THE JOURNAL OF PHYSICAL CHEMISTRY B

ARTICLE

pubs.acs.org/JPCB

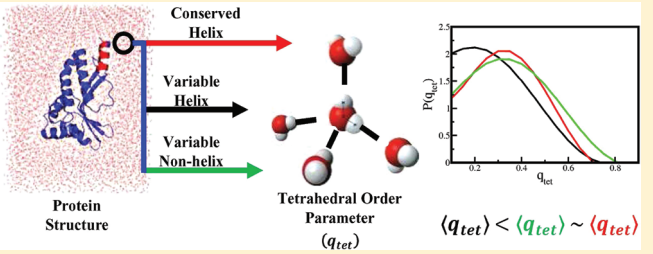
Local Order and Mobility of Water Molecules around Ambivalent Helices

Nicholus Bhattacharjee and Parbati Biswas*

Department of Chemistry, University of Delhi, Delhi 110007 India

Supporting Information

ABSTRACT: Water on a protein surface plays a key role in determining the structure and dynamics of proteins. Compared to the properties of bulk water, many aspects of the structure and dynamics of the water surrounding the proteins are less understood. It is interesting therefore to explore how the properties of the water within the solvation shell around the peptide molecule depend on its specific secondary structure. In this work we investigate the orientational order and residence times of the water molecules to characterize the structure, energetics, and dynamics of the hydration shell water around ambivalent peptides. Ambivalent sequences are identical sequences which display multiple secondary structures in different proteins. Molecular dynamics simulations of representative proteins containing variable helix, variable nonhelix, and conserved helix are also used to explore the local structure and mobility of water molecules in their vicinity. The results, for the first time, depict a different water distribution pattern around the conserved and variable helices. The water molecules surrounding the helical segments in variable helices are found to possess a less locally ordered structure compared to those around their corresponding nonhelical counterparts and conserved helices. The long conserved helices exhibit extremely high local residence times compared to the helical conformations of the variable helices, whereas the residence times of the nonhelical conformations of the variable helices are comparable to those of the short conserved helices. This differential pattern of the structure and dynamics of water molecules in the vicinity of conserved/variable helices may lend valuable insights for understanding the role of solvent effects in determining sequence ambivalency and help in improving the accuracy of water models used in the simulations of proteins.



INTRODUCTION

Water is an integral part of almost all biomolecules, which plays a pivotal role in determining their structure, dynamics, and function. The secondary, tertiary, and quaternary structures of proteins in solution are strongly influenced by the solvent.[1−3] Water molecules lubricate large-scale conformational changes of proteins through the continuous forming and breaking of hydrogen bonds involving the long flexible lateral side chains[4] or act as a ratchet to drive or hinder the directional movements.[5] The organization of water at biomolecular interfaces involves the understanding of protein−water interactions and establishes its significance as a solvent in biological processes.[6−10] Particularly, protein−water interactions mold and control the free-energy landscape that governs the folding, stability, and functional specificity of proteins.[11]

The solvent environment around a protein is characterized by three different types of water molecules:[12] (1) bulk water surrounds the protein molecule at a distance greater than van der Waals contact, (2) individually bound water forms hydrogen bonds with charged or polar protein atoms present in cavities inside the protein,[13] and (3) hydration/biological water at the protein surface, directly interacting with the protein.[14,15] Bulk water moves freely, helping the protein to diffuse by random movement in solution, relative to the other interacting molecules. Individually bound water stabilizes the protein structure by mediating multiple contacts between proteins and their respective ligands.[16−18] Biological water is mostly protein-bound water which interacts with the protein surface to maintain the local structure of the protein in solution. These water molecules act as a glue to stabilize tertiary,[19] higher order structures[20,21] and molecular complexes of the protein[22,23] and regulates its function.[24−27]

Many aspects of the structure and dynamics of biological water are not completely understood. Evidence points out that the properties of the surrounding water molecules induced by the presence of the protein differ considerably from that of bulk water especially its heterogeneous dynamical behavior, which is due to the interactions with the solvent-exposed protein atoms with different chemical composition and topological disorder at the surface of the protein. Furthermore, biological water is visible in protein crystal structures, but bulk water is not observed because there is no defined location for these water molecules.[28,29] Attempts to unravel the structure and dynamics of biological water on the protein surface involve various physical techniques like X-ray crystallography,[30] NMR,[31,32] neutron scattering,[33,34]

Received: July 12, 2011
Revised: September 7, 2011
Published: September 14, 2011

ACS Publications © 2011 American Chemical Society

dielectric relaxation,[35−37] magnetic resonance dispersion,[11] time-resolved fluorescence,[7] and solvation dynamics.[38] These results indicate that these water molecules are distinctly different with respect to structure, diffusion, and the intramolecular vibration modes compared to bulk water. This motivates a deeper comprehension of the conformation and dynamics of water molecules around peptide segments which may lend valuable insights into enzymatic activity, molecular recognition and protein folding, solubility, and drug docking.[39]

In this work, we explore the measures of the orientational order to characterize the structure, energetics, and dynamics of water in the hydration shell of ambivalent peptides and how it depends on the specific secondary structure of the peptide. Ambivalent sequences conform to two different secondary structures (helix−sheet, helix−random coil, sheet−random coil, etc.) in two different proteins. First reported by Kabasch and Sander,[40] these sequences challenge protein structure prediction methods based on sequence homology.[41] Structural plasticity of ambivalent sequences has implications in various misfolding diseases like Alzheimer's disease, CJD scarpie, mad cow disease, etc.[42,43] Recently, we have shown that variable ambivalent helices, i.e., sequences which assume helical conformation in one protein and nonhelical conformation in the other, have different physicochemical properties compared to that of the conserved helices whose conformations are retained throughout the protein database.[44]

This study examines the structure of water in the hydration shell as a function of the distance from helical and nonhelical conformations of variable and conserved helices, respectively. The results, for the first time, depict the different water distribution pattern around the conserved and variable helices; the water map around the nonhelical segments of variable helices closely resembles that of the conserved helices. The water molecules surrounding the helical segments in variable helices are found to be less ordered compared to those around their respective nonhelical counterparts and conserved helices. Water molecules exhibit an uniform spatial arrangement around the nonhelical conformations of variable helices and conserved helices as compared to their helical counterparts. Results from the MD simulations of proteins with ambivalent helices using TIP3P[45] water model also confirms the overestimation of the tetrahedral order parameter as compared to the statistical analysis from the protein crystals.[46] These static measures of local order and energetics are also correlated with the local residence times of water molecules around different amino acid residues of the conserved and variable helices. The results may have far reaching implications in unraveling the effects of water structure and energetics on the conformational changes of ambivalent peptides.

METHODS

Conserved and Variable Helices. The detailed procedure of identifying conserved and variable helices is already outlined in our earlier work.[44] However, a brief recapitulation of this method is also presented here. All α-helices of May-2008 release of PDBselect[47] were compiled to create a database from PDB[48] (Protein Data Bank). The database comprises of protein chains with a sequence identity of 25% or less. Only proteins with X-ray crystallographic structures were considered. All protein chains considered in this study have a resolution of $\leq$3 Å and a crystallographic R-factor of $R \leq 0.3$. The selected database consists of 2586 nonredundant protein chains from 2466 protein structures. Secondary structures in these protein chains were residue-wise annotated with the help of DSSP software.[49] According to the widely used definition, H and G were denoted as helical conformations and all other classes (B, E, I, S, T, -) were considered as nonhelical.[50−52] Neglecting helices less than 5 residues long, there are 11592 helices in the nonredundant database.

Proteins from nine SCOP (Structural Classification of Proteins)[53] classes viz. (I) All alpha proteins, (II) All beta proteins, (III) Alpha and beta proteins(a+b), (IV) Alpha and beta proteins-(a/b), (V) Coiled coiled proteins, (VI) Membrane and cell surface proteins and peptides, (VII) Multidomain proteins (alpha and beta), (VIII) Peptides, and (IX) Small proteins were compiled to obtain exactly identical helical sequences. A structural cutoff resolution $\leq$3 Å and $R \leq 0.3$ were applied on these proteins with the PISCES server.[54] The resultant SCOP database consists of 48244 protein chains from 22309 protein structures. Helical sequences from the nonredundant database were mapped in the SCOP database proteins in the following way to get identical sequences.

For a helix of N residues in the nonredundant database and a protein chain in the SCOP database of M residues an NXM matrix was generated where an element of the matrix, $A(i, j)[i = 1 \rightarrow N, j = 1 \rightarrow M]$, is equal to 1 if ith position of the helix and jth position of the protein chain have identical residues. Otherwise $A(i, j)[i = 1 \rightarrow N, j = 1 \rightarrow M]$ is equal to 0. Now if an element $A(k; l)[k \in i; l \in j] = 1$ and $\Sigma_{m=0}^{N-1} A(k+m, l+m) = N$, where m is a running variable, then the helix from nonredundant database is said to be mapped in position l to $l + N - 1$ of the SCOP protein. Among 11592 helical sequences in the nonredundant database, 6338 helices occur in the SCOP database with varying degrees of conformational changes corresponding to the partially nonhelical conformations. The helices which retain their conformation throughout the SCOP database were termed as conserved helices, whereas the sequences which conform to the nonhelical conformations in any SCOP protein were termed as variable helices. There are 376 variable helices which map into 2785 identical sequences in the SCOP database corresponding to the nonhelical conformations and 4037 conserved helices in the nonredundant database.[44] Among 331 variable helices, 2504 sequences have nonhelical conformations with water of crystallization, whereas 3629 conserved helices have water molecules in their crystal structures.

The present work is performed with this data set of proteins. The solvent contents of these proteins with variable helices, variable nonhelices, and conserved helices are 51.63 ± 0.53%, 51.09 ± 0.17%, and 50.48 ± 0.17%, respectively. The Matthews[55] coefficients are 2.68 ± 0.03, 2.66 ± 0.03, and 2.65 ± 0.02, respectively. For each protein, the sum of the occupancy factor of each water molecule is considered, rather than the actual number of water molecules present.[28,44] Water molecules with low occupancy factor (<0.5) and high B-factors are neglected.

Water Structure. Local tetrahedral arrangement is one of the distinctive features of hydrogen bonded liquid water. The orientational (or tetrahedral) order parameter q_{tet} measures the extent of deviation from this arrangement for a given water molecule with respect to its four nearest neighbors. For pure water, orientational order parameter is defined as[56]

$$q_{tet} = 1 - \frac{3}{8}\sum_{j=1}^{3}\sum_{k=j+1}^{4}\left(\cos(\Psi_{j,k}) + \frac{1}{3}\right)^2 \quad (1)$$

where $\Psi_{j,k}$ is the angle formed between the oxygen atom of a given water molecule with the bond vectors r_{ij} and r_{jk} of its jth and

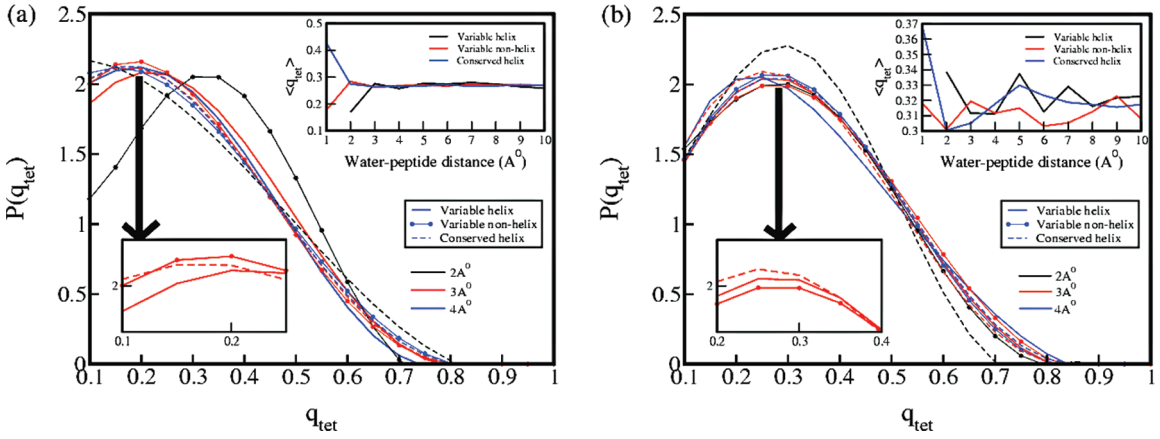


Figure 1. Distribution of the orientational (tetrahedral) order parameter, $P(q_{tet})$, at various water−variable helix/water−variable nonhelix distances calculated from MD simulation trajectory with Δr = 1 Å . Order parameters are calculated (a) including and (b) excluding protein heavy atoms. Distributions at distances with significant variation in $\langle q_{tet} \rangle$ are shown in the figure. A 5th order polynomial curve fitting is done for smoothing of the curves. Inset of the figure shows $\langle q_{tet} \rangle$ at various water−variable helix/water−variable nonhelix distances.

*k*th nearest neighbors of the same type. For a perfect tetrahedral arrangement, q_{tet} = 1. Conversely, for a random arrangement with uncorrelated uniform angular distributions, q_{tet} = 0. For water molecules surrounding the ambivalent peptide, the orientational order parameter is calculated in two ways (i) by considering only water molecules as four nearest neighbors and (ii) by considering both heavy atoms of the peptide (C, N, O, etc.) and water molecules as nearest neighbors of the oxygen atom of the selected water molecule.

Molecular Dynamics Simulation. The structure and dynamics of water molecules around the variable and conserved helices were also studied using molecular dynamics simulations. MD simulations were performed by the AMBER 9 package.[57] The protein crystal structures were chosen from the PDB according to the following selection criteria: The resolution of these crystal structures must be ≤3 Å with the crystallographic *R*-factor, $R \leq 0.3$. The missing hydrogen atoms were added with LEAP subroutine. Each protein structure was solvated in a cubic box with TIP3P water model.[45] During solvation, the buffering distance between edge of the box and the protein was maintained at 10 Å and the closeness parameter was set at 1 Å. Depending upon the charge of the solvated proteins, they were neutralized either by Na+ or by Cl-. The ff99SB force field was used with the periodic boundary conditions. This force field presents a careful reparametrization of the backbone torsion terms in ff99 and achieves a better balance of four basic secondary structure elements (PP_{II}, β, α_L and α_R).[58] This force field also shows the best agreement with experimental data.[59] Nonbonded interactions were treated with a cutoff range of 8 Å and the list of nonbonded neighbors were updated every 50 fs. The long-range electrostatic interactions were calculated using the PME algorithm[60] with a real space cutoff of 8.0 Å and fourth order spline interpolation with a grid spacing of 1.0 Å . The SHAKE algorithm was used to constrain all bond lengths to their equilibrium distances.[61]

Each system was energy minimized twice, the first step involved the conjugate gradient energy minimization of the solvent by keeping the protein constrained followed by minimizing the energy of the whole system. A two stage equilibration was performed. The solvated protein was simulated in a NVT ensemble at a low initial temperature of 100 K and the temperature was gradually raised to 300 K for 10 ps at a constant volume. This was followed by a NPT equilibration for 100 ps at a constant temperature of 300 K and a pressure of 1 bar. Constant temperature was maintained through weak coupling to the Berendsen temperature bath with coupling constant of 2 ps while a constant pressure was maintained through weak coupling to isotropic pressure bath with a coupling constant of 1 ps.[62] For each system, the NPT production run length was 10 ns with a time step of 2 fs.

Technical Details of Calculations. The distribution of the tetrahedral order parameter from both crystallographic data and molecular dynamics simulations are plotted as a function of protein−water distance *r*. At each *r*, the distribution of water molecules are calculated by counting the number of water molecules whose oxygen atoms lie within a distance between *r* and $r - \Delta r$. The bin size, Δr, is chosen to be 1 and 0.5 Å . In the main text, we display the results for Δr = 1 Å, whereas other results are provided in the Supporting Information. At each *r*, the fraction of water molecules with q_{tet} values ranging between $q_{tet} - dq_{tet}/2$ and $q_{tet} + dq_{tet}/2$ is denoted as $P(q_{tet})dq_{tet}$ where dq_{tet} = 0.05. Average value of q_{tet}, $\langle q_{tet} \rangle$, is also plotted at each *r*.

RESULTS AND DISCUSSIONS

Water Structure As a Function of Distance from Ambivalent Helices. The local structure and orientation of the hydration shell water molecules is explored for the first time in the context of ambivalent helices. In the present work with protein crystal structures, both individually bound water and hydration water are considered. Figure 1, panels a and b, illustrate the distribution of q_{tet} of the water molecules as a function of the respective distances from the variable and conserved helices. For variable helices the distribution of water molecules is studied with respect to both helical and nonhelical structures. For helical conformations of the variable helices, water molecules are excluded up to a distance of 1 Å unlike their nonhelical counterparts and conserved helices.

The pattern of variation of q_{tet} of the water molecules are quite different for the helical and nonhelical conformations of the variable helices and the conserved ones, with the inclusion (Figure 1a) and exclusion (Figure 1b) of the protein heavy atoms as four nearest neighbors. This is also evident from the plot of $\langle q_{tet} \rangle$ against the distance from the variable/conserved helices shown as insets in each graph. Figure 1a depicts the distribution

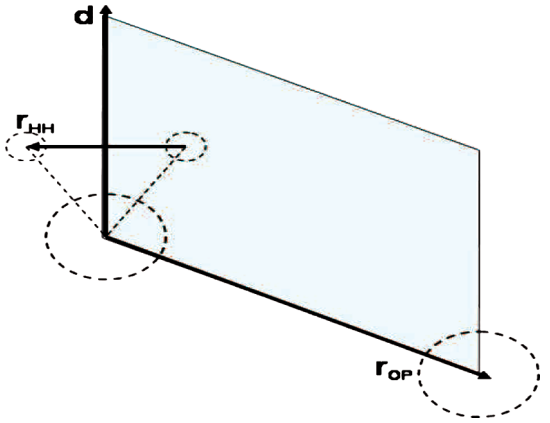


Figure 2. Graphical illustration of the vectors used to calculate cos θ and sin ϕ. **d** is the dipole moment vector, r_{HH} is the hydrogen−hydrogen vector of water molecule and r_{OP} is the vector from water molecule's oxygen atom to nearest oxygen or nitrogen atom of the peptide segment.

of q_{tet}, $P(q_{tet})$, in the vicinity of the variable/conserved helices with the protein heavy atoms as the nearest neighbors. $P(q_{tet})$ has a broad distribution at different distances from the variable helices. The maximum value of q_{tet} is found to be 0.25, indicating that the water molecules are mostly disordered around variable helices. Simulation of bulk water consisting of 364 TIP3P water molecules (results not shown) yields a distribution maxima with q_{tet} value lying between 0.35 to 0.4 which agrees well with the earlier studies.[63] The distributions at different distances reveal a similar characteristic pattern for the chosen range of q_{tet}. The maximum value of q_{tet} is found to be higher when the protein heavy atoms are excluded in calculating q_{tet}. This phenomenon is more pronounced at lower values of Δr i.e., (water−variable helix/variable nonhelix/conserved helix distance) Δr = 0.5 Å as shown in the Supporting Information, Figures 1 and 2. The water molecules are found to have a more tetrahedrally structured local environment when the four nearest neighbors are water molecules rather than the protein heavy atoms. This is also evident from the $\langle q_{tet} \rangle$ plotted against water-variable helix/variable nonhelix/conserved helix distance in the insets of the Figure 1, panels a to b. Values of $\langle q_{tet} \rangle$ are consistently higher when the protein heavy atoms are excluded.

The robustness of the results may be validated by considering another subset of the database with proteins of resolution $\leq$2 Å and R factor of $R \leq 0.3$. The number of proteins containing variable helices, variable nonhelices and conserved helices in this subset are 159, 883, and 1989 respectively. This data set is used to calculate q_{tet} as function of distance from the variable helix, nonhelix and conserved helix. The calculated results of $P(q_{tet})$ vs q_{tet} are displayed in the Supporting Information, Figures 3−6. The general trend of these results qualitatively matches with that of Figures 1a and 1b, despite the lower number of proteins present due to the stringent selection criteria.

With the inclusion of the protein heavy atoms, the value of $\langle q_{tet} \rangle$ shows a peak between 2 and 3 Å from the variable/conserved helical sequences as shown in Figures 1a. Statistical analysis of the radial distribution of water molecules in the high resolution protein structures also reveal the existence of the first hydration layer at this distance.[28] Beyond this distance, $\langle q_{tet} \rangle$ reaches a constant value asymptotically. However, a similar trend is not observed when only water molecules are considered for the calculation of q_{tet} as shown in Figure 1b. The value of $\langle q_{tet} \rangle$ fluctuates considerably even at a large distance from the variable/conserved helices. The q_{tet} values account for the ordered water molecules which are present very near to the protein surface but far off from the variable/conserved helices. These water molecules incorporate the protein heavy atoms to satisfy their tetrahedral geometry and hence have lower values of q_{tet} compared to those calculated with the inclusion of the protein heavy atoms as nearest neighbors. These water molecules with lower values of q_{tet}, present at a large distance from the variable/conserved helices may account for the fluctuating nature of $\langle q_{tet} \rangle$ in the insets of the Figure 1b.

The water molecules surrounding the nonhelical conformations of the variable helices are found to be more ordered compared to their corresponding helical counterparts. Statistical analysis and MD simulation studies of the hydrogen bonding pattern of the buried water molecules show that they prefer to form hydrogen bonds with residues forming irregular secondary structures, i.e., turns, coils, loops, etc.[64] The increased tendency of buried water to form hydrogen bonds with irregular secondary structures justifies the higher values of q_{tet} near the nonhelical conformations of the variable helices as compared to their complementary helical conformations.

The spatial arrangement of water molecules around the helical and nonhelical conformations of variable helices are also probed up to a distance of 4 Å from the protein surface. This spatial arrangement is obtained from the distribution of cos θ and sin ϕ, where θ and ϕ are the orientation angles between the oxygen and nitrogen atoms of the peptide segment closest to each water molecule.[65] θ is the angle between the dipole moment of the water molecule and the vector from the water molecule's oxygen atom to the nearest oxygen or nitrogen atom of the peptide segment. ϕ is obtained by subtracting $\pi/2$ from the angle formed between the plane spanned by the water molecule and the plane spanned by dipole moment of the water molecule and the vector from the water molecule's oxygen atom to the nearest oxygen or nitrogen atom of the peptide segment. Mathematically if **d** is the dipole moment vector of the water molecule, r_{OP} is the vector from water molecule's oxygen atom to nearest oxygen or nitrogen atom of the peptide segment and r_{HH} is the hydrogen−hydrogen vector of water molecule (refer to Figure 2) then $\cos\theta = e_d \cdot e_{OP}$ and $\sin\phi = (e_d \times e_{OP}) \cdot (e_d \times e_{HH})$, where **e** is an unit vector in the given direction. The missing hydrogen atoms in the water molecules are added by the LEAP module of AMBER 9 in the PDB files.

Figure 3 illustrates the distribution of fraction of water molecules as a function of cosθ and sinϕ for both helical and nonhelical conformations of variable helices. The results show that the water molecules around the nonhelical conformations are uniformly distributed compared to those around the helical conformations of variable helices. Water molecules surrounding the helical conformations have a narrow range of cosθ and sinϕ. Homogeneous distribution of water molecules around the nonhelical conformations prevents steric congestion for a given orientation. This leads to an effective tetrahedral arrangement of these water molecules with their respective neighbors and may account for a locally ordered water structure around the nonhelical conformations. Unlike the variable helices, water molecules are found to be highly ordered near the conserved helices. Figure 1 also depicts the distribution of q_{tet} as a function of the distance from conserved helices. Figure 3 portrays the spatial arrangement of water around the conserved helices; the water molecules are found to be uniformly distributed, which is

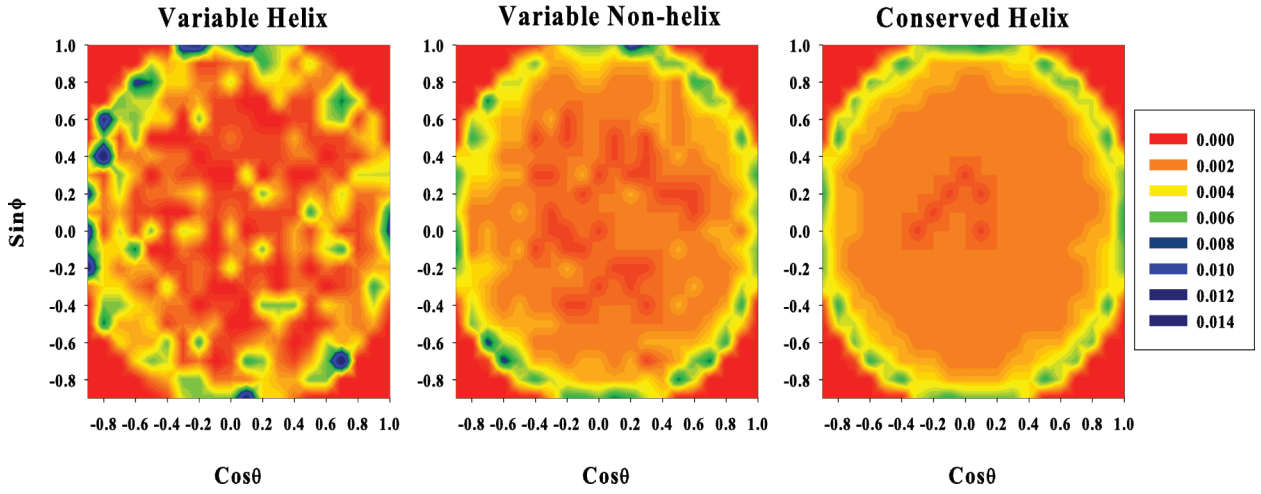


Figure 3. Distribution of fraction of water molecules as function of cos θ and sin ϕ for helical and nonhelical conformations of variable helices and conserved helices.

necessary for a well-defined tetrahedral arrangement of the water molecules and increases the orientational order near the conserved helices.

Water Structure from MD Simulations of Proteins. The distributions of q_{tet} as a function of the distance from the variable and conserved helices are also studied using molecular dynamics simulations and compared with those obtained from the database. For the simulation results, this distance dependent order parameter q_{tet} is computed by averaging over the entire run for all water molecules whose oxygen atoms lie between r and $r - \Delta r$. Proteins containing a variable helical sequence in helical (PDB: 1H4LD) and nonhelical (PDB: 1UNGE) conformations are simulated. The sequence fragment 148−162 (TSELLRCLGEFLCRR) in both proteins are identical. This sequence adopts a helical conformation in 1H4LD while it is found as a random coil in 1UNGE. Conserved helices are simulated using the model protein crystal 1BH8B which contains four conserved helices at positions 3−14 (EEQLNRYEMYRR), 19−30 (KAAIKRLIQSIT), 36−63 (QNVVIAMSGISKVFVGEVVEEALDVCEK), and 72−84 (PKHMREAVRRLKS). In our previous work,[44] the longer conserved helices are found to display different conformational variability under molecular dynamics simulations compared to the smaller ones. The variable helices retain their conformations, whereas their respective nonhelical counterparts may conform to helical, partially helical or complete nonhelical conformations. The distribution of q_{tet} are studied as a function of the distance from the conserved helices for both the long helix (36−63 residues of 1BH8B) and the short helix (72−84 residues of 1BH8B).

Figure 4 depicts the time evolution of the backbone RMSD and radius of gyration (R_G) of these three proteins with respect to their initial structures. It is evident that the structural deviation are minimal for the variable helix 1H4LD followed by the nonhelical conformation 1UNGE, whereas it shows a fluctuating increase for the long conserved helix 1BH8B as it breaks into two shorter helices. The radius of gyration, R_G, also reveals a similar trend with a marked decrease for 1BH8B, while for the helical and nonhelical conformations of the variable helix it remains almost constant. The results show that the values of RMSD and R_G are stable throughout the simulation trajectory.

Extensive simulations of all representative proteins containing variable helical/nonhelical segments and conserved helices from

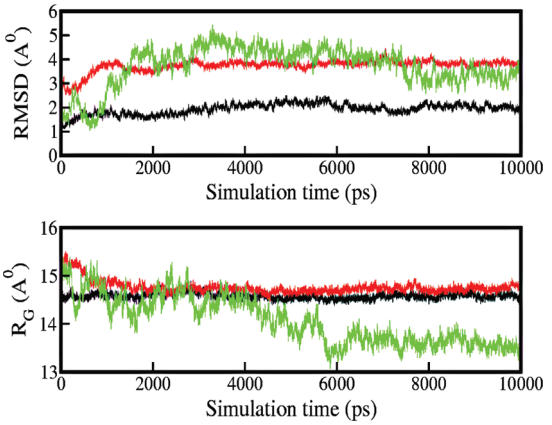


Figure 4. RMSD and radius of gyration (R_G) plotted against the simulation time for protein chains 1H4LD (black), 1UNGE (red), and 1BH8B (green).

all SCOP classes are performed in our previous work.[44] The simulations of three proteins containing four different types of sequences, i.e., variable helix (residues 148−162 of 1H4LD), variable nonhelix (residues 148−162 of 1UNGE), short conserved helix (residues 72−84 of 1BH8B), and long conserved helix (residues 36−63 of 1BH8B)) are highlighted only as representative examples in the present study. Figure 5 exhibits the evolution of the secondary structure of these proteins as a function of the simulation time. The results show that the variable helix (residues 148−162 of 1H4LD) retains its structure after a simulation time of 10 ns, whereas its nonhelical counterpart (residues 148−162 of 1UNGE) conforms to a partial helical structure. The structure of the short conserved helix (residues 72−84 of 1BH8B) is conserved, while the long conserved helix (residues 36−63 of 1BH8B) is partially transformed to the nonhelical conformation, as it frays into two shorter helices.[44] These simulations sample a general trend of the behavior of variable helix/nonhelix and conserved helices.

The mobility of water molecules surrounding the ambivalent helices is measured in terms of residence times (τ) of water molecules, which is defined by the survival probability function $F_s(t)$.

dx.doi.org/10.1021/jp2066106 |J. Phys. Chem. B 2011, 115, 12257–12265

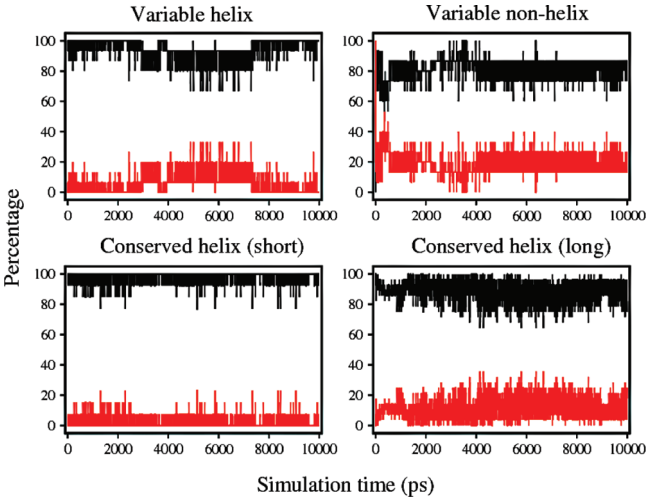


Figure 5. Time evolution of the helical (black) and nonhelical (red) conformations for variable helix (sequence 148−162 of 1H4LD), variable nonhelix (sequence 148−162 of 1UNGE), short conserved helix (sequence 72−84 of 1BH8B), and long conserved helix (sequence 36−63 of 1BH8B).

Survival probability function evaluates the lifetime and population of water molecules which remains within a particular region continuously for a given period of time t. It is assumed here that the time interval between the consecutive stored configurations (1 ps) is sufficiently small for a significant displacement of the water molecules. The survival probability may be defined as

$$F_s(t) = \sum_{j=1}^{N_w} \frac{1}{t_{run} - t + 1} \sum_{t_0}^{t_{run} - t_0} f_j(r, t_0, t_0 + t) \qquad (2)$$

where $f_j(r,t_0,t_0 + t)$ takes a value of 1 if the jth water molecule is present at any given volume element r continuously from time t_0 to $t_0 + t$ and 0 otherwise. N_w is the total number of water molecules present in the system, and t_{run} is the length of the simulation. The region r is defined as the volume surrounding the peptide sequence (conserved/variable helix) up to a distance of 4 Å. The survival probability function $F_s(t)$ shows a typical sharp initial fall followed by an intermediate region which may be fitted to the empirical expression $\ln(F_s(t)) = -t/\tau + c$ to obtain the residence time. Although alternative ways to calculate the residence time has been used in previous studies,[66−69] they do not impose continuous presence of water molecule over a definite time interval t at any given location. Figure 6 shows the survival probability function, $F_s(t)$, for water molecules around variable (both helical and nonhelical conformations) and conserved (both long and short) helices. The initial decay is ignored and the intermediate portion of the survival probability function is fitted to an exponential function to extract the residence time.[63,70] Residence times of water molecules around the nonhelical conformation of variable helices and short conserved helices are comparable to the local residence times of water present in the neighborhood of small peptides for a given temperature.[63] The water molecules around the helical conformations of variable helices records a higher value of the residence time, while it exhibits approximately a 7-fold increase for the long conserved helices. Previous results have shown that long conserved helices fray into two or more smaller helices giving rise to random coil

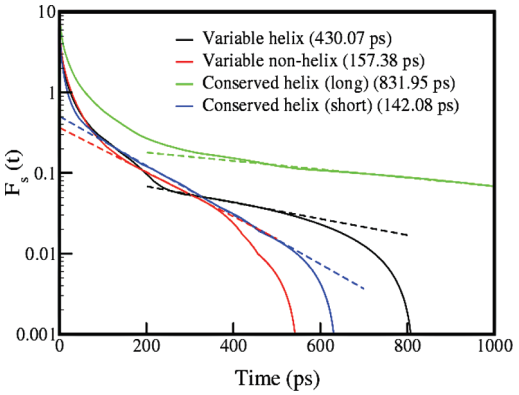


Figure 6. Survival probability function of water molecules around variable and conserved helices. The dotted lines denote the fitting of the intermediate regions of the curves. Residence times for each curve are provided along with the legends.

structures connecting the smaller helices.[44] Since bound water molecules can readily form hydrogen bonds with irregular coiled structures of proteins,[64] the fraying of long conserved helices into randomly coiled structures may account for the extremely high residence times of water molecules around them. This is also consistent with the results of earlier studies[71] which indicate that the residence time of bound water molecules increases due to the presence of the protein.

The high residence time of water molecules in the vicinity of the protein surface may be related to two very different explanations for rationalizing its slow dynamics. The model proposed by Nandi and Bagchi postulates a dynamic equilibrium between bound water incapable of independent reorientation and free water molecules capable of rapid reorientation.[72] Water molecules at the surface of a protein may be considered as distinctly different from that of the bulk because of their strong hydrogen bonding with the protein surface. Based on this assumption, Bagchi and Zewail groups have proposed that slow dynamics of water near protein is an inherent feature of water in the potential field of a protein.[7,35,37,72−76] In a very different second explanation, Halle and Nilsson assumed that water−protein interactions are comparable to water−water interactions.[11,77] The local dynamics of water near the surface of protein is assumed to be almost similar to the time scale of bulk water. The slow dynamics of the protein in aqueous solvent is responsible for the observed long time scale of water in time-dependent Stokes shift experiments. Recent simulation studies have also validated the occurrence of slow dynamics of water molecules around the protein.[78]

Figures 7 and 8 show the distribution of water molecules as a function of q_{tet} at various distances from variable (both for helical and nonhelical conformations) and conserved (both short and long) helices. Figures 7a and 8a portray q_{tet} of water molecules including the protein heavy atoms, whereas Figures 7b and 8b show the same excluding protein heavy atoms. Due to the unavailability of any water molecule within 2 Å from any of the peptide segments (both for variable helical and nonhelical sequences and short and long conserved helices), comparison of q_{tet}, for water molecules very near to the ambivalent helices, with those of the crystal structures is not possible.

However, there are distinct differences observed in the distribution pattern of water molecules with respect to q_{tet}, calculated from the simulation results as compared to those from the crystal structure database. The calculated distribution from the

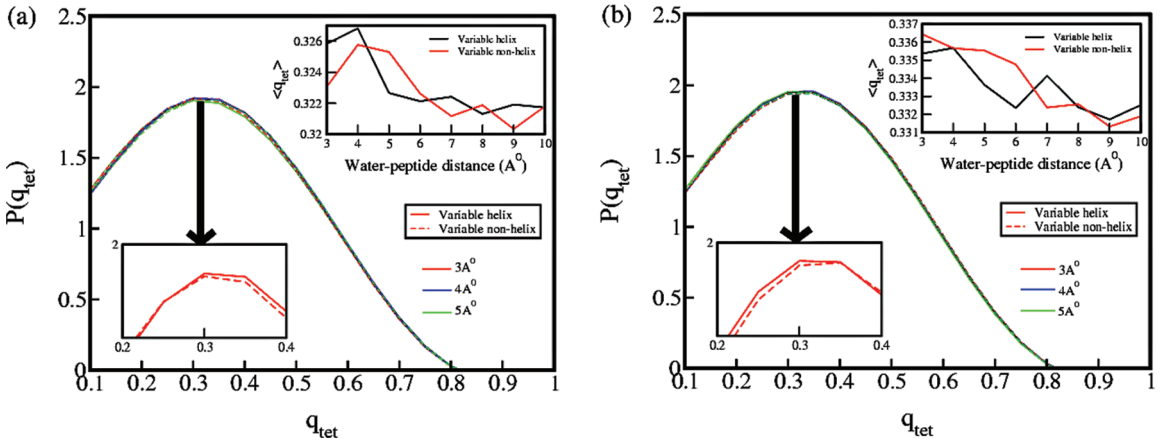


Figure 7. Distribution of the orientational (tetrahedral) order parameter, $P(q_{tet})$, at various water−variable helix/water−variable nonhelix distances calculated from MD simulation trajectory with Δr = 1 Å . Order parameters are calculated (a) including and (b) excluding protein heavy atoms. Distributions at distances with significant variation in $\langle q_{tet} \rangle$ are shown in the figure. A 5th order polynomial curve fitting is done for smoothing of the curves. Inset of the figure shows $\langle q_{tet} \rangle$ at various water−variable helix/water−variable nonhelix distances.

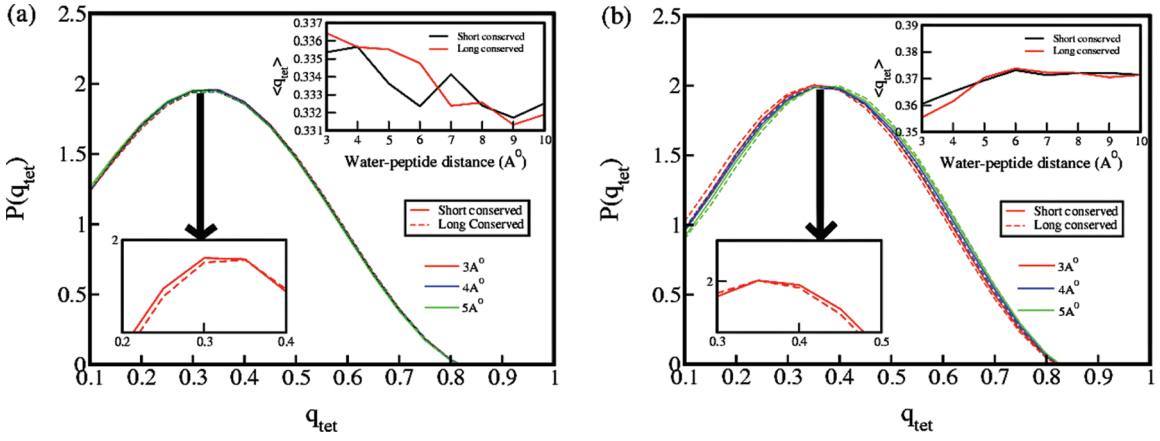


Figure 8. Distribution of the orientational (tetrahedral) order parameter, $P(q_{tet})$, at various water-short conserved helix/water-long conserved helix distances calculated from MD simulation trajectory with Δr = 1 Å . Order parameters are calculated (a) including and (b) excluding protein heavy atoms. Distributions at distances with significant variation in $\langle q_{tet} \rangle$ are shown in the figure. A 5th order polynomial curve fitting is done for smoothing of the curves. Inset of the figure shows $\langle q_{tet} \rangle$ at various water-short conserved helix/water-long conserved helix distances.

simulation trajectory is much smoother compared to those calculated from the crystal structure database of variable and conserved helices. This may be due to the fact that MD trajectory encompasses a large fraction of the phase space in comparison to the database of experimental structures and hence possibly includes all possible water structures. Moreover, analogous to previous simulation results on small peptides,[63] q_{tet} has a slightly higher value in comparison to those calculated from the crystal structure database at different distances from the protein surface. This observation confirms our earlier results of q_{tet} overestimation by MD simulation in comparison to crystal structures.[46] Average tetrahedral order parameter with respect to the distance from variable/conserved helices behaves differently when protein heavy atoms are included or excluded. The average tetrahedral order parameter records a higher value near the protein surface and decays to an average value asymptotically at long distances when the protein heavy atoms are included, while an opposite trend is observed in the other case. Water molecules near the protein surface opt the protein heavy atoms as nearest neighbors to attain a locally ordered tetrahedral geometry. By excluding these heavy atoms the orientational order of water molecules is disrupted near the protein surface due to the less availibility of neighboring water molecules. This justifies the lower values of $\langle q_{tet} \rangle$ near the protein surface shown in Figures 7b and 8b compared to the Figures 7a and 8a.

CONCLUSIONS

The present study attempts a systematic analysis of the structure and dynamics of hydration shell water molecules in the vicinity of the ambivalent proteins. The structure of water in the hydration shell depicted as a function of the distance from helical and nonhelical conformations of variable and conserved helices, show for the first time, different water distribution pattern around the conserved and variable helices; the water molecules are uniformly distributed around the nonhelical segments of the variable helices and have a locally ordered structure which closely resemble that of the conserved helices. The water molecules are more distorted from their ordered tetrahedral geometry either in the vicinity of the protein or far off from the

protein surface. Near the protein surface, it may be due to distance less than that required for optimum H-bonding while at larger distances there are less water molecules available for a perfectly ordered tetrahedral configuration. The MD simulation results of ambivalent proteins with variable/conserved helices overestimates the tetrahedral order parameters of water molecules in proteins.

The average tetrahedral order parameter records a higher value near the protein surface and decays to an average value asymptotically at long distances when the protein heavy atoms are included, whereas the opposite trend is observed when the protein heavy atoms are excluded. The long conserved helices exhibit extremely high local residence times compared to the helical conformations of the variable helices while the residence times of the nonhelical conformations of the variable helices are comparable to those of the short conserved helices. Long conserved helices are found to fray into two or more smaller helices giving rise to random coil structures connecting the smaller helices.[44] These long conserved helices fraying into randomly coiled structures may account for the extremely high residence times of water molecules around them as bound water molecules readily form hydrogen bonds with irregular coiled structures of proteins.[64] The results establish that the structure and dynamics of water molecules within the hydration shell of the protein molecule is influenced by its secondary structure.

■ ASSOCIATED CONTENT

Supporting Information. Tetrahedral order parameter distributions of water calculated from crystal structures and simulation trajectories with lower binning interval (Δr = 0.5 Å). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: pbiswas@chemistry.du.ac.in.

■ ACKNOWLEDGMENT

The authors gratefully acknowledge the financial assistance from Department of Science and Technology (Project No. SR/S1/PC-07/06), India and Delhi University research grant. The authors also acknowledge Bioinformatics Resources and Applications Facility (BRAF) of Centre for Development of Advanced Computing (CDAC), India for providing computational facility in Biogene cluster. N.B. acknowledges CSIR, INDIA for providing financial assistance in form of SRF.

■ REFERENCES

(1) Prabhu, N.; Sharp, K. *Chem. Rev.* **2006**, *106*, 1616.
(2) Ball, P. *Chem. Rev.* **2008**, *108*, 74.
(3) Levy, Y.; Onuchic, J. N. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 389.
(4) Nakasako, M.; Oka, T.; Mashumo, M.; Takahashi, H.; Shimada, I.; Yamaguchi, Y.; Kato, K.; Arata, Y. *J. Mol. Biol.* **2005**, *351*, 627.
(5) Nakasako, M.; Fujisawa, T.; Adachi, S.; Kudo, T.; Higuchi, S. *Biochemistry* **2001**, *40*, 3069.
(6) Bellissent-Funel, M. C., Ed.; *Hydration Processes in Biology*; IOS Press: Amsterdam, The Netherlands, 1999.
(7) Bagchi, B. *Chem. Rev.* **2005**, *105*, 3197.
(8) Li, Z.; Lazaridis, T. *Phys. Chem. Chem. Phys.* **2007**, *9*, 573.
(9) Qvist, J.; Persson, E.; Mattea, C.; Halle, B. *Faraday Discuss.* **2009**, *141*, 131.
(10) Born, B.; Kim, S. J.; Ebbinghaus, S.; Gruebele, M.; Havenith, M. *Faraday Discuss.* **2009**, *141*, 161.
(11) Halle, B. *Phil. Trans. R. Soc. London B* **2004**, *359*, 1207.
(12) Purkiss, A.; Skoulakis, S.; Goodfellow, J. M. *Phil. Trans. R. Soc. London A* **2001**, *359*, 1515.
(13) Roberts, B. C.; Mancera, R. L. *J. Chem. Inf. Model.* **2008**, *48*, 397.
(14) Nakasako, M. *J. Mol. Biol.* **1999**, *289*, 547.
(15) Higo, J.; Nakasako, M. *J. Comput. Chem.* **2002**, *23*, 1323.
(16) Rashin, A. A.; Iofin, M.; Honig, B. *Biochemistry* **1986**, *25*, 3619.
(17) Meyer, E. *Protein Sci.* **1992**, *1*, 1543.
(18) Raymer, M. L.; Sanschagrin, P. C.; Punch, W. F.; Venkataraman, S.; Goodman, E. D.; Kuhn, L. A. *J. Mol. Biol.* **1997**, *265*, 445.
(19) Bottoms, C. A.; White, T. A.; Tanner, J. J. *Proteins* **2006**, *64*, 404.
(20) Nakasako, M.; Odaka, M.; Yohda, M.; Dohmae, N.; Takio, K.; Kamiya, N.; Endo, I. *Biochemistry* **1999**, *38*, 9887.
(21) Rodier, F.; Bahadur, R. P.; Chakrabarti, P.; Janin, J. *Proteins* **2005**, *60*, 36.
(22) Bhat, T. N.; Bentley, G. A.; Boulot, G.; Greene, M. I.; Tello, D.; Acqua, W. D.; Souchon, H.; Schwarz, F. P.; Mariuzza, R. A.; Poljak, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 1089.
(23) Nakasako, M.; Zikihara, K.; Matsuoka, D.; Katsura, H.; Tokutomi, S. *J. Mol. Biol.* **2008**, *381*, 718.
(24) Ferrand, M.; Dianoux, A. J.; Petry, W.; Zaccaï, G. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 9668.
(25) Chen, S.-H.; Liu, L.; Fratini, E.; Baglioni, P.; Faraone, A.; Mamontov, E. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 9012.
(26) Graber, M.; Bousquet-Dubouch, M.-P.; Lamare, S.; Legoy, M.-D. *Biochim. Biophys. Acta* **2003**, *1648*, 24.
(27) Lind, P. A.; Daniel, R. M.; Monk, C.; Dunn, R. V. *Biochim. Biophys. Acta* **2004**, *1702*, 103.
(28) Chen, X.; Weber, I.; Harrison, R. W. *J. Phys. Chem. B* **2008**, *112*, 12073.
(29) Shou, J.-J.; Zeng, G.; Zhang, Y.-H.; Lu, G. Q. M. *J. Phys. Chem. B* **2009**, *113*, 9633.
(30) Huber, R.; Kukla, D.; Bode, W.; Schwager, P.; Bartels, K.; Deisenhofer, J.; Steigemann, W. *J. Mol. Biol.* **1974**, *89*, 73.
(31) Otting, G.; Liepinsh, E.; Wuthrich, K. *Science* **1991**, *254*, 974.
(32) Modig, K.; Liepinsh, E.; Otting, G.; Halle, B. *J. Am. Chem. Soc.* **2004**, *126*, 102.
(33) Bellissent-Funel, M.-C. *J. Mol. Liq.* **2000**, *84*, 39.
(34) Dellerue, S.; Bellissent-Funel, M.-C. *Chem. Phys.* **2000**, *258*, 315.
(35) Pal, S. K.; Peon, J.; Bagchi, B.; Zewail, A. H. *J. Phys. Chem. B* **2002**, *106*, 12376.
(36) Denisov, V. P.; Halle, B. *Faraday Discuss.* **1996**, *103*, 227.
(37) Nandi, N.; Bhattacharyya, K.; Bagchi, B. *Chem. Rev.* **2000**, *100*, 2013.
(38) Grant, E. H.; Sheppard, R. J.; South, G. P. *Dielectric behaviour of biological molecules in solution*; Oxford University Press: Oxford, U.K., 1978.
(39) Raschke, T. M. *Curr. Opin. Struct. Biol.* **2006**, *16*, 152.
(40) Kabsch, W.; Sander, C. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 1075.
(41) Rost, B. *J. Struct. Biol.* **2001**, *134*, 204.
(42) Kelly, J. W. *Curr. Opin. Struct. Biol.* **1998**, *8*, 101.
(43) Serpell, L. C. *Biochim. Biophys. Acta* **2000**, *1502*, 16.
(44) Bhattacharjee, N.; Biswas, P. *BMC Bioinformat.* **2010**, *11*, 519.
(45) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(46) Bhattacharjee, N.; Biswas, P. *Biophys. Chem.* **2011**, *158*, 73.
(47) Hobohm, U.; Sander, C. *Protein Sci.* **1994**, *3*, 522.
(48) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J., Jr.; E., F. M.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.
(49) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.
(50) Guo, J.; Chen, H.; Sun, Z.; Lin, Y. *Proteins* **2004**, *54*, 738.

(51) Hua, S.; Sun, Z. *J. Mol. Biol.* **2001**, *308*, 397.
(52) Kim, H.; Park, H. *Protein Eng.* **2003**, *16*, 553.
(53) Murzin, A. G.; Brenner, S. E.; Hubbard, T.; Chothia, C. *J. Mol. Biol.* **1995**, *247*, 536.
(54) Wang, G.; Dunbrack, R. L., Jr. *Bioinformatics* **2003**, *19*, 1589.
(55) Matthews, B. W. *J. Mol. Biol.* **1968**, *33*, 491.
(56) Errington, J. R.; Debenedetti, P. G. *Nature* **2001**, *409*, 318.
(57) Case, D. A. et al. *AMBER 9*; University of California: San Francisco, 2006.
(58) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712.
(59) Best, R. B.; Hummer, G. *J. Phys. Chem. B* **2009**, *113*, 9004.
(60) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
(61) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(62) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(63) Agarwal, M.; Kushwaha, H. R.; Chakravarty, C. *J. Phys. Chem. B* **2010**, *114*, 651.
(64) Park, S.; Saven, J. G. *Proteins* **2005**, *60*, 450.
(65) Bergman, D. L.; Lyubartsev, A. P.; Laaksonen, A. *Phys. Rev. E* **1999**, *60*, 4482.
(66) Handgraaf, J.-W.; Zerbetto, F. *Proteins* **2006**, *64*, 711.
(67) Lee, S. H.; Rasaiah, J. C. *J. Phys. Chem.* **1996**, *100*, 1420.
(68) Laage, D.; Hynes, J. T. *J. Phys. Chem. B* **2008**, *112*, 7697.
(69) Koneshan, S.; Rasaiah, J. C.; Lynden-Bell, R. M.; Lee, S. H. *J. Phys. Chem. B* **1998**, *102*, 4193.
(70) Merzel, F.; Smith, J. C. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5378.
(71) Makarov, V. A.; Andrews, B. K.; Smith, P. E.; Pettitt, B. M. *Biophys. J.* **2000**, *79*, 2966.
(72) Nandi, N.; Bagchi, B. *J. Phys. Chem. B* **1997**, *101*, 10954.
(73) Pal, S. K.; Zewail, A. H. *Chem. Rev.* **2004**, *104*, 2099.
(74) Nandi, N.; Bagchi, B. *J. Phys. Chem. A* **1998**, *102*, 8217.
(75) Bandyopadhyay, S.; Chakraborty, S.; Balasubramanian, S.; Pal, S.; Bagchi, B. *J. Phys. Chem. B* **2004**, *108*, 12608.
(76) Bandyopadhyay, S.; Chakraborty, S.; Balasubramanian, S.; Bagchi, B. *J. Am. Chem. Soc.* **2005**, *127*, 4071.
(77) Nilsson, L.; Halle, B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13867.
(78) Li, T.; Hassanali, A. A.; Singer, S. J. *J. Phys. Chem. B* **2008**, *112*, 16121.