Biophysical Chemistry 158 (2011) 73–80

Contents lists available at ScienceDirect

Biophysical Chemistry

journal homepage: http://www.elsevier.com/locate/biophyschem

Structure of hydration water in proteins: A comparison of molecular dynamics simulations and database analysis

Nicholus Bhattacharjee, Parbati Biswas *

Department of Chemistry, University of Delhi, Delhi, India 110007

ARTICLE INFO

Article history:
Received 9 April 2011
Received in revised form 7 May 2011
Accepted 8 May 2011
Available online 27 May 2011

Keywords:
Orientational/tetrahedral order parameter
Radial distribution
Solvent models
Molecular dynamics simulation

ABSTRACT

Hydration layer water molecules play important structural and functional roles in proteins. Despite being a critical component in biomolecular systems, characterizing the properties of hydration water poses a challenge for both experiments and simulations. In this context we investigate the local structure of hydration water molecules as a function of the distance from the protein and water molecules respectively in 188 high resolution protein structures and compare it with those obtained from molecular dynamics simulations. Tetrahedral order parameter of water in proteins calculated from previous and present simulation studies show that the potential of bulk water overestimates the average tetrahedral order parameter compared to those calculated from crystal structures. Hydration waters are found to be more ordered at a distance between the first and second solvation shell from the protein surface. The values of the order parameter decrease sharply when the water molecules are located very near or far away from the protein surface. At small water–water distance, the values of order parameter of water are very low. The average order parameter records a maximum value at a distance equivalent to the first solvation layer with respect to the water–water radial distribution and asymptotically approaches a constant value at large distances. Results from present analysis will help to get a better insight into structure of hydration water around proteins. The analysis will also help to improve the accuracy of water models on the protein surface.

1. Introduction

Water in the protein interior has long been known to be crucial for the assembly and three-dimensional structure of proteins and nucleic acids. Particularly, protein–water interactions delineate and mould the free energy landscape that stabilize the structure of a protein by bridging between the protein hydrogen bond acceptors and donors and govern the folding of proteins by expelling core water in the rate-limiting step [1]. Water also participates in enzyme catalysis by facilitating conformational transitions and mediating proton transfer reactions [2]. Properties of hydration water surrounding biomolecules have been extensively studied both experimentally and computationally [3]. These studies have clearly established the distinction between the properties of water in the hydration layer surrounding the biomolecules and the bulk water [4–26]. Attempts to characterize the structure and dynamics of the hydration water on a protein surface are implemented through a number of techniques, which include X-ray crystallography [27], NMR [28,29], neutron scattering [30,31], dielectric relaxation [32–34], magnetic resonance dispersion [1], time-resolved fluorescence [5], solvation dynamics [35] and molecular simulations [6,10,36]. These results indicate that the structure, diffusion and the intramolecular vibration modes of water molecules in the hydration layer are distinctly different compared to bulk water. While most experimental tools access the average effects of the solute over the entire hydration layer, molecular dynamics provides valuable insight to a site-specific picture of the hydration dynamics leading to a better comprehension of the conformation and dynamics of water molecules around peptide segments.

This remarkable complexity of the structure and dynamics of hydration water is not completely understood through experiments or simulations. In experiments, it is often difficult to detect and characterize hydration water against the dominant background of bulk water [37–39,33,40,41]. Direct observation of hydration water in proteins is difficult except for crystal structures of atomic resolution better than 1 Å. Computationally this difference is reflected in a variety of different potential functions, such as TIP3P [42], TIP4P [43], TIP5P [44], SPC/E [45], etc., commonly used in molecular dynamics simulations, which are parametrized to reproduce the experimental data of bulk water [46]. Since the potential derived from bulk water may not accurately represent properties of hydration water, the differences between the properties of hydration water and bulk water need to be better defined in molecular dynamics simulation programs. MD simulations are also limited by the availability of computational resources. In such circumstances, crystal structures of proteins deposited in Protein Data Bank (PDB) [47] may provide an

* Corresponding author. Tel.: +91 11 27666646.
E-mail address: pbiswas@chemistry.du.ac.in (P. Biswas).

0301-4622/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2011.05.009

alternative way for investigating the hydration water structure. Direct measurement of visible hydration water has become possible with the availability of significant amount of high resolution protein crystal structures as well as protein structures elucidated by cryogenic X-ray diffraction techniques. Averaging over a number of crystal structures reduce the errors caused by the differences in crystallization conditions and other sources of experimental variations.

In this study, we have explored radial distribution pattern of water tetrahedral structure in the hydration shell of 188 high resolution proteins (resolution ≤ 1 Å). The results show that water molecules placed very near and very far away from the protein surface loose their tetrahedral geometry. Average tetrahedral order parameter of water molecules calculated from high resolution crystal structures of proteins shows an oscillating behavior at long distance from protein surface. The results are also found to be robust with respect to an additional database of protein structures elucidated by high-resolution cryogenic X-ray crystallography. Previous simulation results of peptide sequences [48] are found to overestimate the average tetrahedral order parameter at various distances from protein surface while simulation of α-lactalbumin in the present work shows slight overestimation of tetrahedral ordering of water molecules at long distances from the protein surface. Distribution of the tetrahedral order parameter of water at various distances from each water molecule also exhibit vast discrepancy when the results from the crystal structures are compared to those obtained from simulations. Apart from providing a better insight on water structure in hydration shell of proteins the results from present analysis may also help in improving the accuracy of water models used in simulations of proteins.

2. Materials and methods

2.1. High resolution proteins

A total of 188 high resolution protein crystal structures were selected from the Protein Data Bank (PDB) [47] with a cut-off atomic resolution of ≤1 Å. This database consists of 105 high resolution proteins used in a related work [46] to calculate the radial distribution of water molecules and also includes additional ones as listed in Appendix A. The hydration level (gram of water molecules per gram of protein) of these crystal structures range from 0.7 to 0.81. The temperature of isolation of these 188 crystal structure ranges from 85 K to 293 K.

Every water molecule has an occupancy index in the PDB file. Most of the water molecules in these crystal structures have a maximum occupancy factor of 1.0, implying the presence of these water molecules in the protein crystal. A few occupancy values are less than 1.0, which indicates that the respective water molecules are absent in some of the protein molecules. For this study the sum of the occupancy of each water molecule was considered rather than the actual number of water molecules. The selected proteins contain a total of 58919 water molecules including 9230 partially occupied waters.

2.2. Radial distribution

Water–protein radial distribution function describes the density of water as a function of the distance from the closest non-hydrogen atom on the protein surface. The water–protein radial distribution is calculated by dividing the number of water molecules present in different radial shells at various distances from the protein molecule by the volume of the corresponding shell [46]. The water–protein radial distribution function or surface distribution function (SDF) describes the density of water as a function of the distance h from the protein surface which is determined by calculating the distance of the water oxygen atom from the nearest non-hydrogen atom of the protein,

$$g_{\mathrm{wp}}(h) = \frac{\mathrm{d}\rho}{\mathrm{d}h} \tag{1}$$

where ρ is the average density of water around a protein at a distance h, given by

$$\rho = \frac{n}{\frac{4}{3}\pi\left(R_p + h\right)^3 - \frac{4}{3}\pi R_{\mathrm{p}}^3} \tag{2}$$

and n is the number of water molecules present within water–protein distance h and R_{p} is the radius of the protein molecule p. The radius of the protein, R_{p}, is defined as the distance between the center of the protein to the farthest non-hydrogen atom of the protein measured from its center.

The water–water radial distribution function (RDF) gives the density of water as a function of distance from a tagged water molecule. RDF in proteins is different from that of the bulk water due to the interaction of the water with protein molecules. In the presence of protein atoms, the volume occupied by water is restricted, and thus the procedure of normalizing density must be modified. The raw RDF is calculated by counting water molecules around the tagged water molecule at various distances

$$g_{\mathrm{raw}}(r_{\mathrm{w}}) = \frac{\mathrm{d}(\rho_{\mathrm{ww}})}{\mathrm{d}r_{\mathrm{w}}} \tag{3}$$

where $d(\rho_{ww})$ is the water density at a distance between r_w and $r_w - dr_w$ from the tagged water molecule w. The raw counts of water molecules within each radial shell were converted into an average value for the raw RDF. The normalization factor for each shell was calculated as

$$g_{\mathrm{nor}}(r_{\mathrm{w}}) = \frac{\mathrm{d}\left(\rho_{\mathrm{wp}}\right)}{\mathrm{d}r_{\mathrm{w}}} \tag{4}$$

where $\mathrm{d}(\rho_{\mathrm{wp}})$ is the average of the water density of each small cell in the shell between r_{w} and $r_{\mathrm{w}} - \mathrm{d}r_{\mathrm{w}}$, where density of each small cell is determined by its water–protein distance. Both raw RDF as well as the normalization factor were averaged over the number of water molecules in each crystal structure and the number of crystal structures.

$$\left\langle g_{\mathrm{raw}}\left(r_{\mathrm{wp}}\right)\right\rangle = \frac{1}{P}\sum_{p=1}^{P}\left(\frac{1}{W_p}\sum_{w=1}^{W_{\mathrm{p}}} g_{\mathrm{raw}}\left(r_{\mathrm{wp}}\right)\right) \tag{5}$$

$$\left\langle g_{\mathrm{nor}}\left(r_{\mathrm{wp}}\right)\right\rangle = \frac{1}{P}\sum_{p=1}^{P}\left(\frac{1}{W_{\mathrm{p}}}\sum_{w=1}^{W_{\mathrm{p}}} g_{\mathrm{nor}}\left(r_{\mathrm{wp}}\right)\right) \tag{6}$$

where P is the number of high resolution protein crystal structures chosen for this study i.e. $P = 188$, W_{p} is the number of water molecules in protein p and r_{wp} is the distance of the shell from the tagged water molecule in protein p. The normalized RDF was calculated as

$$g_{\mathrm{normalized}}\left(r_{\mathrm{wp}}\right) = \frac{\left\langle g_{\mathrm{raw}}\left(r_{\mathrm{wp}}\right)\right\rangle}{\left\langle g_{\mathrm{nor}}\left(r_{\mathrm{wp}}\right)\right\rangle} \tag{7}$$

3. Water structure

Orientational order parameter characterizes the local structure of hydrogen bonded liquid water. The orientational (or tetrahedral) order parameter q_{tet} measures the extent to which a given water

molecule and its four nearest neighbors adopt a tetrahedral arrangement. For pure water, orientational order parameter is defined as [49]

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

where $\Psi_{j,k}$ is the angle formed between the oxygen atom of a given water molecule and the bond vectors r_{ij} and r_{jk} of its four nearest neighbor atoms j and k of the same type. For a perfect tetrahedral arrangement, $q_{\text{tet}} = 1$. Conversely, for a random arrangement with uncorrelated uniform angular distributions, $q_{\text{tet}} = 0$. For water molecules surrounding the proteins, the orientational order parameter is calculated either (i) by considering only water molecules as four nearest neighbors or (ii) by considering heavy atoms of the peptide (C, N, O etc.) as nearest neighbors of the oxygen atom of the selected water molecule.

3.1. Molecular dynamics simulation

Molecular dynamics simulations of α-lactalbumin are used to explore the structure and properties of water around protein. MD simulation was performed by the AMBER 9 package [50]. The crystal structure was extracted from the PDB (id: 1A4V.pdb) and the missing hydrogen atoms were added with the LEAP subroutine. The crystallographic water molecules were striped off initially and then the protein was solvated in a cubic box with TIP3P water molecules [42]. The buffering distance between edge of the box and the protein was maintained at 10 Å during solvation and the closeness parameter was set at 1 Å. The overall charge of the system was neutralized by adding 7 Na + ions. The ff99SB force field was used with the periodic boundary conditions. This force field represents a better balance of four basic secondary structure elements (PP_{II}, β, α_L and α_R) by incorporating an improved distribution of the backbone dihedrals [51]. This force field also displays the best agreement with experimental data [52]. The nonbonded cutoff was set to be 8 Å. Particle Mesh Ewald (PME) summation was used to treat the long range electrostatic interactions [53] with the nonbonded real space cutoff set at 8.0 Å and fourth order spline interpolation with grid spacing of 1.0 Å. The SHAKE algorithm was used to constrain all bond lengths to their equilibrium distances [54]. The energy of the system was minimized twice by the conjugate gradient method, the solvent was energy minimized first by keeping the protein constrained followed by the energy minimization of the whole system. A two-step equilibration was performed by simulating the solvated protein in a NVT ensemble at an initial temperature of 100 K, which was gradually raised up to 300 K for 50 ps at a constant volume. This was followed by a NPT equilibration for 300 ps at a constant temperature of 300 K and a pressure of 1 bar. Constant temperature was maintained through weak coupling to Berendsen thermostat with a coupling constant of 2 ps while constant pressure was maintained through weak coupling to isotropic barostat with a coupling constant of 1 ps [55]. The production run was executed with the NPT system for 10 ns with a time step of 2 fs. Snapshots of the trajectory were saved at each picosecond. All properties were extracted from the trajectory of the final production run. Previous studies [46] on water–water radial distribution function in protein crystal structures using similar database compares the results with that of the bulk water at room temperature [56]. However, for low temperature simulations, water is neither frozen nor glassy but remains supercooled [57–60] and the Stokes–Einstein relation is valid [61]. Hence it is expected that the conclusions drawn from the simulation results (as analyzed in Results and discussion section) may have negligible effect on the temperature at which the simulation is performed.

3.2. 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 or water–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 width, Δr is chosen to be 1 Å, 0.5 Å and 0.1 Å. In the main text, we display the results for $\Delta r = 1$ Å, while the results for other values of Δr are provided in the Supplementary data. At each r, the fraction of water molecules with q_{tet}-values ranging between $q_{\text{tet}} - dq_{\text{tet}}/2$ and $q_{\text{tet}} + dq_{\text{tet}}/2$ is denoted as $P(q_{\text{tet}})dq_{\text{tet}}$ where $dq_{\text{tet}} = 0.05$. Average value of q_{tet}, $\langle q_{\text{tet}}\rangle$, is also plotted at each r.

4. Results and discussion

4.1. Water structure as a function of water–protein distance

188 high resolution (≤1 Å) protein crystal structures are compiled from the PDB for analyzing the distribution of tetrahedral order parameter of water molecules as a function of distance from the protein surface. Fig. 1(a) and (b) show the distribution of tetrahedral order parameter of water molecules, $P(q_{\text{tet}})$, as a function of distance from surface of the protein molecule. In Fig. 1(a) q_{tet} values are calculated by including protein heavy atoms as four nearest neighbors, while in Fig. 1(b) only water molecules are considered as nearest neighbors. Insets of Fig. 1(a) and (b) depict the average distribution of tetrahedral order of water molecules $<q_{\text{tet}}>$ at different distances from the protein surface. Water molecules which are very close or far off from the protein surface are found to be less tetrahedrally ordered compared to the water molecules situated at any intermediate distance. This result is more pronounced for smaller Δr values (Supplementary Figs. 3, 4, 5 and 6). To verify robustness of these results an additional database of protein structures elucidated by high-resolution cryogenic X-ray crystallography is also considered. There are 11964 protein structures in this database whose structures are refined using diffraction data collected bellow 150 K and at resolution better than 2.2 Å. The structures contain 3920740 water molecules including 110811 partially occupied waters. Supplementary Figs. 11 to 16 show the distribution of $P(q_{\text{tet}})$ values at various distances from protein surface in this additional database. The pattern of results is found to be similar to that shown in the main manuscript.

Earlier results from molecular dynamics simulation of small peptides have shown less ordered water molecules near protein surface [48]. The results of two-particle approximation for calculating the change in entropy of water in the vicinity of the peptide also confirm that the water molecules near the protein are less ordered than those in bulk [62]. A possible explanation for the loss of tetrahedral structure of water molecules, which are close to the protein surface may be explained in terms of the optimum distance of hydrogen bonding. Any protein atom located closer than the optimum H-bond distance may lead to the distortion of water molecules from their tetrahedral arrangements. Results from the molecular dynamics simulation of collagen triple helix also show that the number of hydrogen bonds formed by each water molecule near the peptide is less than four, leading to the disruption of its required tetrahedral geometry [63].

The order parameter, q_{tet}, calculated by including and excluding protein heavy atoms (i.e. Fig. 1(a) and (b) respectively) are qualitatively similar, except for the differences at close and intermediate distances from the protein atoms. Higher tetrahedral order is observed at closer distances when the heavy atoms of the protein are excluded from the calculation, while an opposite trend is observed at intermediate distances. From the inset of Fig. 1(a) and (b) it is found that the peak in $\langle q_{\text{tet}}\rangle$ is shifted from 4 Å to 5 Å when protein heavy atoms are not considered. Both these peaks are at a distance beyond

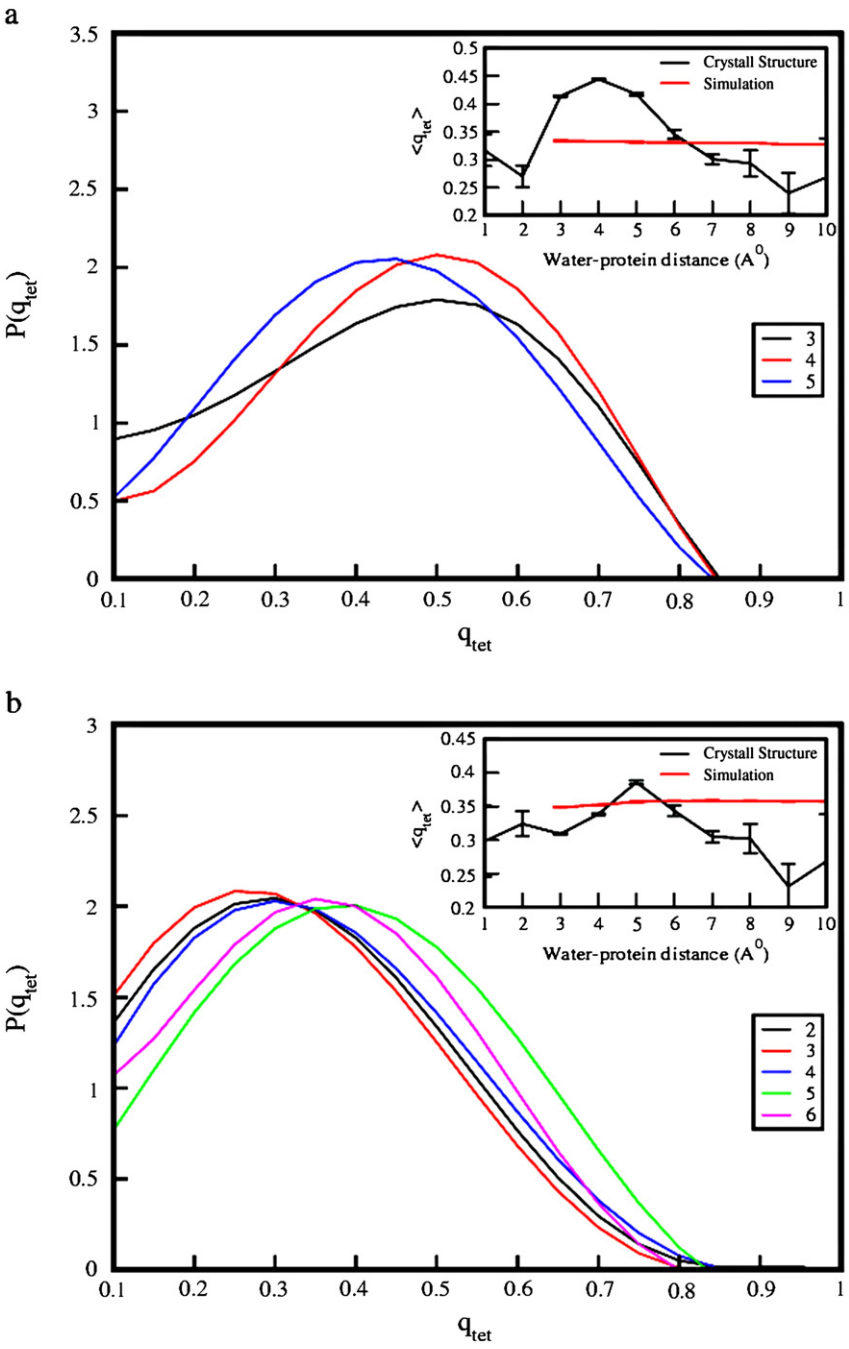

Fig. 1. Distribution of the orientational (tetrahedral) order parameter, $P(q_{tet})$, at various water–protein distances in high resolution protein structures with $\Delta r = 1$ Å. Order parameters are calculated (a) including and (b) excluding protein heavy atoms. Distributions at distances with significant 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–protein distances in high resolution crystal structures and from trajectory of α-lactalbumin simulation. The standard errors are plotted as error bars.

the first and second solvation shell as depicted in Supplementary Fig. 1. The position of this peak is similar to that of small peptides simulated using different solvent models [48]. Previous simulation results on antifreeze proteins [64] exhibit a similar peak for the local tetrahedrality of water at a position similar to that of the present work even though a different parameter was used to define the tetrahedral ordering of water molecules. The plot of $\langle q_{tet} \rangle$ calculated excluding protein heavy atoms (inset of Fig. 1(b)) shows a small hump near 2 Å from the protein surface which is absent in the case where the protein heavy atoms are included.

The distribution of tetrahedral order parameter, $P(q_{tet})$, in the limit of small (Δr) values are depicted in the Supplementary Figs. 3, 4, 5 and 6. At a distance beyond 6 Å from the protein surface, the average tetrahedral order parameter, $\langle q_{tet} \rangle$, shows an oscillatory behavior. This observation is also supported by the MD simulation of small peptides where the oscillations are found to be more pronounced. This oscillatory behavior is due to the fact that $\langle q_{tet} \rangle$ is anti-correlated with average tagged potential energy which in turn is found to be oscillatory in nature up to a distance of 10 Å from the protein surface [48]. MD simulation of collagen triple helix with TIP3P water model reveals the existence of alternating hydration layer of lower and higher water density around the collagen molecule [63]. This may account for an oscillating behavior of $\langle q_{tet} \rangle$ at various distances from the protein surface.

The $\langle q_{tet} \rangle$ values calculated from the high resolution crystal structures are found to be lower than those calculated from the earlier MD simulation results [48]. For the sake of comparison, $\langle q_{tet} \rangle$ values at various distances from the protein surface calculated from 10 ns MD simulation trajectory of α-lactalbumin are also plotted in the insets of Figs. 1(a) and (b). The results show that at long distances from the protein surface $\langle q_{tet} \rangle$ show higher values for simulation trajectory as compared to those obtained from crystal structures when protein heavy atoms are included in calculating q_{tet} (Fig. 1(a)). When protein heavy atoms are excluded, (Fig. 1(b)), $\langle q_{tet} \rangle$ values are consistently higher for simulation except for intermediate protein–water distance. Earlier MD simulation studies [64,65] confirm that the tetrahedral ordering of water molecules in proteins increases with the decrease in temperature. Our simulation results at room temperature establish the overestimation of the tetrahedral order parameter of water in proteins as compared to that calculated from the high resolution crystal structures. Since the tetrahedral order of water increases with the decrease of temperature, the qualitative trend will remain the same but only the degree of overestimation will increase with the decrease of temperature.

A comparison of the radial distribution function of bound water in protein crystals with that of the bulk water from experiment results [56] shows that the density of hydration shell water molecules in protein crystal structures is slightly higher in value [46] compared to the bulk water. This higher density of water in protein crystals may lead to the crowding of water molecules resulting in deviation from the preferred tetrahedral geometry. Molecular dynamics simulation of lysozyme in TIP3P water model with NPT ensemble also records an increase of the density of water in the first hydration shell as a result of the geometric constraints which is a primary contributing factor for unperturbed bulk water [66]. Since water models used in simulations are parametrized to reproduce the experimental data of bulk water, the density profile of hydration water is quite similar to that of bulk water with slightly higher $\langle q_{tet} \rangle$. A noticeable difference of these results compared to those obtained from MD simulations is reflected in the decrease of the average tetrahedral order parameter, $\langle q_{tet} \rangle$, at distances far off from the protein surface rather than approaching a constant value. This may be explained by the fact that in MD simulations, the hydration shell water primarily exhibits bulk properties at large distances from the protein surface and $\langle q_{tet} \rangle$ asymptotically approaches a constant value. However, the values of $\langle q_{tet} \rangle$, calculated from the high resolution protein crystal structures at distances far off from the protein surface, decrease due to a decrease in the density of the hydration shell water molecules exhibiting a considerable deviation from its tetrahedral geometry, as shown in Supplementary Fig. 1.

4.2. Water structure as a function of water–water distance

The distribution of the tetrahedral order parameter of water in proteins is also explored as a function of the water–water distance. Analogous to the protein–water distance, the tetrahedral order parameter of water molecules are binned at various water–water distances with different values of Δr i.e. 1 Å, 0.5 Å and 0.1 Å. The distribution of the tetrahedral order parameter of water and the average tetrahedral order parameter, $\langle q_{tet} \rangle$, of each bin are calculated. Fig. 2(a) and (b) depict the distribution of water molecules having different values of q_{tet} with respect to the distance from each water molecule in presence and absence of the protein heavy atoms as nearest neighbors. The figure insets show $\langle q_{tet} \rangle$ as a function of the distance from each water molecule. The results show that a water molecule tends to be less ordered when it is very close to another water molecule. Beyond a distance equivalent to the first water–water solvation layer, the water molecules are found to be more ordered with the values of q_{tet} showing similar distribution patterns and $\langle q_{tet} \rangle$ approaching constant values of 0.43 and 0.33 in insets of Fig. 2(a) and (b) respectively. The average tetrahedral order parameter, $\langle q_{tet} \rangle$, has a maximum value at a distance of nearly 3 Å (inset of Fig. 2) which represents the peak value in the first solvation layer (Supplementary Fig. 2). Since this distance is lower than the optimum H-bond distance, it may lead to the distortion of the water structure from its preferred tetrahedral geometry. This may account for the decrease of $\langle q_{tet} \rangle$ at a distance less than the first water–water solvation layer. A small hump is also observed in $\langle q_{tet} \rangle$ at a distance of the second solvation shell. The $\langle q_{tet} \rangle$ values also display an oscillating behavior at large water–water distances as shown in Supplementary Figs. 7, 8, 9 and 10 even though the amplitude of oscillation is very low. Fig. 2(a) and (b) depict the results calculated by including and excluding the protein heavy atoms as four nearest neighbors respectively differ in the position of the maximum in the distribution of water molecules with different values of q_{tet}. The plot shows lower values of q_{tet} when the protein heavy atoms are excluded. The results validate that most of the water molecules have ordered tetrahedral arrangement with respect to the protein heavy atoms. This is also emphasized in the insets of the Fig. 2 (a) and (b) where the values of $\langle q_{tet} \rangle$ turns out to be consistently higher when protein heavy atoms are included.

A comparison with the simulation results is also performed by plotting $\langle q_{tet} \rangle$ at various distances from water molecules calculated from the 10 ns simulation trajectory of α-lactalbumin shown in the insets of Fig. 2(a) and (b). The values of $\langle q_{tet} \rangle$ are found to be higher in crystal structures compared to the simulation results when the protein heavy atoms are included in calculation of q_{tet}; while the trend is exactly opposite when protein heavy atoms are excluded. A closer look at the figures shows that $\langle q_{tet} \rangle$ values do not differ much from the simulation results when the protein heavy atoms are either included or excluded from calculations. The difference in the trends of insets of Fig. 2(a) and (b) is mostly due to different $\langle q_{tet} \rangle$ values calculated from crystal structures. A similar trend is found in the insets of Fig. 1(a) and (b) where the $\langle q_{tet} \rangle$ values from simulation results vary minimally with protein–water distance. This confirms the need of improved water models to accurately portray the fluctuations in q_{tet} values of crystal structures at various distances from protein surface and water molecules.

5. Conclusions

The present study deals with the distribution of water tetrahedral/orientational order parameter as a function of protein–water and water–water distances in protein crystal structures. The work is done with 188 high resolution (≤1Å) protein crystal structures from Protein Data Bank (PDB). It is found that water molecules are more distorted from their tetrahedral geometry when present in the vicinity or far off from the protein surface. At distances close to the protein surface, the value of q_{tet} is low due to distances less than optimum H-bonding while at larger distances from the protein surface there are less number of water molecules. Previous peptide simulation results [48] as well as present simulation results of α-lactalbumin using solvent models based on experimental bulk water potentials are found to overestimate tetrahedral order parameters of water molecules in proteins. However, the average tetrahedral order parameter calculated from protein crystal structures shows similar oscillating pattern at long distance from protein surface as suggested previously by MD simulation results [48]. Similar study with an additional database of cryogenic proteins also validates the robustness of the results from high resolution proteins. Present analysis may help in improving the accuracy of water models near protein surface to be used for future simulations.

Acknowledgments

The authors gratefully acknowledge the financial assistance from Department of Science and Technology (project no. SR/S1/PC-07/06),

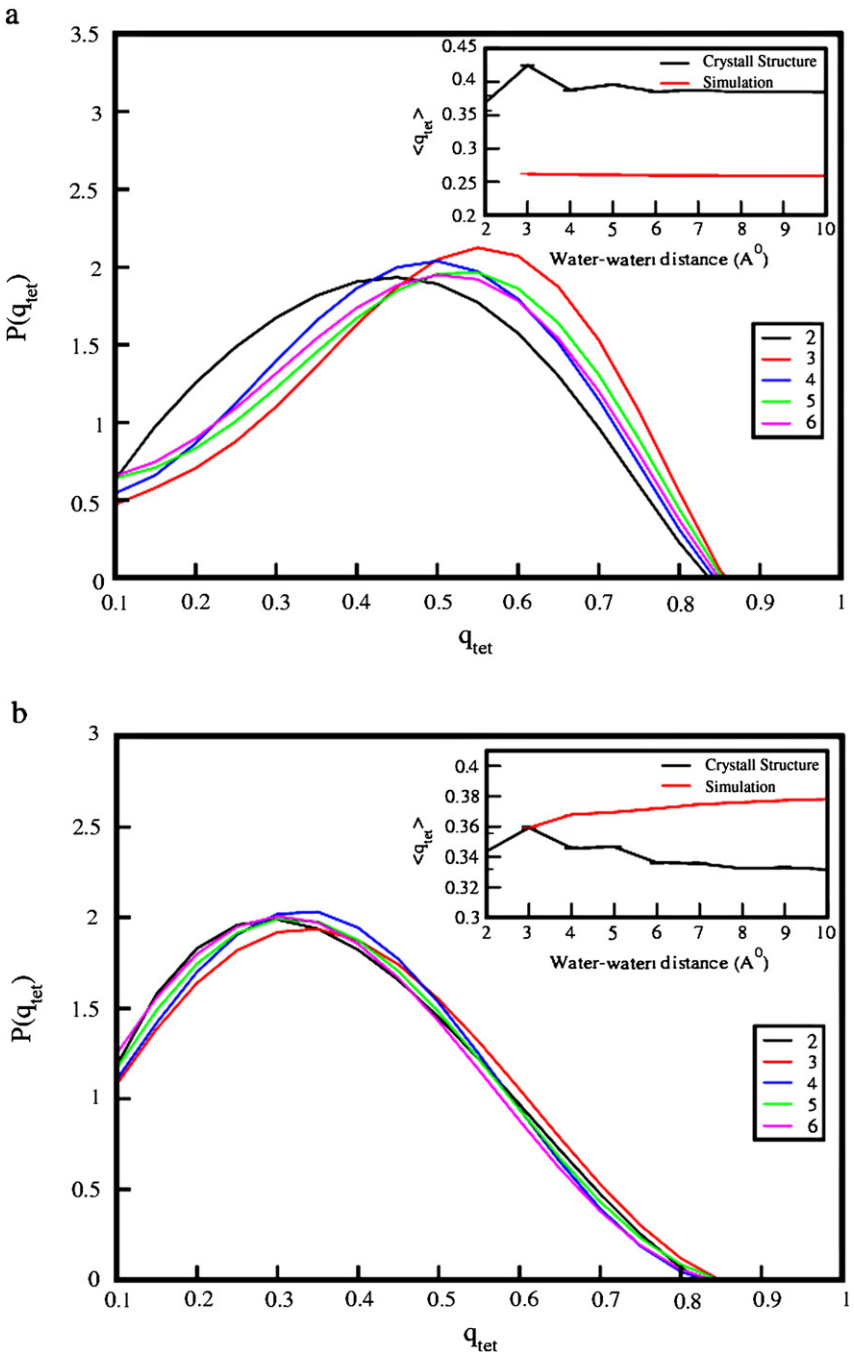

Fig. 2. Distribution of the orientational (tetrahedral) order parameter, $P(q_{tet})$, at various water–water distances in high resolution protein structures with $\Delta r = 1$ Å. Order parameters are calculated (a) including and (b) excluding protein heavy atoms. Distributions at distances with significant 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–water distances in high resolution crystal structures and from trajectory of α-lactalbumin simulation. The standard errors accumulated while averaging are plotted as error bars.

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.

Appendix A. PDB IDs of 188 high resolution protein crystals used in the present work

1A6M, 1AHO, 1B0Y, 1BRF, 1BXO, 1BYI, 1C75, 1C7K, 1CEQ, 1CEX, 1DY5, 1EA7, 1EB6, 1ETL, 1ETM, 1ETN, 1EXR, 1F94, 1G4I, 1G66, 1G6X, 1GA6, 1GCI, 1GKM, 1GQV, 1GWE, 1HJ8, 1HJ9, 1HJE, 1I1W, 1IC6, 1IEE, 1IQZ, 1IR0, 1IUA, 1IX9, 1J0P, 1JFB, 1K2A, 1K4I, 1K4P, 1K5C, 1K6U, 1KCC, 1KTH, 1KWF, 1L9L, 1LKK, 1LNI, 1LUG, 1M1Q, 1M40, 1MC2, 1MJ5, 1MN8, 1MNZ, 1MUW, 1MWQ, 1MXT, 1N1P, 1N4U, 1N4V, 1N4W, 1N55, 1N9B, 1NKI, 1NLS, 1NQJ, 1NWZ, 1O7J, 1OAI, 1OB4, 1OB7, 1OD3, 1OEW, 1OK0, 1OT6, 1OT9, 1P1X, 1P9G, 1PJX, 1PQ5, 1PQ7, 1PWM, 1Q6Z, 1R6J, 1RB9, 1RTQ, 1S5M, 1S5N, 1SSX, 1SY2, 1SY3, 1TQG, 1TT8, 1U2H, 1UCS, 1UFY, 1UG6, 1UNQ, 1US0, 1V0L, 1V6P, 1VB0, 1VBW, 1VL9, 1VYR, 1W0N, 1X6Z, 1X8P, 1X8Q, 1XG0, 1XMK, 1ZK4, 1ZUU, 1ZZK, 2A6Z, 2B97, 2BF6, 2BF9, 2BT9, 2CE2, 2CHH, 2CNQ, 2CWS, 2DDX, 2DSX, 2E4T, 2ERL, 2F01, 2FDN, 2FMA, 2FVY, 2FWH, 2GGC,

2GKG, 2GUD, 2H5C, 2HS1, 2JFR, 2JHF, 2NRL, 2O7A, 2O9S, 2OV0, 2P5K, 2PND, 2PNE, 2PPN, 2PVB, 2QCP, 2QSK, 2R31, 2RBK, 2RH2, 2V1M, 2V8T, 2VB1, 2VHA, 2VHK, 2VXN, 2WFI, 2ZPM, 3A39, 3BWH, 3CCD, 3D1P, 3DHA, 3DK9, 3E4G, 3E7R, 3EA6, 3EO6, 3F1L, 3F7L, 3FIL, 3FSA, 3G21, 3G46, 3G63, 3GOE, 3IP0, 3LZT, 3PYP, 4LZT, 7A3H, 8A3H, 8RXN.

Appendix B. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.bpc.2011.05.009.

References

[1] B. Halle, Protein hydration dynamics in solution: a critical survey, Philos. Trans. R. Soc. London, Ser. B 359 (2004) 1207–1224.
[2] J.C. Smith, F. Merzel, A.-N. Bondar, A. Tournier, S. Fischer, Structure, dynamics and reactions of protein hydration water, Philos. Trans. R. Soc. London, Ser. B 359 (2004) 1181–1190.
[3] P. Ball, Water as an active constituent in cell biology, Chem. Rev. 108 (2008) 74–108.
[4] S.K. Pal, J. Peon, A.H. Zewail, Biological water at the protein surface: Dynamical solvation probed directly with femtosecond resolution, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 1763–1768.
[5] B. Bagchi, Water dynamics in the hydration layer around proteins and micelles, Chem. Rev. 105 (2005) 3197–3219.
[6] D.M. Leitner, J.E. Straub (Eds.), Proteins: Energy, Heat and Signal Flow, Taylor and Francis, New York, 2009.
[7] H.E. Stanley, P. Kumar, S. Han, M.G. Mazza, K. Stokely, S.V. Buldyrev, G. Franzese, F. Mallamace, L. Xu, Heterogeneities in confined water and protein hydration water, J. Phys. Condens. Matter 21 (2009) 504105.
[8] A.R. Bizzarri, S. Cannistraro, Molecular dynamics of water at the protein–solvent interface, J. Phys. Chem. B 106 (2002) 6617–6633.
[9] M.E. Johnson, C. Malardier-Jugroot, R.K. Murarka, T. Head-Gordon, Hydration water dynamics near biological interfaces, J. Chem. Phys. B 113 (2009) 4082–4092.
[10] L. Hua, X. Huang, R. Zhou, B.J. Berne, Dynamics of water confined in the interdomain region of a multidomain protein, J. Chem. Phys. B 110 (2006) 3704–3711.
[11] S. Wiesner, E. Kurian, F.G. Prendergast, B. Halle, Water molecules in the binding cavity of intestinal fatty acid binding protein: dynamic characterization by Water ^{17}O and ^{2}H magnetic relaxation dispersion, J. Mol. Biol. 286 (1999) 233–246.
[12] D.M. Leitnera, M. Havenithb, M. Gruebele, Biomolecule large-amplitude motion and solvation dynamics: modelling and probes from THz to X-rays, Int. Rev. Phys. Chem. 25 (2006) 553–582.
[13] D.S. Grebenkov, Y.A. Goddard, G. Diakova, J.-P. Korb, R.G. Bryant, Dimensionality of diffusive exploration at the protein interface in solution, J. Chem. Phys. B 113 (2009) 13347–13356.
[14] F. Pizzitutti, M. Marchi, F. Sterpone, P.J. Rossky, How protein surfaces induce anomalous dynamics of hydration water, J. Chem. Phys. B 111 (2007) 7584–7590.
[15] K.E. Furse, S.A. Corcelli, Molecular dynamics simulations of DNA solvation dynamics, J. Phys. Chem. Lett. 1 (2010) 1813–1820.
[16] K.E. Furse, S.A. Corcelli, Effects of an unnatural base pair replacement on the structure and dynamics of DNA and neighboring water and ions, J. Chem. Phys. B 114 (2010) 9934–9945.
[17] I. Yu, T. Tasaki, K. Nakada, M. Nagaoka, Influence of hydrostatic pressure on dynamics and spatial distribution of protein partial molar volume: time-resolved surficial Kirkwood–Buff approach, J. Phys. Chem. B 114 (2010) 12392–12397.
[18] I. Yu, M. Nagaoka, Slowdown of water diffusion around protein in aqueous solution with ectoine, Chem. Phys. Lett. 388 (2004) 316–321.
[19] L. Mitra, N. Smolin, R. Ravindra, C. Royer, R. Winter, Pressure perturbation calorimetric studies of the solvation properties and the thermal unfolding of proteins in solution—experiments and theoretical interpretation, Phys. Chem. Chem. Phys. 8 (2006) 1249–1265.
[20] D.N. LeBard, D.V. Matyushov, Ferroelectric hydration shells around proteins: electrostatics of the protein water interface, J. Phys. Chem. B 114 (2010) 9246–9258.
[21] V.M. Dadarlat, C.B. Post, Decomposition of protein experimental compressibility into intrinsic and hydration shell contributions, Biophys. J. 91 (2006) 4544–4554.
[22] X. Yu, J. Park, D.M. Leitner, Thermodynamics of protein hydration computed by molecular dynamics and normal modes, J. Phys. Chem. B 107 (2003) 12820–12828.
[23] T.V. Chalikian, Volumetric properties of proteins, Annu. Rev. Biophys. Biomol. Struct. 32 (2003) 207–235.
[24] A. Lervik, F. Bresme, S. Kjelstrup, D. Bedeaux, J.M. Rubi, Heat transfer in protein water interfaces, Phys. Chem. Chem. Phys. 12 (2010) 1610–1617.
[25] C. Carey, Y.-K. Cheng, P.J. Rossky, Hydration structure of the α–chymotrypsin substrate binding pocket: the impact of constrained geometry, Chem. Phys. 258 (2000) 415–425.
[26] H. Yu, S.W. Rick, Free energies and entropies of water molecules at the inhibitor–protein interface of DNA gyrase, J. Am. Chem. Soc. 131 (2009) 6608–6613.
[27] R. Huber, D. Kukla, W. Bode, P. Schwager, K. Bartels, J. Deisenhofer, W. Steigemann, Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor: II. Crystallographic refinement at 1.9 *A* resolution, J. Mol. Biol. 89 (1974) 73–101.
[28] G. Otting, E. Liepinsh, K. Wuthrich, Protein hydration in aqueous solution, Science 254 (1991) 974–980.
[29] K. Modig, E. Liepinsh, G. Otting, B. Halle, Dynamics of protein and peptide hydration, J. Am. Chem. Soc. 126 (2004) 102–114.
[30] M.C. Bellissent-Funel, Hydration in protein dynamics and function, J. Mol. Liq. 84 (2000) 39–52.
[31] S. Dellerue, M.-C. Bellissent-Funel, Relaxational dynamics of water molecules at protein surface, Chem. Phys. 258 (2000) 315–325.
[32] S.K. Pal, J. Peon, B. Bagchi, A.H. Zewail, Biological water: femtosecond dynamics of macromolecular hydration, J. Phys. Chem. B 126 (2002) 12376–12395.
[33] V.P. Denisov, B. Halle, Protein hydration dynamics in aqueous solution, Faraday Discuss. 103 (1996) 227–244.
[34] N. Nandi, K. Bhattacharyya, B. Bagchi, Dielectric relaxation and solvation dynamics of water in complex chemical and biological systems, Chem. Rev. 100 (2000) 2013–2046.
[35] E.H. Grant, R.J. Sheppard, G.P. South, Dielectric Behaviour of Biological Molecules in Solution, Oxford, U.K., 1978.
[36] V.A. Makarov, M. Feig, B.K. Andrews, B.M. Pettitt, Diffusion of solvent around biomolecular solutes: a molecular dynamics simulation study, Biophys. J. 75 (1998) 150–158.
[37] A.E. Eriksson, W.A. Baase, X.J. Zhang, D.W. Heinz, M. Blaber, E.P. Baldwin, B.W. Matthews, Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect, Science 255 (1992) 178–183.
[38] G. Otting, K. Wuethrich, Studies of protein hydration in aqueous solution by direct NMR observation of individual protein-bound water molecules, J. Am. Chem. Soc. 111 (1989) 1871–1875.
[39] G.M. Clore, A. Bax, P.T. Wingfield, A.M. Gronenborn, Identification and localization of bound internal water in the solution structure of interleukin 1.beta. by heteronuclear three-dimensional proton rotating-fram Overhauser nitrogen-15-proton multiple quantum coherence NMR spectroscopy, Biochemistry 29 (1990) 5671–5676.
[40] G. Otting, E. Liepinsh, B. Halle, U. Frey, NMR identification of hydrophobic cavities with ow water occupancies in protein structures using small gas molecules, Nat. Struct. Biol. 4 (1997) 396–404.
[41] B.W. Matthews, L. Liu, A review about nothing: are apolar cavities in proteins really empty? Protein Sci. 18 (2009) 494–502.
[42] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.
[43] W. Jorgensen, J. Madura, Temperature and size dependence for Monte Carlo simulations of TIP4P water, Mol. Phys. 56 (1985) 1381–1392.
[44] M.W. Mahoney, W.L. Jorgensen, A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions, J. Chem. Phys. 112 (2000) 8910–8922.
[45] H. Berendsen, J. Postma, W. van Gunsteren, J. Hermans, Intermolecular Forces, Reidel, Dordrecht, 1981.
[46] X. Chen, I. Weber, R.W. Harrison, Hydration water and bulk water in proteins have distinct properties in radial distributions calculated from 105 atomic resolution crystal structures, J. Phys. Chem. B 112 (2008) 12073–12080.
[47] F.C. Bernstein, T.F. Koetzle, G.J. Williams, E.F. Meyer Jr., M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, The protein data bank: a computer-based archival file for macromolecular structures, J. Mol. Biol. 112 (1977) 535–542.
[48] M. Agarwal, H.R. Kushwaha, C. Chakravarty, Local order, energy, and mobility of water molecules in the hydration shell of small peptides, J. Phys. Chem. B 114 (2010) 651–659.
[49] J.R. Errington, P.G. Debenedetti, Relationship between structural order and the anomalies of liquid water, Nature 409 (2001) 318–321.
[50] D.A. Case, T.A. Darden, T.E. Cheatham III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, D.A. Pearlman, M. Crowley, R.C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K.F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D.H. Mathews, C. Schafmeister, W.S. Ross, P.A. Kollman, AMBER 9, University of California, San Francisco, 2006.
[51] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of multiple amber force fields and development of improved protein backbone parameters, Proteins 65 (2006) 712–725.
[52] R.B. Best, G. Hummer, Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides, J. Phys. Chem. B 113 (2009) 9004–9015.
[53] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: an N.log(N) method for Ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089–11009.
[54] J.-P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, J. Comput. Phys. 23 (1977) 327–341.
[55] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684–3690.
[56] A.K. Soper, The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa, Chem. Phys. 258 (2000) 121–137.
[57] M. Matsumoto, S. Saito, I. Ohmine, Molecular dynamics simulation of the ice nucleation and growth process leading to water freezing, Nature 416 (2002) 409–413.

[58] I. Brovchenko, A. Geiger, A. Oleinikova, Liquid–liquid phase transitions in supercooled water studied by computer simulations of various water models, J. Chem. Phys. 123 (2005) 0445151-16.
[59] V. Velikov, S. Borick, C.A. Angell, The glass transition of water, based on hyperquenching experiments, Science 294 (2001) 2335–2338.
[60] C.A. Angell, Amorphous water, Annu. Rev. Phys. Chem. 55 (2004) 559–583.
[61] P. Kumar, S.V. Buldyrev, S.R. Becker, P.H. Poole, F.W. Starr, H.E. Stanley, Relation between the Widom line and the breakdown of the Stokes Einstein relation in supercooled water, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 9575–9579.
[62] D. Czapiewski, J. Zielkiewicz, Structural properties of hydration shell around various conformations of simple polypeptides, J. Phys. Chem. B 114 (2010) 4536–4550.
[63] J.-W. Handgraaf, F. Zerbetto, Molecular dynamics study of onset of water gelation around the collagen triple helix, Proteins 64 (2006) 711–718.
[64] D.R. Nutt, J.C. Smith, Dual function of the hydration layer around an antifreeze protein revealed by atomistic molecular dynamics simulations, J. Am. Chem. Soc. 130 (2008) 13066–13073.
[65] N. Smolin, V. Daggett, Formation of ice-like water structure on the surface of an antifreeze protein, J. Chem. Phys. B 112 (2008) 6193–6202.
[66] F. Merzel, J.C. Smith, Is the first hydration shell of lysozyme of higher density than bulk water? Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 5378–5383.