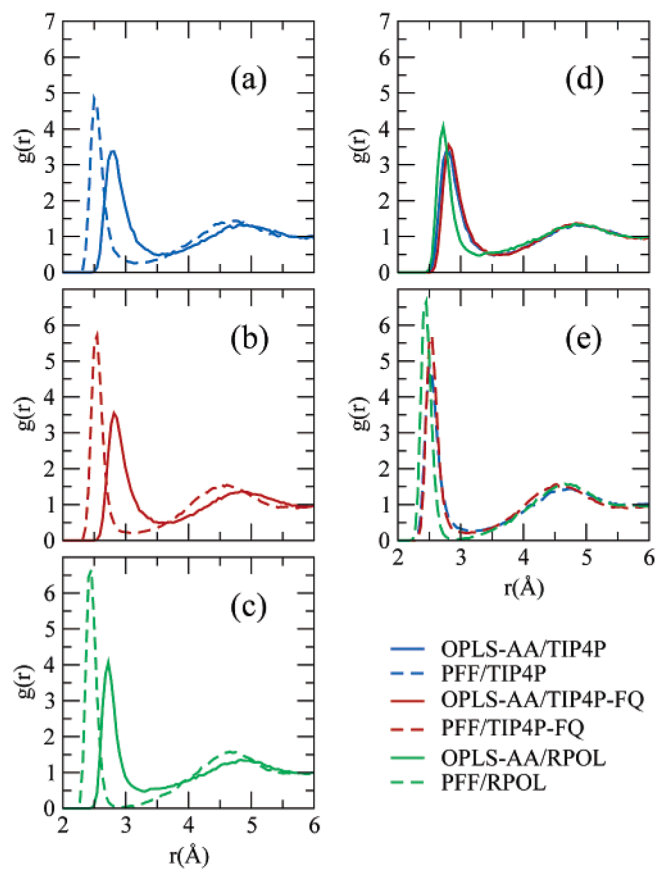


**Figure 3.**

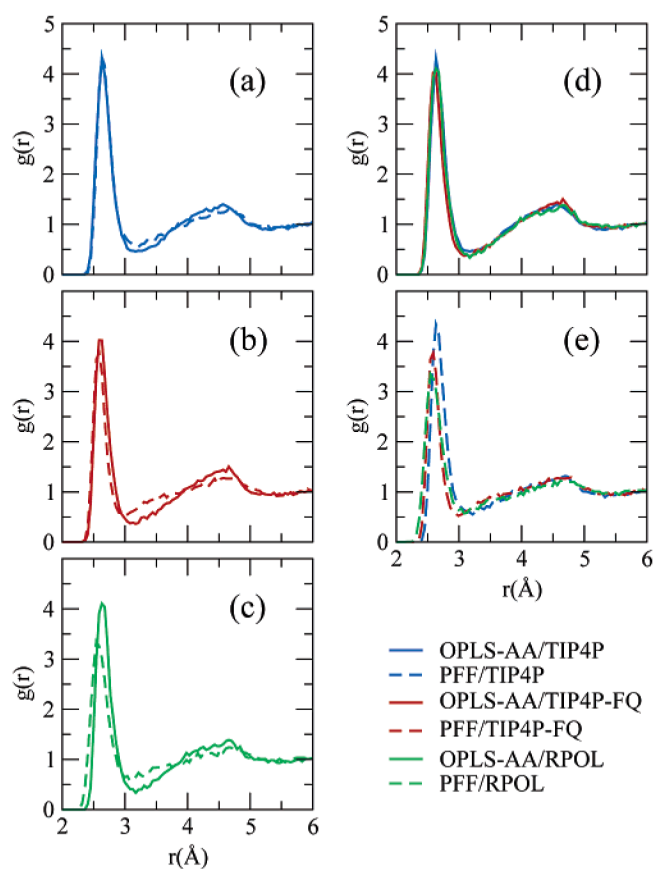
Local radial distribution functions of the water oxygens with respect to the protein carbons in the hydrophobic region. Parts a, b, and c are the curves for systems with the TIP4P (blue), TIP4P-FQ (red), and RPOL (green) models of water, respectively. The solid and dashed lines correspond to the OPLS-AA and PFF models of the protein, respectively. Parts d and e show the data for systems with the OPLS-AA and PFF models of the protein, respectively.



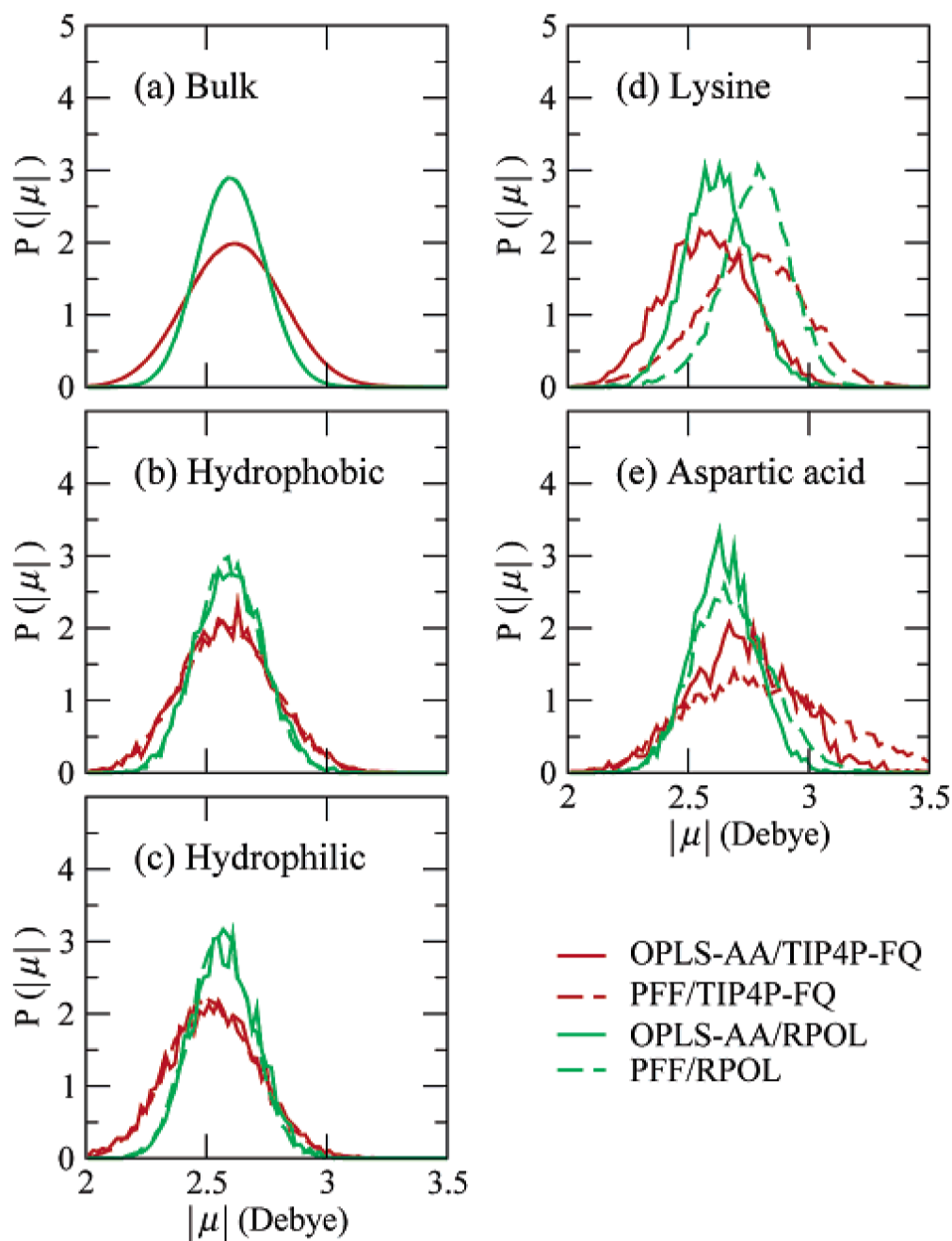
**Figure 4.** Local radial distribution functions of the water oxygens with respect to the protein carbonyl oxygens in the hydrophilic region. Parts a, b, and c are the curves for systems with the TIP4P (blue), TIP4P-FQ (red), and RPOL (green) models of water, respectively. The solid and dashed lines correspond to the OPLS-AA and PFF models of the protein, respectively. Parts d and e show the data for systems with the OPLS-AA and PFF models of the protein, respectively.



**Figure 5.** Local radial distribution functions of the water oxygens with respect to the protein nitrogens in the positively charged region. Parts a, b, and c are the curves for systems with the TIP4P (blue), TIP4P-FQ (red), and RPOL (green) models of water, respectively. The solid and dashed lines correspond to the OPLS-AA and PFF models of the protein, respectively. Parts d and e show the data for systems with the OPLS-AA and PFF models of the protein, respectively.



**Figure 6.** Local radial distribution functions of the water oxygens with respect to the carboxylic oxygens in the negatively charged region. Parts a, b, and c are the curves for systems with the TIP4P (blue), TIP4P-FQ (red), and RPOL (green) models of water, respectively. The solid and dashed lines correspond to the OPLS-AA and PFF models of the protein, respectively. Parts d and e show the data for systems with the OPLS-AA and PFF models of the protein, respectively.



**Figure 7.** Dipole moment distribution of waters in the solvation shell of the different electrostatic regions of the protein. All four systems with polarizable water are shown.

**TABLE 1**RMSD between the Average NMR Structure and the Average Simulation Structure<sup>a</sup>

model	C <sub>α</sub>	heavy atoms
OPLS-AA/TIP4P	0.70	1.17
PFF/TIP4P-FQ	0.85	1.47
PFF/RPOL	0.91	1.39
OPLS-AA/RPOL	0.67	1.16
PFF/TIP4P	0.89	1.47
OPLS-AA/TIP4P-FQ	0.66	1.23

<sup>a</sup>The results include only C<sub>α</sub> atoms and all heavy atoms of the protein. Terminal residues are excluded from the comparison

Mean Dipole Moments of the Solvating Water in the Differing Electrostatic Regions of the Protein<sup>a</sup>

TABLE 2

model	bulk	hydrophobic	hydrophilic	lysine	aspartic acid
OPLS-AA/TIP4P	2.18	2.18	2.18	2.18	2.18
PFF/TIP4P	2.18	2.18	2.18	2.18	2.18
OPLS-AA/TIP4P-FQ	2.62	2.59	2.53	2.58	2.72
PFF/TIP4P-FQ	2.62	2.59	2.53	2.76	2.83
OPLS-AA/RPOL	2.60	2.60	2.57	2.62	2.65
PFF/RPOL	2.61	2.59	2.56	2.78	2.68

<sup>a</sup>The units are Debye.

**TABLE 3**  
Hydrogen-Bond Relaxation Times in the Bulk Region<sup>a</sup>

model	$\tau_{\text{rlx}}$	$\tilde{\tau}_{\text{ret}}$	$\tilde{\tau}_{\text{adj}}$	$\eta_{\text{adj}}$
OPLS-AA/TIP4P	3.10	1.00	1.00	1.60
PFF/TIP4P	3.35	1.00	1.00	1.57
OPLS-AA/TIP4P-FQ	5.01	1.00	1.00	1.51
PFF/TIP4P-FQ	4.97	1.00	1.00	1.51
OPLS-AA/RPOL	4.28	1.00	1.00	1.51
PFF/RPOL	4.46	1.00	1.00	1.49

<sup>a</sup>Units are in unscaled, retarded, and scaled retarded units. The number of adjacent non-hydrogen-bonded waters is also presented.



**TABLE 4**Hydrogen-Bond Relaxation Times in the Bulk Region and from Simulations of Neat Water<sup>30a</sup>

model	$\tau_{\text{rlx}}^{\text{bulk}}$	$2\langle\text{KE}\rangle/Nk$	$\tau_{\text{rlx}}^{\text{neat}}$	T
OPLS-AA/TIP4P	3.10	301.9	3.32	298.15
PFF/TIP4P	3.35	298.5	3.32	298.15
OPLS-AA/TIP4P-FQ	5.10	299.8	5.26	298.15
PFF/TIP4P-FQ	4.97	300.2	5.26	298.15
OPLS-AA/RPOL	4.28	301.5	N/A	N/A
PFF/RPOL	4.46	299.6	N/A	N/A

<sup>a</sup>Units are in picoseconds. The temperature in Kelvin for the neat systems and the average kinetic energy (KE) per degree of freedom ( $N$ ) in the protein simulations is also shown.

**TABLE 5**Hydrogen-Bond Relaxation Times in the Hydrophobic Region of the Protein<sup>a</sup>

model	$\tau_{\text{rlx}}$	$\tilde{\tau}_{\text{ret}}$	$\tilde{\tau}_{\text{adj}}$	$\eta_{\text{adj}}$
OPLS-AA/TIP4P	4.33	1.40	0.96	1.10
PFF/TIP4P	4.69	1.40	1.00	1.12
OPLS-AA/TIP4P-FQ	7.10	1.42	0.95	1.01
PFF/TIP4P-FQ	6.99	1.41	0.96	1.03
OPLS-AA/RPOL	6.37	1.49	1.09	1.11
PFF/RPOL	6.47	1.45	1.08	1.11

<sup>a</sup>The relaxation times are represented in unscaled, retarded, and scaled retarded units. The number of adjacent non-hydrogen-bonded waters is also presented.

**TABLE 6**Hydrogen-Bond Relaxation Times in the Hydrophilic Region of the Protein<sup>a</sup>

model	$\tau_{\text{rlx}}$	$\tilde{\tau}_{\text{ret}}$	$\tilde{\tau}_{\text{adj}}$	$\eta_{\text{adj}}$
OPLS-AA/TIP4P	5.49	1.77	1.25	1.13
PFF/TIP4P	5.61	1.68	1.20	1.13
OPLS-AA/TIP4P-FQ	7.88	1.57	1.17	1.12
PFF/TIP4P-FQ	7.74	1.56	1.16	1.13
OPLS-AA/RPOL	7.51	1.75	1.23	1.06
PFF/RPOL	8.16	1.83	1.25	1.02

<sup>a</sup>The relaxation times are represented in unscaled, retarded, and scaled retarded units. The number of adjacent non-hydrogen-bonded waters is also presented.

**TABLE 7**Hydrogen-Bond Relaxation Times in the Positively Charged Lysine Region of the Protein<sup>a</sup>

model	$\tau_{\text{rlx}}$	$\tilde{\tau}_{\text{ret}}$	$\tilde{\tau}_{\text{adj}}$	$\eta_{\text{adj}}$
OPLS-AA/TIP4P	3.28	1.06	1.32	2.00
PFF/TIP4P	3.59	1.07	1.35	1.98
OPLS-AA/TIP4P-FQ	5.45	1.09	1.29	1.80
PFF/TIP4P-FQ	8.54	1.72	1.71	1.51
OPLS-AA/RPOL	5.06	1.18	1.28	1.64
PFF/RPOL	6.81	1.53	1.52	1.49

<sup>a</sup>The relaxation times are represented in unscaled, retarded, and scaled retarded units. The number of adjacent non-hydrogen-bonded waters is also presented.

**TABLE 8**Hydrogen-Bond Relaxation Times in the Negatively Charged Aspartic Acid Region of the Protein<sup>a</sup>

model	$\tau_{\text{rtx}}$	$\tilde{\tau}_{\text{ret}}$	$\tilde{\tau}_{\text{adj}}$	$\eta_{\text{adj}}$
OPLS-AA/TIP4P	4.64	1.50	1.67	1.79
PFF/TIP4P	4.48	1.34	1.69	1.98
OPLS-AA/TIP4P-FQ	10.60	2.12	1.72	1.23
PFF/TIP4P-FQ	9.45	1.90	2.06	1.64
OPLS-AA/RPOL	7.95	1.86	1.80	1.46
PFF/RPOL	7.74	1.74	2.21	1.90

<sup>a</sup>The relaxation times are represented in unscaled, retarded, and scaled retarded units. The number of adjacent non-hydrogen-bonded waters is also presented.