Hydration Water and Bulk Water in Proteins Have Distinct Properties in Radial Distributions Calculated from 105 Atomic Resolution Crystal Structures

Xianfeng Chen,[†] Irene Weber,[†] and Robert W. Harrison*[,‡]

Department of Biology and Department of Computer Science, Georgia State University, Atlanta, Georgia 30302-4010

Received: April 1, 2008; Revised Manuscript Received: June 13, 2008

Water plays a critical role in the structure and function of proteins, although the experimental properties of water around protein structures are not well understood. The water can be classified by the separation from the protein surface into bulk water and hydration water. Hydration water interacts closely with the protein and contributes to protein folding, stability, and dynamics, as well as interacting with the bulk water. Water potential functions are often parametrized to fit bulk water properties because of the limited experimental data for hydration water. Therefore, the structural and energetic properties of the hydration water were assessed for 105 atomic resolution ($\leq$1.0 Å) protein crystal structures with a high level of hydration water by calculating the experimental water−protein radial distribution function or surface distribution function (SDF) and water radial distribution function (RDF). Two maxima are observed in SDF: the first maximum at a radius of 2.75 Å reflects first shell and hydrogen bond interactions between protein and water, and the second maximum at 3.65 Å reflects second shell and van der Waals interactions between water and nonpolar atoms of protein-forming clathrate-hydrate-like structures. Thus, the two shells do not overlap. The RDF showed the features of liquid water rather than solid ice. The first and second maxima of RDF at 2.75 and 4.5 Å, respectively, are the same as for bulk water, but the peaks are sharper, indicating hydration water is more stable than bulk water. Both distribution functions are inversely correlated with the distribution of B factors (atomic thermal factors) for the waters, suggesting that the maxima reflect stable positions. Therefore, the average water structure near the protein surface has experimentally observable differences from bulk water. This analysis will help improve the accuracy for models of water on the protein surface by providing rigorous data for the effects of the apparent chemical potential of the water near a protein surface.

Introduction

Water is essential to the structure and function of proteins. Generally, water around proteins can be divided into three categories:[1] (1) bulk water that surrounds the protein molecule at a separation of greater than van der Waals contact, (2) individually bound water that forms hydrogen bonds with charged or polar protein atoms in cavities inside the protein,[2] and (3) hydration water at the protein surface with direct interactions with the protein.[3,4] The three categories have different functions. Bulk water is free to move, and it assists in protein diffusion relative to the other interacting molecules by random movement in solution. Hydration water forms water networks around the protein surface to keep protein in solution. Individually bound water has multiple contacts that stabilize the protein structure. For example, individually bound water stabilizes cavities in thrombin and trypsin[5,6] and mediates contacts between proteins and their ligands.[7] The presence of individually bound water depends on the local structure of the protein and is often conserved among related proteins. However, a water molecule can share the properties of both individually bound water and hydration water if it has multiple contacts with protein and contributes to a hydration water network. Both individually bound water and hydration water are also known as protein-bound water and form strong interactions with proteins. Bound water is visible in protein crystal structures, but bulk water is not observed because there is no stable location for the water molecules.

Hydration water is a major form of protein-bound water that contributes to many properties of proteins such as protein folding, solubility, drug docking, and oligomer formation as reviewed in ref 8 Hydration water directly interacts with the protein surface and affects the local protein structure. Protein enzymes without a hydration layer will lose their catalytic abilities.[9] Direct observation of hydration water in proteins is difficult, except for visualization in atomic resolution ($\sim$1.0 Å) crystal structures of proteins where analysis of neutron and X-ray crystal structures of three proteins showed differences from bulk water.[10] Elegant calorimetric experiments on hydrated collagen show that there is a specific binding energy or chemical potential associated with hydration.[11] Terahertz spectroscopy has shown that the effects of the hydration environment can extend to 10 Å from the surface of a protein.[12] Neutron scattering of dipeptide solutions showed small differences between bulk and hydration water,[13] but the defects may be different with proteins because of their larger size. X-ray scattering has been used with myglobin to study hydration water.[14]

Although hydration water is very important for protein structure and function, there are no specific parameters for the potential of hydration water in simulations such as molecular dynamics simulations (MD). The water potential function is defined by a set of parameters in simulations that represent the potential energy of water. The water potentials currently used

* Corresponding author. Phone: 404-413-5724; fax: 404-413-5717; e-mail: rharrison@cs.gsu.edu.
[†] Department of Biology.
[‡] Department of Computer Science.

10.1021/jp802795a CCC: $40.75
Published on Web 08/28/2008

in MD programs are based on water models such as TIP3P,[15] TIP4P,[16] TIP5P, [17] and SPC[18] that are parametrized to reproduce the experimental data of bulk water. Recently, neutron scattering data on dipeptides were used as a basis for generating water potentials.[13] The potentials derived from bulk water may not accurately represent the properties of hydration water. For example, when TIP3P was used with CSFF[19] and GLYCAM-2000a,[20] the new versions of CHARMM[21] and AMBER[22] force fields, respectively, the modeled carbohydrate−water interactions on the surface of disaccharides were less structured than in experimental data.[23] Theoretical calculations using TIP potentials have been decomposed into polar, apolar, and charged interactions[24] in order to fit compressibility data. Moreover, although the compressibility data were well reproduced in ref 24 the radial distribution curves show less detail than is seen experimentally. Since incomplete sampling of phase space may be an issue in theoretical calculations, solution thermodynamics of proteins and amino acids have been analyzed by Fourier integral methods in statistical mechanics, as implemented in 3D-RISM [25−27] and DYNAMA.[28] These methods may resolve sampling issues to produce a better structural agreement between theory and experiment for the same potentials. Hence, the differences between the properties of hydration water and bulk water need to be better defined in order to improve the water potentials and especially to reproduce the critical interactions between water and proteins or other soluble macromolecules.

The structural and energetic factors of hydration water can be studied using the water−protein radial distribution function or surface distribution function (SDF) and water radial distribution function (RDF). The SDF describes the density of water as a function of the distance from the protein surface (the closest non-hydrogen atom). The RDF describes the density of water as a function of the distance from a particular water molecule. The RDF is determined by averaging over the many different local structures formed in liquid water.[29−31] Local structures include neutral (uncharged) water structures, and two-water−molecule structures with one proton attached (H_5+O_2).[30] Water molecules also form higher order clusters, such as 10-molecule tetrahedral clusters[29] and 280-molecule expanded icosahedral water clusters comprising 14-molecule tetrahedral units.[31]

Previous studies on the hydration water were highly dependent on computational simulations because of the difficulty of direct measurements. Direct measurements of visible hydration water became possible when the atomic resolution crystal structures of proteins became available in significant numbers. Moreover, an average over a number of crystal structures will reduce the deviation caused by the differences in crystallization conditions and other sources of experimental variation.

This study examined 105 crystal structures from the Protein Data Bank[32] solved with X-ray data at resolutions better than 1.0 Å. The SDFs and RDFs were calculated and averaged. Two density maxima representing two shells are observed in SDF, the water molecules in the first shell are much denser than those in the second shell. The maxima of RDF are higher than those of bulk water and the minima are lower. In both SDF and RDF, the higher density of water corresponds to the lower B-factor value of water, confirming that water molecules are stable in each maximum.

Methods and Algorithms

The Water−Protein Distance. The ideal model of a protein is a sphere surrounded by first and second water shells (Figure 1a). The positions and densities of the two shells are determined

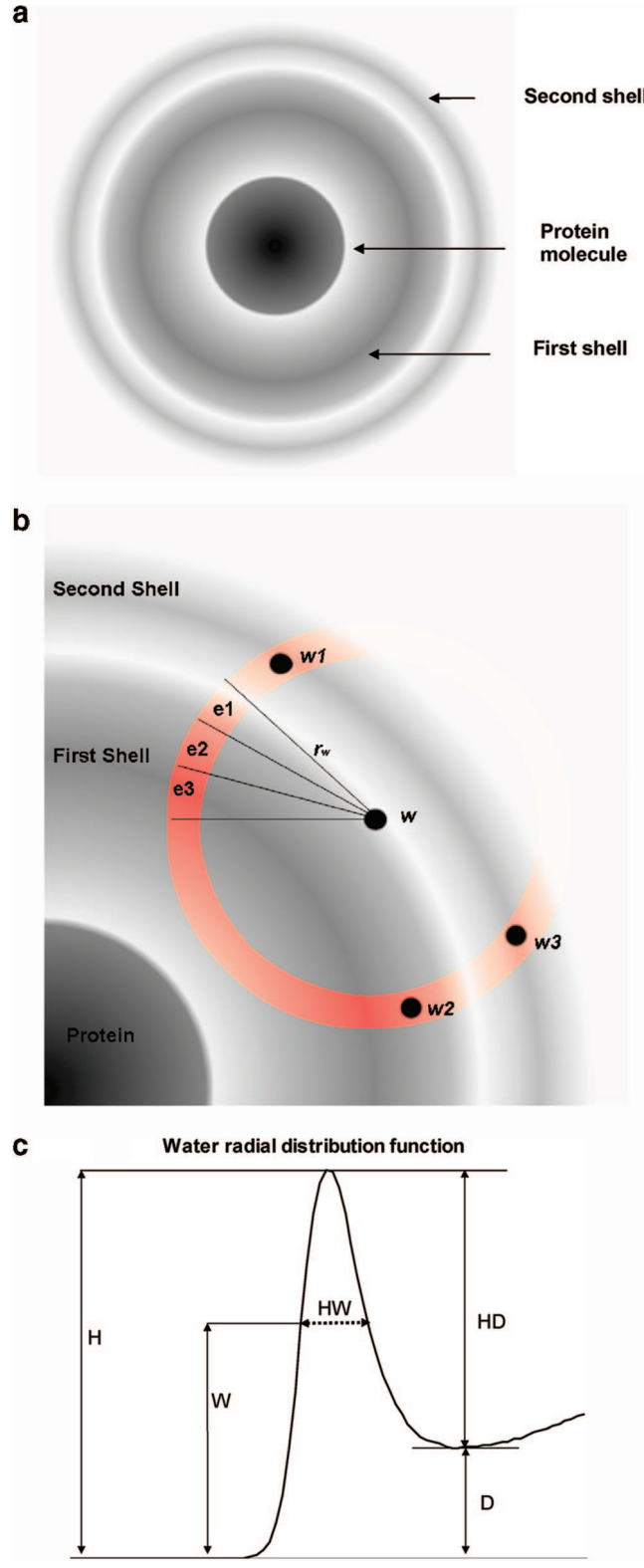

Figure 1. (a) The ideal model of water distributed around a protein molecule. (b) Representation of terms. The RDF is the density of water as a function of distance from a particular water molecule w. The actual density of water in the orange shell (UDV) is the number of waters (w1, w2, and w3 represent three of the water molecules) divided by the volume of the UDV. The small cells (e1, e2, e3, etc.) are expected to have different water density because they have different distances from the protein surface. The observed density of water in the UDV is the average of the expected density of each small cell. (c) The first maximum of an RDF. The sharpness of a maximum of an RDF is defined as HD/HW, where HD = H − D, HW (dashed line) is the width of the peak at height W, and W = HD/2 + D.

by the interactions between water and the protein. In practice, the protein−water distance is defined as the distance between the water oxygen and the closest non-hydrogen atom in the protein molecule.

The Water−Protein Radial Distribution. In protein crystal structures, the average density of water around protein (ρ) within any water−protein distance h is

$$\rho = \frac{n}{\frac{4}{3}\pi(R_p + h)^3 - \frac{4}{3}\pi R_p^3}$$

where n is the number of water molecules with water−protein distance less than h, R_p is the radius of the protein molecule p.

The bulk water is invisible in protein crystal structures because it is not bound to protein. The density of "visible" water (mostly hydration water) around the protein decreases sharply with distance from the protein surface. The SDF is the water density as a function of the water−protein distance (h) and termed g_{wp}

$$g_{wp}(h) = \frac{d\rho}{dh} \quad (1)$$

Water Radial Distribution. The RDF in the protein crystal structure must be adjusted because of the presence of protein molecules, unlike the RDF in bulk water where only water−water interactions are considered. In the presence of protein atoms, the volume over which the water can be observed is restricted and thus the manner in which the density is normalized must be modified. The raw counts of water molecules within each radial shell are converted into an expected value for the RDF based on the restriction of the shell because of interactions between water and protein molecules. This process is shown in Figure 1b where the thin orange shell represents a volume shell at a distance r_w from the water w. The thin orange shell illustrates the unimpeded differential volume element (UDV). The water density at any point within the area surrounded by the orange shell is not only related to the distance to water w but also related to the distance to protein surface. Steric interactions between the protein and water prevent or impede water from binding in all of the positions in the UDV; thus, the area surrounded by the orange shell defines the available differential volume element (ADV). Thus, two water densities can be calculated within ADV: the water density (ρ_{ww}) which is related to the distance from the water w and caused by water−water the interaction and the water density (ρ_{wp}) which is related to the distance from the protein surface and caused by the water−protein interaction.

$$\rho_{ww} = \frac{n_{ww}}{\upsilon_w}$$

$$\rho_{wp} = \frac{n_{wp}}{\upsilon_w} = \frac{\int_{\varphi=0}^{2\pi}\int_{\theta=0}^{\pi}\int_{r=0}^{r_w} r^2 g_{wp}(h_{wr\varphi\theta})\sin\varphi\, d\varphi d\theta dr}{\upsilon_w}$$

where n_{ww} is the actual total number of water molecules within ADV, υ_w is the volume of ADV, r_w is the distance from UDV to water w, n_{wp} is the number of water molecules within ADV evaluated by SDF (g_{wp}), $h_{wr\varphi\theta}$ is the distance from each point within ADV to the protein surface.

Thus, the raw RDF (g_{raw}) and the water density distribution for normalization (g_{nor}) can be calculated as

$$g_{raw}(r_w) = \frac{d(\rho_{ww})}{dr_w}$$

$$g_{nor}(r_w) = \frac{d(\rho_{wp})}{dr_w}$$

where $d(\rho_{ww})$ is the water density in the UDV calculated from the actual number of waters (w1, w2, and w3 indicate the three water molecules in Figure 1b) divided by the volume of the UDV. $d(\rho_{wp})$ is the average of the water density of each small cell in the UDV (e1, e2, e3, etc.), which is determined by its water−protein distance ($h_{wr\varphi\theta}$). dr_w is the thickness of the UDV.

Thus, the averaged raw RDF ($\langle g_{raw}\rangle$) and the averaged water density for normalization ($\langle g_{nor}\rangle$) over all sampled crystal structures can be expressed as

$$\langle g_{raw}(r_{wp})\rangle = \frac{1}{P}\sum_{p=1}^{P}\left(\frac{1}{W_p}\sum_{w=1}^{W_p} g_{raw}(r_{wp})\right) \quad (2)$$

$$\langle g_{nor}(r_{wp})\rangle = \frac{1}{P}\sum_{p=1}^{P}\left(\frac{1}{W_p}\sum_{w=1}^{W_p} g_{nor}(r_{wp})\right) \quad (3)$$

where P is the total number of protein structures in our sample, W_p is the total number of water molecules in protein p, r_{wp} is the distance from the UDV to water w in protein p.

Therefore, in crystal structures the normalized RDF (g_{nmlzed}) can be expressed as

$$g_{nmlzed}(r_{wp}) = \frac{\langle g_{raw}(r_{wp})\rangle}{\langle g_{nor}(r_{wp})\rangle} \quad (4)$$

The Number of Water Molecule. Every water molecule has a refined occupancy in the PDB (Protein Data Bank) file. The majority of water has an occupancy value of 1.0, indicating the presence of this water in all protein molecules forming the crystal. Some waters have occupancies of less than 1.0, indicating the water is not present in all protein molecules in the crystal. In this study, the sum of the occupancy of each water molecule was used instead of the actual number of waters.

The Sharpness of a RDF. A numerical measure based on the width at half-height was defined in order to describe the sharpness of the features of the RDF. The peaks in the RDF are asymmetric, especially the first peak, which makes the standard symmetric half-height measure difficult to apply. Because of this asymmetry, the measure was defined as the amplitude between a maximum and its next minimum (HD) divided by the width of the maximum at the half-value of HD (Figure 1c). This corresponds to the width at the smaller of the two peak heights.

The Water−protein Radial Distribution of B-Factors. The surface (or water−protein radial) distribution function of B-factors (SDFB) is the averaged B-factor of water as a function of water−protein distance (h) which can be calculated from

$$B(h) = \frac{d}{dh}\left(\frac{1}{W}\sum_{w=1}^{W}\frac{b_w}{o_w}\right)$$

where W is the total number of water molecules in the sphere with distance ($<h$) from protein p, b_w is B-factor of each water oxygen in the sphere, and o_w is the occupancy of each water molecule in the sphere. Thus, the averaged SDFB over all sampled crystal structures can be expressed as

$$\langle B(h_p)\rangle = \frac{1}{P}\sum_{p=1}^{P}\left(\frac{1}{W_p}\sum_{w=1}^{W_p} B(h_p)\right) \quad (5)$$

where P is the total number of protein structures in our samples, W_p is the total number of water molecules in protein p, h_p is

the distance from the surface of the sphere to the closest atom in protein p.

The Water Radial Distribution of B-Factor. The water radial distribution of the B-factor is the averaged B-factor of water as a function of water−water distance. At any distance from water w (r_w), the radial distribution function of the water B-factor (RDFB) can be calculated as

$$B(r_w) = \frac{d}{dr_w}\left(\frac{1}{W}\sum_{w=1}^{W}\frac{b_w}{o_w}\right)$$

where W is the number of water and partial water molecules in the ADV with radial ($<r_w$) from water w (Figure 1b), b_w is the B-factor of each water in the ADV, and o_w is the occupancy of each water in the ADV. Thus, the averaged RDFB over all sample crystal structures can be expressed as

$$\langle B(r_{wp})\rangle = \frac{1}{P}\sum_{p=1}^{P}\left(\frac{1}{W_p}\sum_{w=1}^{W_p}B(r_{wp})\right) \quad (6)$$

where P is the total number of protein structures, W_p is the total number of water molecules in protein p, and r_{wp} is the distance from the UDV to water w in protein p.

The above calculations are performed with PDBAnalyzer, a JAVA program developed in-house, and using Microsoft SQL Server 2000 as the database engine. The data for each water molecule such as occupancy, B factor, PDB file name, atom number, distance to the closest atom in protein, and the polarity of this atom in protein were input to a database table as one record. Then, SQL scripts were used to extract data for SDF from more than ten thousand records. Similarly, the data for each pair of water molecules with separation less than 10.5 Å were input into a database table as one record, along with both occupancies, B factors, PDB file names, atom numbers, and distances to protein surface. The SQL scripts were used to extract data for RDF from over one million records. Hydrogen atoms can rarely be seen because of the presence of a single electron, even in subatomic resolution crystal structures of proteins. Therefore, the position of the water molecule was defined by the position of the oxygen atom, and the RDF is the O−O distribution. The water−protein distance is also calculated as the distance from the O to the non-hydrogen atom of the protein.

Results

One hundred and five protein crystal structures refined with water molecules and diffraction data at resolutions ≤1.0 Å were analyzed (the structures are listed in the Appendix). These structures contain a total of 33376 water molecules, including 9026 partially occupied waters. The variable crystallization conditions such as temperature, ion concentration, and pH value, and differences in crystallographic methodology, are assumed to have little or no effect on the calculated radial distributions.

Water−Protein Radial Distribution. The calculated SDF showed two maxima, representing two water shells around proteins, at radial distances of 2.75 Å and 3.65 Å (Figure 2a). The first maximum of 0.19 atoms/Å^3 is twice the average density of bulk water at 267 K temperature.[33] The second maximum of 0.024 atoms/Å^3 is much lower. The SDF (blue) is the sum of the water−polar atom radial distribution (SDFpol, red) and the water−nonpolar atom radial distribution (SDFnon, green), since the protein surface consists of polar atoms (or charged atoms, mostly oxygen and nitrogen) and nonpolar atoms (mostly carbon). SDF overlaps with SDFpol almost perfectly between 2.25−3.25 Å, indicating that the first shell consists of water

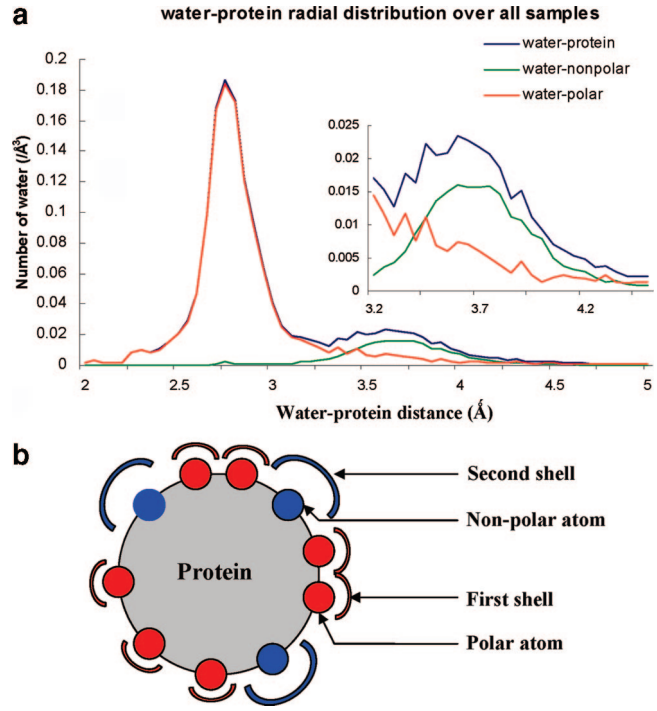

Figure 2. (a) The SDF (blue), water−polar atom radial distribution function (SDFpol, red), and water−nonpolar atom radial distribution function (SDFnon, green). The number of waters is averaged over 105 crystal structures. The plots are calculated from eq 1. The inset shows the enlarged second shell. (b) The model of the first and second water shells on the protein surface. Red and blue beads indicate the polar (or charged) and nonpolar atoms on protein surface, respectively. The red and blue arcs represent the first and second shell, respectively. The two shells do not overlap.

molecules with hydrogen bonds to the polar (or charged) atoms of the protein. The second maximum between 3.2−4.2 Å has a larger contribution from SDFnon, indicating the importance of van der Waals interactions between water molecules and protein. Thus, the first shell is formed around polar atoms and the second shell is formed around the nonpolar atoms of the protein, and the two shells are distinct (Figure 2b). The first shell is much denser than the second shell, because polar and charged atoms predominate on the solvent-accessible surface of proteins.

The water−water interactions for each water molecule were calculated from all samples. More than 98% of water molecules in the second shell, compared to 82% of all waters, have hydrogen bond interactions with at least one additional water molecule (with average 1.7), indicating further stabilization of these waters by formation of networks. According to Petrenko and Whitworth 1999,[34] water molecules interact to form small cage-like structures called clathrate hydrates around methane, ethane, $(CH_2)_2O$, and other hydrophobic molecules. Similar cages around methyl groups are observed in a protein crystal structure.[35] Thus, the majority of water molecules in the second shell of SDF form clathrate-hydrate-like structures.

The Water Radial Distribution Function. Equation 2 used to calculate the actual RDF was developed for calculation of RDF in a system of bulk water where the expected density of water is constant. The observed density of water as a function of water−water distance in proteins is not constant (Figure 3a, dashed pink) which causes the uncorrected or observed RDF to decrease rather than remaining constant with increasing distance from the protein surface (Figure 3a, blue). To eliminate this bias, the normalized RDF (g_{nmlzed}) (Figure 3b, red) is used as calculated from eq 4. The normalized RDF is compared to

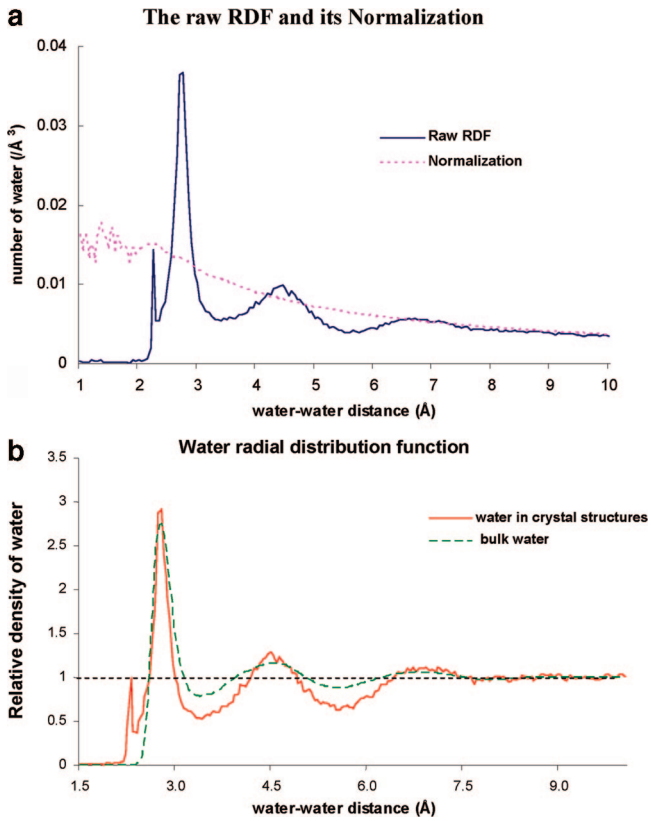

Figure 3. (a) The raw RDF (blue) from eq 2 and the water density distribution for normalization (pink) from eq 3 among 105 crystal structures. 3b. The normalized RDF (red) of water in crystal structures and the RDF of pure bulk water at 298K (green dashed line).[37]

the RDF of bulk water from ALS X-ray scattering and neutron diffraction at 298 K [36,37] (Figure 3b). The normalized RDF in crystal structures fluctuates more sharply than in bulk water, although the positions of the maxima and minima are the same. The amplitude of each maximum or minimum in the normalized RDF is larger than in the RDF of pure bulk water (Figure 3b) and the peaks are sharper. This suggests that steric hindrance and specific interactions between the protein surface and the water lead to a more ordered water structure in the vicinity of the protein surface. Unlike bulk water where three peaks are seen in the RDF,[37] these surface effects extend to a (weak) fourth peak at about 8.5 Å distance (Figure 3b, Table 1) and evidence is seen of ordering at 10 Å.

The Effect of Individually Bound Water on the Normalized RDF. Individually bound water is a small component of the water observed in protein structures. Some individually bound water molecules are deeply buried inside the protein with no surface accessibility. For example, there is a water molecule deeply embedded in inverting glycosidase (water 513 in PDB entry 1KWF), which is 9.3 Å away from the closest water. Such water molecules can bias the analysis of hydration water because they do not belong to the hydration water network. Methodologically, individually bound water was defined as water with no other water molecules within 3.2 Å (the maximum distance to form a hydrogen bond) in this study. Using this definition, individually bound water formed 18% of our sample. The normalized RDF of "pure" hydration water, after removing the individually bound water molecules, also is sharper than that for bulk water and even more than the normalized RDF (Table 1). Similar results were obtained by changing the water−water distance that defines individually bound water to 3.0 Å or 3.4 Å. However, a discontinuity in the RDF is introduced at the cutoff which is not physically realistic. Therefore, the RDF of "pure" hydration water in the experimental data is sharper than bulk water indicating hydration water is less mobile and more ordered than bulk water.

The Relationship between Water Density and Its B-Factor. Atoms in molecular structures move because of thermal vibrations. In crystal structures, the atomic B factor or thermal factor is used to describe this motion, as well as the effects of static disorder.[28] The distribution of atomic B factors can be used to study protein flexibility,[38] thermal stability [39,40] and protein dynamics.[41] The relationship between water density and water thermal factor could shed light on the dynamics of hydration water.

The radial distribution of averaged B factor of water molecules as a function of the water−protein distance (SDFB) is shown in Figure 4a. The maxima and minima are the inverse of those in the SDF (Figure 4b). Similarly, the radial distribution of averaged B factor of water molecules as a function of the water−water distance (RDFB) show inversed positions of maxima and minima relative to the RDF. In conclusion, water density and its B-factor are negatively correlated, confirming that the water molecules in the shells of high density have relatively low mobility.

Discussion

What Is the Small Maximum in the Normalized RDF. The small maximum in the vicinity of 2.3 Å (Figure 3a,b.) has not been reported previously. One possibility is just sodium misassigned as water, because sodium is often abundant in the crystallization solutions and can easily resemble a "water" molecule during the refinement of the crystal structure. Harding[42] analyzed the radial distribution of sodium-water in protein crystal structures and found the peak to be between 2.3−2.5 Å. However, the small maximum is very sharp within the narrow range of 2.28−2.3 Å and distinctly different from the maximum because of the assigned sodium−water distribution. The water molecules with separation of 2.28−2.3 Å were checked manually, and none resembled sodium ions with short interactions with several water molecules or polar atoms of protein as described previously.[42−44] Therefore, the small maximum is unlikely to be caused by sodium mistakenly assigned as water.

It is possible the small maximum near 2.3 Å arises from a particular water structure, although known water structures such as ice (Ih)[37] produce a maxima at 2.5 Å and H_5+O_2[30] has a O···O distance of 2.52 Å. Alternatively, 90% of the crystal structures in this sample were refined with SHELX (http://shelx.uni-ac.gwdg.de/SHELX/). SHELX picks the maximum of electron density with a distance cutoff of 2.3 Å by default. In addition, most of the waters in the small maximum were observed to have poorly defined electron density and high B factors. Therefore, the small maximum is possibly caused by the distance cutoff of SHELX, or overinterpretation of the electron density maps. However, the possibility that the small maximum represents protonated or another specific water structure cannot be excluded.

How Closely Does the Hydration Water in Protein Crystals Represent Water in Solvated Proteins. Water and protein structures are affected by many factors such as temperature and ion concentration. Protein crystals are grown from solutions with various buffers, salts, and/or organic precipitating agents and additives. Many protein crystals are grown at room temperature and atomic resolution X-ray data recorded at very low temperature (liquid nitrogen). However, these factors were

TABLE 1: Comparison of Maxima and Minima in RDF Calculated for Protein Structures and Pure Bulk Water[a]

	proteins[b]			hydration water in proteins[c]			pure bulk water[d]		
	position	height	sharpness	position	height	sharpness	position	height	sharpness
first maximum	2.76	2.92	10.08	2.76	4.04	16.95	2.73	2.75	6.35
first minimum	3.39	0.50		3.33	0.48		3.33	0.78	
second maximum	4.47	1.29	0.84	4.47	1.48	0.99	4.53	1.16	0.27
second minimum	5.52	0.68		5.52	0.68		5.58	0.88	
third maximum	6.90	1.15	N/A	6.81	1.15	N/A	6.87	1.06	N/A
forth maximum	8.85	1.05	N/A	8.82	1.05	N/A	N/A	1.0	N/A

[a] The position (Å), height (relative density), and the sharpness of each maximum or minimum are noted. [b] The normalized RDF from figure 3b (red line). [c] The normalized RDF of "pure" hydration water. [d] The RDF of pure bulk water at 298 K[37] from Figure 3b (dashed green line).

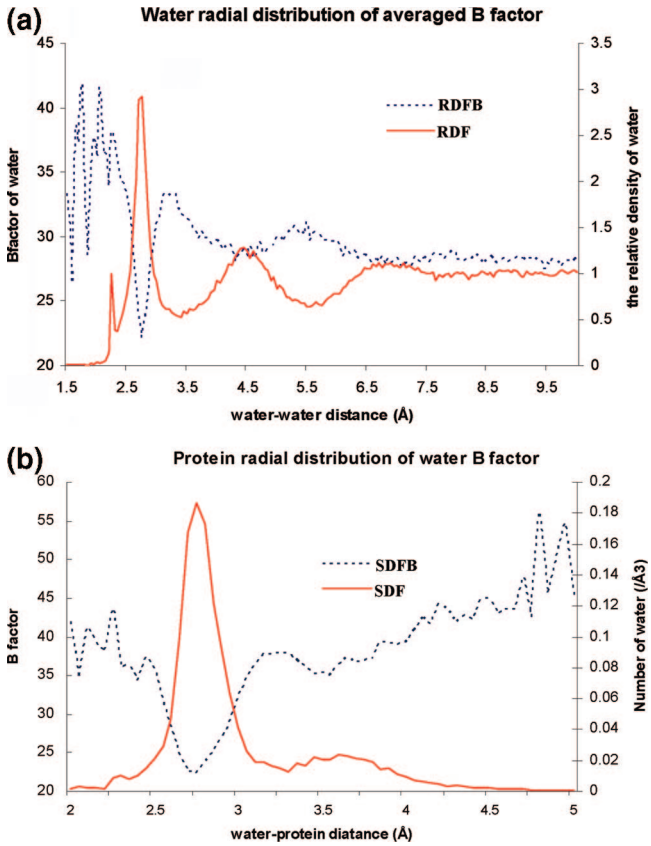

Figure 4. (a) The normalized RDF (red line) and RDFB (dotted blue line) calculated from eq 5. (b). The SDF (red line) and SDFB (dotted blue line) calculated from eq 6.

ignored in the calculations of the normalized RDF in protein crystal structures, which raises several questions as discussed below.

Crystals are often grown in high ionic strength solutions which could affect the RDF. The high ion concentration in the hydration water of protein crystals, based on previous studies,[45,46] does not increase the sharpness of RDF maxima. The RDF of 10 M lithium chloride solution has no change relative to the RDF of pure water.[46] However, with an 8 M NaCl solution the sharpness of the RDF maxima decreased about 30% from that of pure water.[45] Salts and buffers tend to reduce the sharpness if they have any affect. The RDF in this study shows an increase in sharpness for RDF maxima (Figure 3b, red), which strongly suggests that the effects of the salts are not causing the difference between hydration water and bulk water.

One possible explanation for the increase in the order in the RDF is the formation of ice. The ice specific maxima (from RDF of ice (Ih)[37]) are not observed in the normalized RDF of water in protein crystals, indicating that the hydration water on the protein surface is still liquid-like even at very low temperature. Even though the normalized RDF in protein crystal structures is sharper than in bulk liquid water at 298 K or at 275 K,[47] it is unclear how much the low data temperature of data collection affects the normalized RDF. Experimental studies[48] have shown that the disorder in protein crystal structures is dominated by temperature-independent static disorder rather than disorder because of thermal motion. Therefore, the effects of temperature on the sharpness of the RDF for protein crystal structures are difficult to extract from the data and may differ from those of an isolated system. It would be necessary to obtain atomic resolution protein structures with data collected at room temperature to answer this question, which is very difficult for most protein crystals using current technology.

Because of the difficulties of experimental measurement, the current studies on hydration water on the protein surface highly depend on computational simulations. Many simulations use water models such as TIP3P,[17] TIP4P, TIP5P,[17] and SPC[18] derived from bulk water. This study, with experimental atomic resolution data from 105 protein crystal structures, showed a significant difference in the water structure and water potentials between hydration water and bulk water in protein crystals. The results agree with previous studies on crystals of three proteins.[10] The RDF of liquid water in vycor, a kind of porous glass with a hydrophilic surface similar to proteins, has higher first and second maximum and lower first minimum than that of bulk water at 300 K.[49] The RDF of water around the myoglobin molecule showed that both the depth of the first minimum and the height of the second maximum are larger than those for pure bulk water at 298 K.[14] Neutron scattering on dipeptide solutions shows smaller but distinct effects (McLain, 2008),[13] as expected due to the smaller size of the solute. The SDFpol and SDFnon in this study are similar to those seen in computer simulations of trypsin, ribonuclease A, hen egg white lysozyme, and α-lactalbumin.[24]

Conclusions

There are two layers of hydration water in protein crystals as demonstrated by the two maxima observed in SDF calculated for 105 atomic resolution structures. This study expands on the analysis of three protein structures by neutron and X-ray diffraction[10] by using a much larger and more complete sample that allows for better normalization. Similar to the earlier study[10] the first significant maximum is sharper than is seen in bulk water, but our analysis showed ordered structure at a larger distance and a correlation between RDF and thermal factors. The first maximum at 2.75 Å represents the center of the inner layer of hydration water formed by hydrogen bonds between water and polar atoms of the protein. The second maximum at 3.65 Å represents the outer layer of hydration water formed by

water−water hydrogen bonds and van der Waals interactions between water and nonpolar atoms of the protein. The structure of water in the outer layer resembles that of a clathrate-hydrate. The RDF displays the characteristics of liquid water rather than solid ice. The first and second maxima of the RDF are at 2.75 and 4.5 Å, respectively, the same distances as those of bulk water. However, the maxima/minima of the RDF of hydration water are higher/lower than those of bulk water, indicating that hydration water is denser and narrowly defined, likely due to stronger interactions induced by the protein. The observed experimental differences between hydration water and bulk water will serve as a solid foundation for theoretical calculations to analyze the apparent chemical potential (or zeta-potential) of water near protein surfaces.

Acknowledgment. We thank Dr. Long Wang for discussing the equations and Dr. Alan K. Soper for data on bulk water. X.C. was supported in part by the Georgia State University Research Program Enhancement award. I.T.W. and R.H.W. are Georgia Cancer Coalition Distinguished Cancer Scholars. This research was supported in part by the Molecular Basis of Disease Program of Georgia State University, the Georgia Research Alliance, the Georgia Cancer Coalition, and the National Institutes of Health awards GM065762 and GM062920.

Appendix

Structural data for 105 protein crystal structures have been deposited in the Protein Data Bank: 1a6m, 1aho, 1b0y, 1brf, 1bxo, 1c75, 1ceq, 1ea7, 1eb6, 1etl, 1etm, 1etn, 1f94, 1f9y, 1g66, 1gkm, 1gwe, 1hje, 1i1w, 1ic6, 1iee, 1ir0, 1iua, 1k2a, 1k4p, 1k6u, 1kwf, 1m40, 1mc2, 1mnz, 1mxt, 1n1p, 1n4u, 1n4v, 1n4w, 1n55, 1n9b, 1ob4, 1ob7, 1ok0, 1ot6, 1ot9, 1p9g, 1pq5, 1pwm, 1rtq, 1s5n, 1ssx, 1sy3, 1tt8, 1ucs, 1unq, 1us0, 1v0l, 1vb0, 1vyr, 1w0n, 1 × 8p, 2fdn, 4lzt, 7a3h, 8rxn, 1byi, 1c7k, 1cex, 1exr, 1g4i, 1g6x, 1gci, 1gqv, 1hj8, 1hj9, 1iqz, 1j0p, 1jfb, 1k4i, 1k5c, 1kcc, 1kth, 1l9l, 1lug, 1m1q, 1mj5, 1muw, 1nls, 1nwz, 1od3, 1pjx, 1pq7, 1q6z, 1r6j, 1rb9, 1s5m, 1sy2, 1tqg, 1ufy, 1ug6, 1vl9, 1 × 8q, 2bf9, 2erl, 2pvb, 3lzt, 3pyp, 8a3h.

References and Notes

(1) Purkiss, A. The protein-solvent interface: a big splash. *Philos. Trans. R. Soc. London, Ser. A:* **2001**, *359* (1785), 1515–1527.

(2) Roberts, B. C.; Mancera, R. L. Ligand-Protein Docking with Water Molecules. *J. Chem. Inf. Model.* **2008**, *48* (2), 397–408.

(3) Nakasako, M. Large-scale networks of hydration water molecules around bovine beta-trypsin revealed by cryogenic X-ray crystal structure analysis. *J. Mol. Biol.* **1999**, *289* (3), 547–564.

(4) Higo, J.; Nakasako, M. Hydration structure of human lysozyme investigated by molecular dynamics simulation and cryogenic X-ray crystal structure analyses: on the correlation between crystal water sites, solvent density, and solvent dipole. *J. Comput. Chem.* **2002**, *23* (14), 1323–1336.

(5) Rashin, A. A.; Iofin, M.; Honig, B. Internal cavities and buried waters in globular proteins. *Biochemistry* **1986**, *25* (12), 3619–25.

(6) Meyer, E. Internal water molecules and H-bonding in biological macromolecules: a review of structural features with functional implications. *Protein Sci.* **1992**, *1* (12), 1543–1562.

(7) Raymer, M. L.; Sanschagrin, P. C.; Punch, W. F.; Venkataraman, S.; Goodman, E. D.; Kuhn, L. A. Predicting conserved water-mediated and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm. *J. Mol. Biol.* **1997**, *265* (4), 445–464.

(8) Raschke, T. M. Water structure and interactions with protein surfaces. *Curr. Opin. Struct. Biol.* **2006**, *16* (2), 152–9.

(9) Rupley, J. A.; Careri, G. Protein hydration and function. *Adv. Protein Chem.* **1991**, *41*, 37–172.

(10) Svergun, D. I.; Richard, S.; Koch, M. H.; Sayers, Z.; Kuprin, S.; Zaccai, G. Protein hydration in solution: experimental observation by x-ray and neutron scattering. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95* (5), 2267–72.

(11) Cameron, I. L.; Short, N. J.; Fullerton, G. D. Verification of simple hydration/dehydration methods to characterize multiple water compartments on tendon type 1 collagen. *Cell Biol. Int.* **2007**, *31* (6), 531–9.

(12) Ebbinghaus, S.; Kim, S. J.; Heyden, M.; Yu, X.; Heugen, U.; Gruebele, M.; Leitner, D. M.; Havenith, M. An extended dynamical hydration shell around proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (52), 20749–52.

(13) McLain, S. E.; Soper, A. K.; Watts, A. Water Structure Around Dipeptides in Aqueous Solutions. *Eur. Biophys. J.* **2008**, *37* (5), 647–655.

(14) Seki, Y.; Tomizawa, T.; Khechinashvili, N. N.; Soda, K. Contribution of solvent water to the solution X-ray scattering profile of proteins. *Biophys. Chem.* **2002**, *95* (3), 235–252.

(15) 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*, 10.

(16) Jorgensen, W. L.; Madura, J. D. Temperature and size dependence for monte carlo simulations of TIP4P water. *Mol. Phys.* **1985**, *56*, 11.

(17) Mahoney, M. W.; Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* **2000**, 112.

(18) Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. v.; Hermans, J. *Intermolecular Forces*; Reidel: Dordrecht, 1981.

(19) Kuttel, M.; Brady, J. W.; Naidoo, K. J. Carbohydrate solution simulations: producing a force field with experimentally consistent primary alcohol rotational frequencies and populations. *J. Comput. Chem.* **2002**, *23* (13), 1236–1243.

(20) Basma, M.; Sundara, S.; Calgan, D.; Vernali, T.; Woods, R. J. Solvated ensemble averaging in the calculation of partial atomic charges. *J. Comput. Chem.* **2001**, *22* (11), 1125–1137.

(21) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *CHARMM: a program for macromolecular energy, minimization, and dynamics calculations*; Wiley: New York, 1983; Vol. 4, pp 187−217.

(22) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(23) Corzana, F.; Motawia, M. S.; Du Penhoat, C. H.; Perez, S.; Tschampel, S. M.; Woods, R. J.; Engelsen, S. B. A hydration study of (1→4) and (1→6) linked α-glucans by comparative 10 ns molecular dynamics simulations and 500-MHz NMR. *J. Comput. Chem.* **2004**, *25* (4), 573–586.

(24) Dadarlat, V. M.; Post, C. B. Decomposition of protein experimental compressibility into intrinsic and hydration shell contributions. *Biophys. J.* **2006**, *91* (12), 4544–4554.

(25) Imai, T.; Hiraoka, R.; Kovalenko, A.; Hirata, F. Water molecules in a protein cavity detected by a statistical-mechanical theory. *J. Am. Chem. Soc.* **2005**, *127* (44), 15334–15335.

(26) Harano, Y.; Imai, T.; Kovalenko, A.; Kinoshita, M.; Hirata, F. Theoretical study for partial molar volume of amino acids and polypeptides by the three-dimensional reference interaction site model. *J. Chem. Phys.* **2001**, *114*, 9506.

(27) Imai, T.; Kovalenko, A.; Hirata, F. Solvation thermodynamics of protein studied by the 3D-RISM theory. *Chem. Phys. Lett.* **2004**, *395* (1−3), 1–6.

(28) Harrison, R. W.; Kourinov, I. V.; Andrews, L. C. The Fourier-Green's function and the rapid evaluation of molecular potentials. *Protein Eng.* **1994**, *7* (3), 359–369.

(29) Hajdu, F. A model of liquid water Tetragonal clusters: description and determination of parameters. *Acta Chim. (Budapest)* **1977**, *93*, 24.

(30) Huang, X.; Braams, B. J.; Bowman, J. M. Ab initio potential energy and dipole moment surfaces of $(H_2O)_2$. *J. Phys. Chem. A* **2006**, *110* (2), 445–451.

(31) Müller, A.; Bögge, H.; Diemann, E. Structure of a cavity-encapsulated nanodrop of water. *Inorg. Chem. Commun.* **2003**, *6*, 52.

(32) Berman, H.; Henrick, K.; Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* **2003**, *10* (12), 980.

(33) Botti, A.; Bruni, F.; Isopo, A.; Ricci, M. A.; Soper, A. K. Experimental determination of the site-site radial distribution functions of supercooled ultrapure bulk water. *J. Chem. Phys.* **2002**, *117*, 6196–6199.

(34) Petrenko, V. F.; Whitworth, R. W., *Physics of Ice*; Oxford University Press: Oxford, 1999.

(35) Head-Gordon, T. Is water structure around hydrophobic groups clathrate-like. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92* (18), 8308–8312.

(36) Sorenson, J. M.; Hura, G.; Glaeser, R. M.; Head-Gordon, T. What can x-ray scattering tell us about the radial distribution functions of water? *J. Chem. Phys.* **2000**, *113*, 9149.

(37) Soper, A. K. The Radial Distribution Functions of Water and Ice from 220 to 673 K and at Pressures up to 400 MPa. *Chem. Phys.* **2000**, *258* (2−3), 121–137.

(38) Vihinen, M.; Torkkila, E.; Riikonen, P. Accuracy of protein flexibility predictions. *Proteins* **1994**, *19* (2), 141–149.

(39) Vihinen, M. Relationship of protein flexibility to thermostability. *Protein Eng.* **1987**, *1* (6), 477–80.

(40) Parthasarathy, S.; Murthy, M. R. Protein thermal stability: insights from atomic displacement parameters (B values). *Protein Eng.* **2000**, *13* (1), 9–13.

(41) Chen, X.; Weber, I. T.; Harrison, R. W. Molecular dynamics simulations of 14 HIV protease mutants in complexes with indinavir. *J. Mol. Model.* **2004**, *10* (5−6), 373–381.

(42) Harding, M. M. Metal-ligand geometry relevant to proteins and in proteins: sodium and potassium. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2002**, *58* (Pt. 5), 872–874.

(43) Nayal, M.; Di Cera, E. Valence screening of water in protein crystals reveals potential Na^+ binding sites. *J. Mol. Biol.* **1996**, *256* (2), 228–234.

(44) Kovalevsky, A. Y.; Tie, Y.; Liu, F.; Boross, P. I.; Wang, Y. F.; Leshchenko, S.; Ghosh, A. K.; Harrison, R. W.; Weber, I. T. Effectiveness of nonpeptide clinical inhibitor TMC-114 on HIV-1 protease with highly drug resistant mutations D30N, I50V, and L90M. *J. Med. Chem.* **2006**, *49* (4), 1379–1387.

(45) Sherman, D. M.; Collings, M. D. Ion association in concentrated NaCl brines from ambient to supercritical conditions: results from classical molecular dynamics simulations. In Geochemical Transactions 2002; Vol. 3, p 102.

(46) Tromp, R. H.; Neilson, G. W.; Soper, A. K. Water structure in concentrated lithium chloride solutions. *J. Chem. Phys.* **1992**, *96*, 8460.

(47) Hura, G.; Russo, D.; Glaeser, R. M.; Head-Gordon, T.; Krack, M.; Parrinello, M.; Matter, S. Water structure as a function of temperature from X-ray scattering experiments and ab initio molecular dynamics. *Phys. Chem. Chem Phys.* **2003**, *5*, 1981–1991.

(48) Kurinov, I. V.; Harrison, R. W. The influence of temperature on lysozyme crystals. Structure and dynamics of protein and water. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1995**, *51* (Pt. 1), 98–109.

(49) Puibasset, J.; Pellenq, R. J. Confinement effect on thermodynamic and structural properties of water in hydrophilic mesoporous silica. *Eur. Phys. J. E* **2003**, *12* (Suppl 1), 67–70.

JP802795A