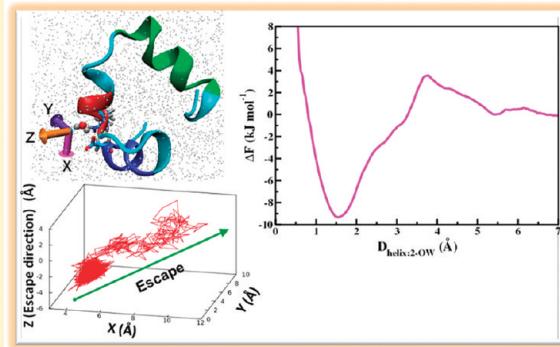


# Free Energy Barriers for Escape of Water Molecules from Protein Hydration Layer

Susmita Roy and Biman Bagchi\*

Solid State and Structural Chemistry Unit, Indian Institute of Science, Bangalore 560012, India

**ABSTRACT:** Free energy barriers separating interfacial water molecules from the hydration layer at the surface of a protein to the bulk are obtained by using the umbrella sampling method of free energy calculation. We consider hydration layer of chicken villin head piece (HP-36) which has been studied extensively by molecular dynamics simulations. The free energy calculations reveal *a strong sensitivity to the secondary structure*. In particular, we find a region near the junction of first and second helix that contains a cluster of water molecules which are slow in motion, characterized by long residence times (of the order of 100 ps or more) and separated by a large free energy barrier from the bulk water. However, these “slow” water molecules constitute only about 5–10% of the total number of hydration layer water molecules. Nevertheless, they play an important role in stabilizing the protein conformation. Water molecules near the third helix (which is the important helix for biological function) are enthalpically least stable and exhibit the fastest dynamics. Interestingly, barrier height distributions of interfacial water are quite broad for water surrounding all the three helices (and the three coils), with the smallest barriers found for those near the helix-3. For the quasi-bound water molecules near the first and second helices, we use well-known Kramers’ theory to estimate the residence time from the free energy surface, by estimating the friction along the reaction coordinate from the diffusion coefficient by using Einstein relation. The agreement found is satisfactory. We discuss the possible biological function of these slow, quasi-bound (but transient) water molecules on the surface.



## 1. INTRODUCTION

In many respects, water appears to be the “fastest” solvent studied so far, with time scales of rotational and translational motions in neat water ranging from a few tens of femtoseconds to a few tens of picoseconds.<sup>1–5</sup> However, the usual ultrafast dynamics of water can become relatively slower in confined systems (like reverse micelles and nanotubes), on the surface of interacting surfaces like silica, in the grooves of DNA, and on protein surfaces.<sup>4–10</sup> This simple fact has spurred a tremendous research effort aimed at elucidating the dynamics of water in and around proteins in aqueous solution. Several experimental and theoretical studies have reported that the dynamics of the protein hydration layer is *heterogeneous with a wide range of time scales*.<sup>11–14</sup> Hence to understand the origin, it has become essential to characterize different species (water) in the hydration layer that contribute to different time scale regions. Simulations show that depending on the number and nature of hydrogen bonds (H-bond) that water molecules make with the charged/polar groups on the protein surface, the water molecules at the interface can be divided broadly into two classes: (i) interfacial quasi-bound (IQBW) and (ii) interfacial free (IFW) water. Interfacial free water (IFW) molecules do not form any hydrogen bonds with the protein residues whereas interfacial quasi-bound water molecules can be either singly (IBW1) or doubly hydrogen bonded (IBW2). Interfacial free water molecules (IFW) of course form hydrogen bonds with neighboring water molecules and experience van der Waals type interactions

with protein atoms if they are within the range of interaction potential. Thus, these water molecules are by no mean free just as the IBW water molecule is not permanently bound. These nomenclatures have been introduced to characterize different states of water molecules on the protein surface.

In addition, a hydrophobic environment around interfacial water molecules may induce more ordered arrangement than that existing in the bulk.<sup>1–5</sup> Since nearly 50% of the residues on the surface are hydrophobic, there are a large number of water molecules in the hydration layer that exhibit dynamics almost as fast as bulk water. However, a quantification of their dynamics in terms of free energy has never been obtained. Similarly, only a semiquantitative understanding of the free energy surface of quasi-bound water molecules, obtained from population distribution, exists.

It has been shown recently that quasi-bound water molecules play a significant role in stabilizing a half-open-half-closed (HOHC) state of adenylyl kinase (ADK) in the catalytic conversion of an adenosine triphosphate (ATP) and an adenosine monophosphate (AMP) to two adenosine diphosphate (ADP) molecules.<sup>15</sup> In general, a few strongly quasi-bound water molecules might play a role in stabilizing certain secondary structures. In addition, these quasi-bound water molecules may also play a crucial role

Received: September 30, 2011

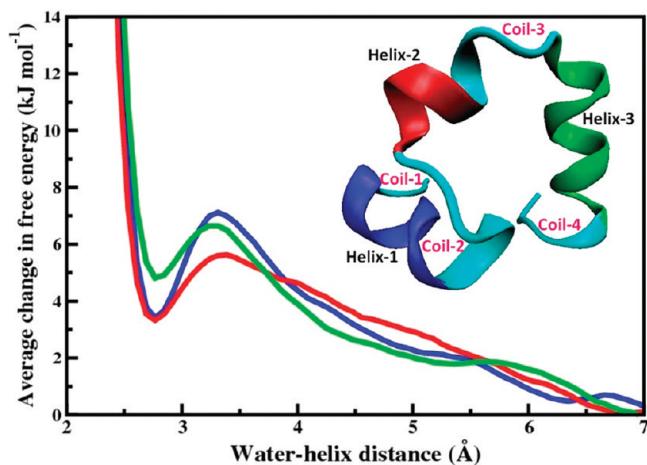
Revised: January 1, 2012

Published: January 30, 2012

in noncovalent association of proteins and small drug compounds, as demonstrated recently.<sup>16,17</sup>

Several studies have been carried out to characterize localized water molecules surrounding the protein surface.<sup>18,19</sup> To understand the hydrophilic nature of protein cavities, Zhang and Hermans used molecular dynamics free energy simulations to calculate the free energy cost for introducing a water molecule into the protein cavities.<sup>20</sup> Subsequently, Roux et al. applied molecular dynamics free energy perturbation methods to calculate the stability of water molecules in the hydrophobic bacteriorhodopsin proton channel.<sup>21</sup> They reported that the transfer of four water molecules from bulk solvent to the channel is thermodynamically feasible which can further explain the mechanism of proton transfer in bacteriorhodopsin. The standard free energy barrier of transferring the water molecule from the bulk to the binding interface was obtained by Hamelberg and McCammon, as 3.1 kcal/mol (12.96 kJ/mol).<sup>16</sup> In the recent past, Zheng and Lazaridis used methods borrowed from statistical mechanics of inhomogeneous fluid solvation theory to study the contribution of a water molecule to the energy, entropy, and heat capacity of protein solvation.<sup>22</sup>

Various aspects of hydration dynamics of HP-36 have been studied by molecular dynamics simulations, mostly from the study of individual trajectories.<sup>12–14</sup> In these studies, dynamical characterization was used to infer, indirectly, the free energy surface of the water molecules at protein surface. Thus, distinction between different states of water was made by studying dynamical characteristics of many hydration water molecules. This led to the identification of slow water molecules near the middle helix (helix 2) of HP-36 (see Figure 1 below). However, free energy barriers were not obtained.



**Figure 1.** Average free energy profile of interfacial water molecules around the three helices of HP-36. A representative configuration of HP-36 highlighting three helices in three different color codes is shown in the inset. Helix-1 is drawn as a blue ribbon, helix-2 as a red ribbon, and helix-3 as a green ribbon. The coils in between the helices are highlighted in cyan. We have taken representative water molecules in the hydration layer (except doubly H-bonded species which are explicitly and separately treated) of each helix and averaged over their free energy profile. Each profile is scaled with respect to the bulk free energy. Note that the average barrier height is the lowest for the third helix. Magnitudes are listed in Table 2.

In this study, we use the method of umbrella sampling to obtain the free energy barriers for the transition of the hydration water to the bulk. The bulk water molecules are those

water molecules that remain in a region beyond 10 Å away from the protein surface atoms. Bulk free energy is estimated as -69.72 kJ/mol.<sup>23</sup> Early experimental and simulation studies found that the free energy of transferring a water molecule from hydration layer to bulk could be in the range of 6–12 kJ/mol.<sup>16,24–26</sup>

Our free energy calculations reveal several interesting results that throw light on different aspects of hydration water. First, we find that there is a distribution of free energy barrier separating hydration water from the bulk. Second, water molecules can indeed be separated into two groups. There are a few water molecules which are separated by a large barrier, but the number of such molecules, at least for HP-36, is few, less than ten. The majority of water molecules are separated by a small barrier, on the order of a couple of  $k_B T$  only ( $k_B T$  is the Boltzmann constant times the temperature of the system). Third, the majority of interfacial water molecules have larger free energy than the bulk. This surprising result means that these water molecules are in an unfavorable environment at the protein surface, being deprived of the hydrogen bond network of the bulk. Most of these water molecules exhibit fast rotational and translational dynamics and may be even faster than those in the bulk!

The organization of the rest of the paper is as follows. In next section we briefly describe the system used (HP-36 and water). Section 3 contains the calculated free energy profile. Section 4 includes the dynamical characterization of water. In section 5, we showed the movement of few interfacial water molecules along the escape direction. From the trajectory analysis, we obtain two-dimensional free energy surfaces of escape of those hydration layer water molecules which are shown in section 6. Section 7 includes an analysis of the rate of escape of quasi-bound water molecules from the bulk, by using well-known Kramers' theory. In section 8, we discuss the heterogeneous solvation dynamics at the protein surface. Section 9 concludes with a brief summary of results. Section 10 contains the simulation details and free energy calculation method.

## 2. STRUCTURAL DETAILS OF THE PROTEIN STUDIED: HP-36 IN WATER

HP-36, a small globular protein with 36 residues, is the thermo-stable subdomain present at the extreme C-terminus of the 76-residue chicken villin headpiece domain.<sup>27,28</sup> Villin is a unique protein which can both assemble and disassemble actin structures.<sup>29,30</sup> HP-36 contains one of the two F-actin binding sites in villin necessary for F-actin bundling activity. The structure of HP-36, as determined by NMR spectroscopy, consists of three short R-helices surrounding a tightly packed hydrophobic core as shown in the inset of Figure 1. These helices are connected and held together by a few turns and loops and a hydrophobic core. In this work we number the residues from 1 to 36 that correspond to residues 41–76 in the NMR structure. We denote the three R-helices as helix-1 (Asp-4 to Lys-8), helix-2 (Arg-15 to Phe-18), and helix-3 (Leu-23 to Glu-32). The biological activity is believed to be centered around helix-3, which contains 10 amino acid residues. This protein subdomain has been studied extensively in recent years, concentrating especially on its folding phenomena.<sup>31,32</sup>

## 3. FREE ENERGY PROFILE OF HYDRATION WATER

Dynamical coupling between protein and hydration layer water is studied by assuming the existence of interfacial water

molecules that have relatively long-lived H-bonds with the polar amino acid residues of protein.<sup>13</sup> Simulations show that these interfacial water molecules can be categorized into two broad classes, free and quasi-bound as we discussed in the introduction. Quasi-bound molecules can again be subdivided into singly H-bonded and doubly H-bonded.<sup>33</sup> Dynamic equilibrium between free and quasi-bound water is an important process occurring at the protein surface. We evaluate the average free energy profile of the interfacial water molecules both for free and quasi-bound water molecules using the umbrella sampling method<sup>34</sup> (for all the technical details, see the section 10.2).

**3.1. Free Energy Barrier along the Escape Coordinate.** The strongly H-bonded water molecules are the minority compared to the rest of interfacial water near a protein surface. However, they might play critical role in the stability of the protein. In Figure 1, we plot the average binding energy of water molecules in the hydration layer of three helices of HP-36 (also shown for clarity). The magnitudes of barrier heights are consistent with the corresponding polar solvent accessible surface area of the three helices (as shown in Table 2), which will be further discussed later. The noticeable difference between the free energy barrier of water molecules near helix-1 and helix-2 also supports the fact that higher the polar solvent accessibility, more bound the adjacent water surface. The lowest barrier height in energy profile of water around helix-3 signifies its loosely bound character. The labile nature of the helix-3 hydration layer also correlates well with the biological activity of protein. This could be important because helix-3 contains the active site residues for actin binding.

**3.2. Distribution of Barrier Heights.** Barrier free energy profiles presented above for three helices are averaged over both the quasi-bound and free water molecules (except doubly H-bonded species) which are within the hydration layer of the corresponding helices. To view the behavior of each of those water molecules, we have calculated the barrier height from their respective free energy profiles. The distribution of barrier heights (Figure 2) of the interfacial water molecules show a

environment. The lowest barrier height distribution near helix-3 again undoubtedly signifies that the corresponding hydration layer water molecules are loosely bound.

### 3.3. Free Energy Profile for Doubly Hydrogen Bonded Water Molecules.

From the simulation trajectory we select two such water molecules that are doubly H-bonded. Among them, one is H-bonded to the residue Asp-4 in helix-1 and also to Arg-15 in helix-2. We denote this water molecule as water-1 (Figure 3a). This water molecule is motionally restricted with long residence times of around 437 ps during which it resides within the hydration layer of helix-1 and helix-2. The second quasi-bound water molecule has been chosen from another location which is doubly H-bonded to Thr-14 near the helix-2 with residence time around 120 ps. We denote this as water-2 (Figure 4a). We evaluate the free energy profile of both for these two water molecules using the Umbrella Sampling method.

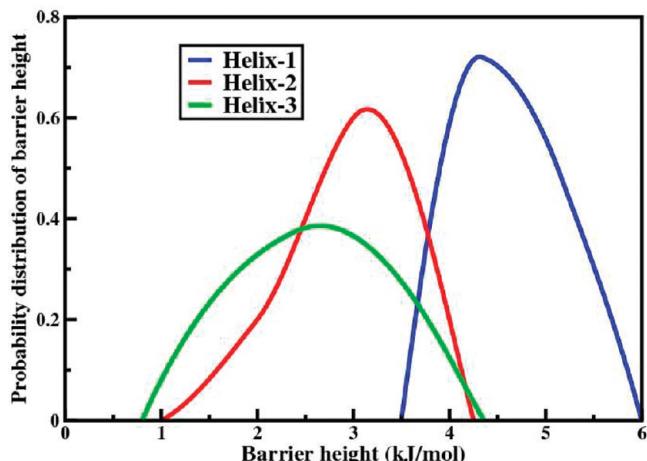
From the above figures, we observe that the quasi-bound water molecules reside in a remarkable deep minimum as they form double hydrogen bonds to the protein residues. If the distance is further decreased, that is, we move toward the protein surface, then due to overlap with the protein atoms, van der Walls repulsion comes to play a role and the energy of the water molecule increases sharply. On the other hand, if the distance is increased, the hydrogen bond formed between the protein and water molecule gets stressed and energy again increases.

The energy difference between the minima and barrier top is 12.75 kJ/mol for water-1 (Figure 3b) and 9.3 kJ/mol for water-2 (Figure 4b) which is in the range of the hydrogen bond energy (5–30 kJ/mol). The noticeably larger barrier height in the case of water-1 signifies that it spends more time near helix-2 than water-2. The interaction energy between this particular water molecule and the residues Asp-4 near helix-1 and Arg-15 near helix-2 plays a crucial role for such large stabilization. The topography of this particular water molecule assists to build a stable environment near helix-2. Thus, only one strongly hydrogen bonded water molecule may be sufficient to affect the dynamics of the whole resulting slow structural decay near helix-2.

These water molecules have high propensity to reform their hydrogen bonds after a breakage. The dangling motion of these water molecules plays the key role behind their long residence time.<sup>35</sup>

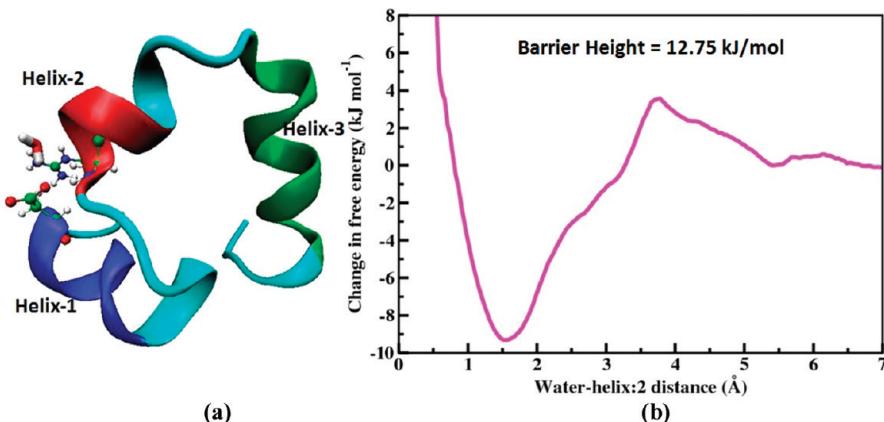
**3.4. Comparison of Free Energy of Quasi-Bound versus Interfacial Free Water.** In Figure 5, we present a comparison between quasi-bound water and the free water. The deep minimum of the quasi-bound water and the significantly large barrier height indicate that the presence of this type of quasi-bound water molecule is enough to affect the dynamics around helix-1 and helix-2. Interestingly, the strongly tied water molecules have considerably lower free energy than the bulk, whereas the free water behaves entirely in the opposite manner.

**3.4.1. Distribution of Interaction Energy of Interfacial Water Molecules.** The single molecule energy distribution for interfacial water molecules are shown in Figure 6. The abscissa gives the total interaction energy of a single water molecule with the rest of the atoms in the whole system. The distribution for free water is found to be similar to that of the bulk. The quasi-bound water molecules are displaying higher stability with more negative interaction potential. Such quasi-bound water molecules are stabilized by about 13 kJ/mol energy compared to the bulk that mostly arises due to doubly hydrogen bond formation and electrostatic interaction with the protein

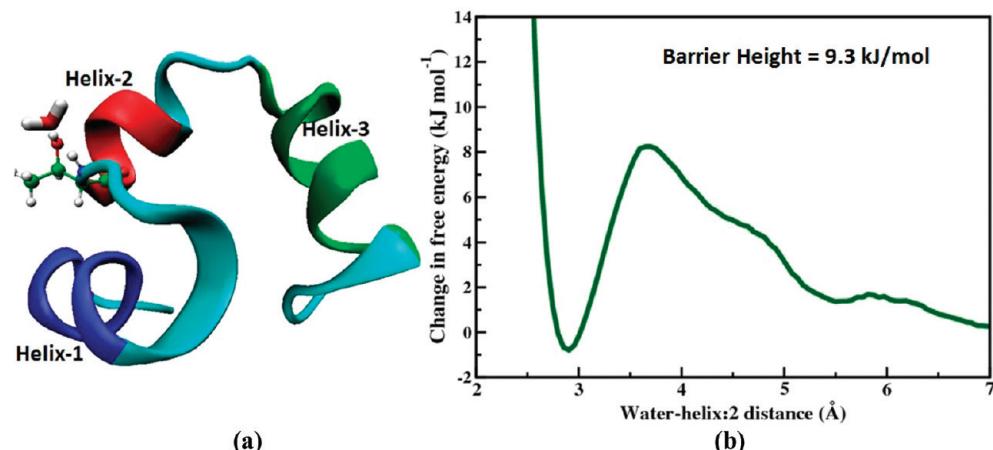


**Figure 2.** Barrier height distribution for interfacial water molecules near the three helices. Barrier heights are evaluated from the free energy profile of hydration layer water molecules. The lowest barrier height is the signature for loosely bound water structure surrounding helix-3.

remarkable diversity depending upon the helix region. The interfacial waters near helix-1 and helix-2 show comparatively higher barrier height indicating that they are in the more stable



**Figure 3.** Free energy profile of the strongly hydrogen bonded water-1. (a) The location of the water-1. The water molecule is hydrogen bonded with the side-chain oxygen atom of Asp-4 in helix-1 and backbone NH group of Arg-15 in helix-2. The helices are drawn in ribbon representation where the color sequences are as follows: helix-1 in blue, helix-2 in red, and helix-3 in green. The coils are colored in cyan. The atoms of the residues Asp-4 and Arg-15 are drawn using a ball-and-stick model, and the quasi-bound water molecule is drawn using a lycorice model. (b) Free energy profile of that quasi-bound water plotted against distance from the helix. The free energy shows a barrier height of 12.75 kJ/mol, and its residence time is 437 ps.



**Figure 4.** Free energy profile of the strongly hydrogen bonded water-2. (a) Location of water-2. The water molecule is hydrogen bonded with the side-chain OH group and backbone NH group of Thr-14 near helix-2. The helices are drawn maintaining the previous color code. The atoms of the residue Thr-14 drawn using a ball-and-stick model, and the quasi-bound water molecule is drawn using lycorice model. (b) Free energy profile of this quasi-bound water molecule with barrier height 9.3 kJ/mol. Its residence time is 120 ps.

residues. However, these quasi-bound water molecules have lower entropy which is found to play an important role in determining their relative stability with respect to the bulk.

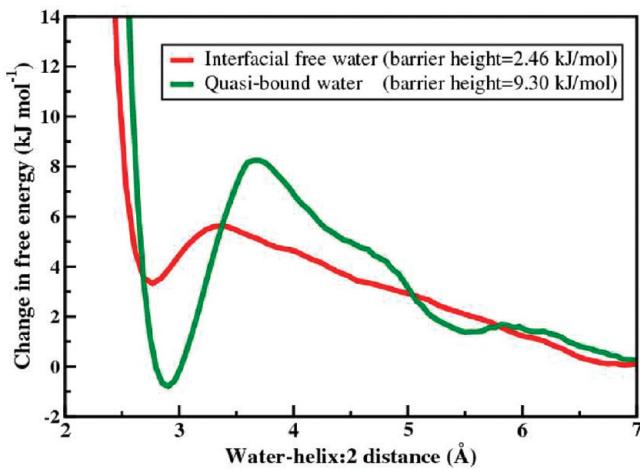
From the free energy and the monomer potential energy data, we can approximately estimate the entropy loss of quasi-bound water over the interfacial free water to be about  $20 \text{ J mol}^{-1} \text{ K}^{-1}$ . The loss in entropy for quasi-bound water is clearly related to the higher constrain on its movement that it faces. Distribution of single molecule interaction energy of interfacial free water indicates negligible change in interaction energy compared to the bulk. These water molecules do not have any hydrogen bond interaction with protein residues. However, neighboring free water molecules prefer to surround a quasi-bound water molecule in a tetrahedral manner and such cluster formation further decreases the freedom of these water molecules. Thus indirect interaction with protein residues lowers the entropy of these interfacial free water molecules which is also accountable for the overall dynamical retardation of the protein hydration layer. So, neither enthalpically nor entropically, these interfacial free water molecules in the hydration layer gain much stability compared to the bulk.

#### 4. CORRELATION BETWEEN STRUCTURE AND DYNAMICS OF PROTEIN HYDRATION LAYER

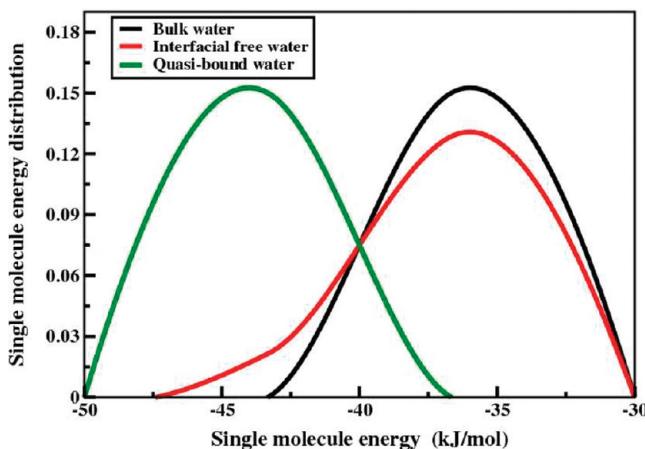
##### 4.1. Slow Dynamics of a Few Water Molecules.

Analyses of the dynamics of protein–water hydrogen bonds around different secondary structures of HP-36 have already revealed that the hydration layer surrounding helix-3 is the most labile among the layers around the secondary structures. This can be correlated with faster structural relaxation of the hydrogen bonds between its residues and water. It also reflects in the faster translational and rotational motions of water near helix-3. This was attributed to the presence of a few strongly hydrogen bonded water molecules that makes the relaxation slow near helix-1 and helix-2.<sup>13,14</sup>

As mentioned earlier our molecular dynamics simulation studies clearly reveal that the interface of HP-36 consists of three different, albeit transient species, namely, IFW, IBW1, and IBW2. To understand the precise reason of emergence of the dynamical heterogeneity in the hydration layer, we have quantified the relative ratio of these different entities in three  $\alpha$ -helical regions accordingly. The ratio, IBW2:IBW1:IFW is listed in Table 1.



**Figure 5.** Comparison between average free energy profiles of interfacial free water molecule and interfacial quasi-bound water molecule. The remarkable difference is evident from the depth of the 1st minima and from the barrier height of the respective profiles. The large barrier height of the quasi-bound water corresponds to its extra stability near the protein surface.



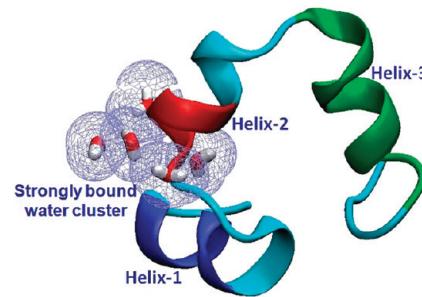
**Figure 6.** Distributions of single molecule interaction energies of interfacial water molecules. Comparison of the relative interaction energy of the free water and the quasi-bound water with respect to the bulk clearly show the higher stability of the quasi-bound water than that of the free water. For free water, the peak position remains unaltered with the bulk.

**Table 1. Relative Ratio of IBW2:IBW1:IFW around Three Helical Regions**

helix	IBW2:IBW1:IFW
helix-1	0.6:3.8:5.6
helix-2	1.0:2.4:6.6
helix-3	0.2:2.6:7.2

We find from this evaluation that the doubly hydrogen bonded species are relatively rare because their formation requires a highly constrained arrangement that is entropically expensive. However, their long-lived nature is largely responsible for the dynamical heterogeneity present and observed in the hydration layer. It is important to note that the abundance of strongly bound water molecules is mainly responsible for dynamical slowing down of hydration layer near helix-2.

We detect a few motionally restricted water molecules located between helix-1 and helix-2 that are doubly hydrogen bonded to the protein residues. In Figure 7, we show such a cluster of water molecules which is captured in one snap.

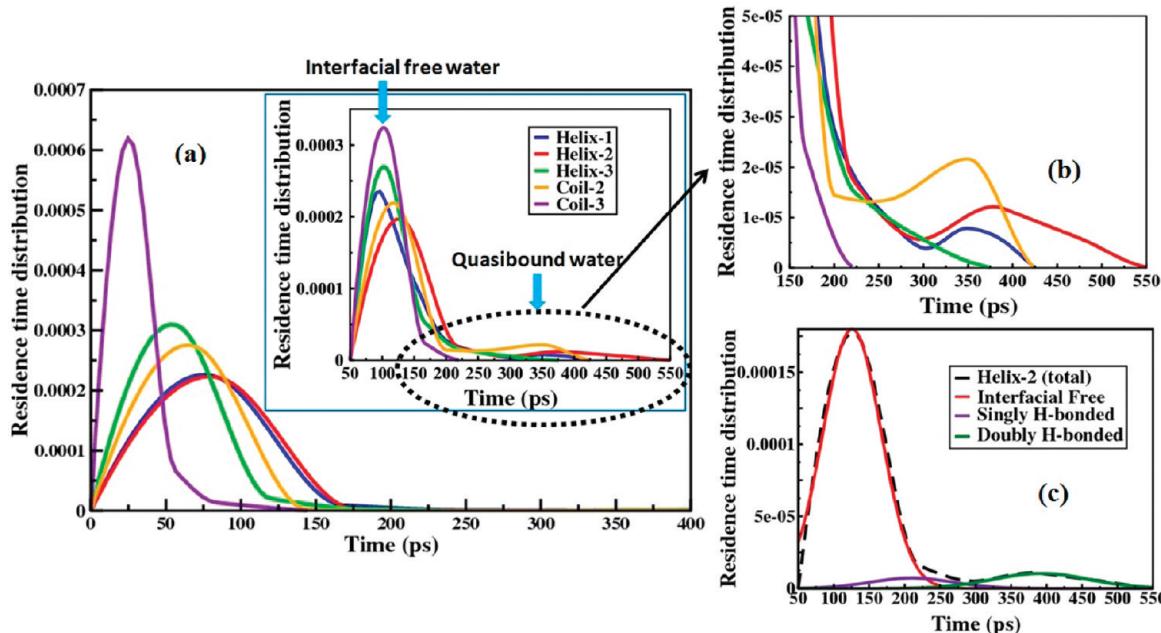


**Figure 7.** Representative snap extracting from the simulation trajectory, showing the location of some water molecules in the first hydration layer which are doubly hydrogen bonded to the protein residues. The helices are drawn as colored ribbons maintaining the previous color code. The quasi-bound water molecules are drawn using the lycorice model.

**4.2. Residence Time Distribution.** Quasi-bound water molecules are motionally restricted with long residence time of 100 ps or more within the hydration layer of HP-36. To distinguish such water molecules in the hydration layer according to their residence time, we evaluate their residence time distribution, shown in Figure 8. The first peak of the distribution shows that by far the largest number of water molecules belong to free molecules with residence times in 50–100 ps range (or, even less for coil-3) (see Figure 8a). These water molecules are mostly either free or singly hydrogen bonded to the protein residues. The population of such water molecules is particularly large near helix-3 and coil-3. The significant population of this particular residence time region (50–100 ps) may suppress the contribution from those interfacial water molecules which are less in number but more bound to the protein surface.

To highlight that region we plot the residence time distribution of only those water molecules that stay more than 50 ps as shown in the inset of Figure 8a. In this plot a clear appearance of the two peak character validates our classification. Here we observe that for all three helices there exists a fraction of water molecules with residence time in the range of 150–450 ps. But a clear distinction of two time zones has been found by zooming on above 150 ps (see Figure 8b). More interestingly, in case of helix-2 we find a long residence time region in the distribution plot indicating the presence of some slow water molecules with residence time more than 400 ps.

Multi-Gaussian fits of the residence time distribution of water molecules near helix-2 reveal the presence of three time zones, indicating the presence of three different types of water molecules in the hydration layer (as shown in Figure 8c). The short time distribution is due to the presence of large number of free interfacial water molecules. The second peak corresponds to the singly hydrogen bonded water. In addition, one finds a third peak, shifted to the long time region due to the presence of doubly hydrogen bonded water molecules. Note the huge overlap between the interfacial free and singly hydrogen bonded water in the distribution which can be attributed to the dynamical equilibrium between the two species. It is also



**Figure 8.** (a) Residence time distribution of the hydration layer water molecules near three helices and two coils of HP-36. The residence time distributions of water molecules which stay in the hydration layer more than 50 ps are highlighted in the inset of part a. Note the bimodal nature in the plots of the residence time distribution near helix-1, helix-2, and coil-2 regions. (b) Close-up of the circled section of the inset of part a. A distinct time zone has been characterized for the quasibound water molecules. (c) Multi-Gaussian fit of the distribution for helix-2. Three different Gaussian components correspond to the presence of three types of water molecules in the hydration layer.

worth mentioning that similar fitting can be done for other distributions having a long time component. However, the fitted parameters for different species are quite sensitive to the details of protein architecture.

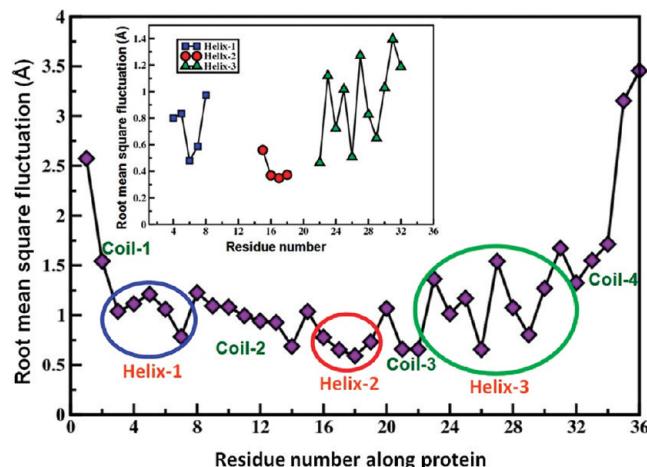
**4.3. Solvent Accessibility near the Three Helices.** It has already been reported that the interfacial water molecules around helix-2 exhibit the slowest orientational dynamics.<sup>12–14</sup> The relative exposure of the polar probe residues in different helices might play a role which determines the essential part of protein–water interaction causing such slow dynamics. To estimate the exposure of the polar probe, we have calculated the relative polar solvent accessible surface area averaged over a 20 ns trajectory (shown in Table 2). We find that the solvent

**Table 2. Relative Polar and Nonpolar SASA (Solvent Accessible Surface Area) and the Average Free Energy Barrier for the Three Helices**

helix	polar SASA	nonpolar SASA	average free energy barrier
helix-1	62%	38%	3.82 kJ/mol
helix-2	44.47%	55.53%	2.46 kJ/mol
helix-3	43.25%	56.75%	1.93 kJ/mol

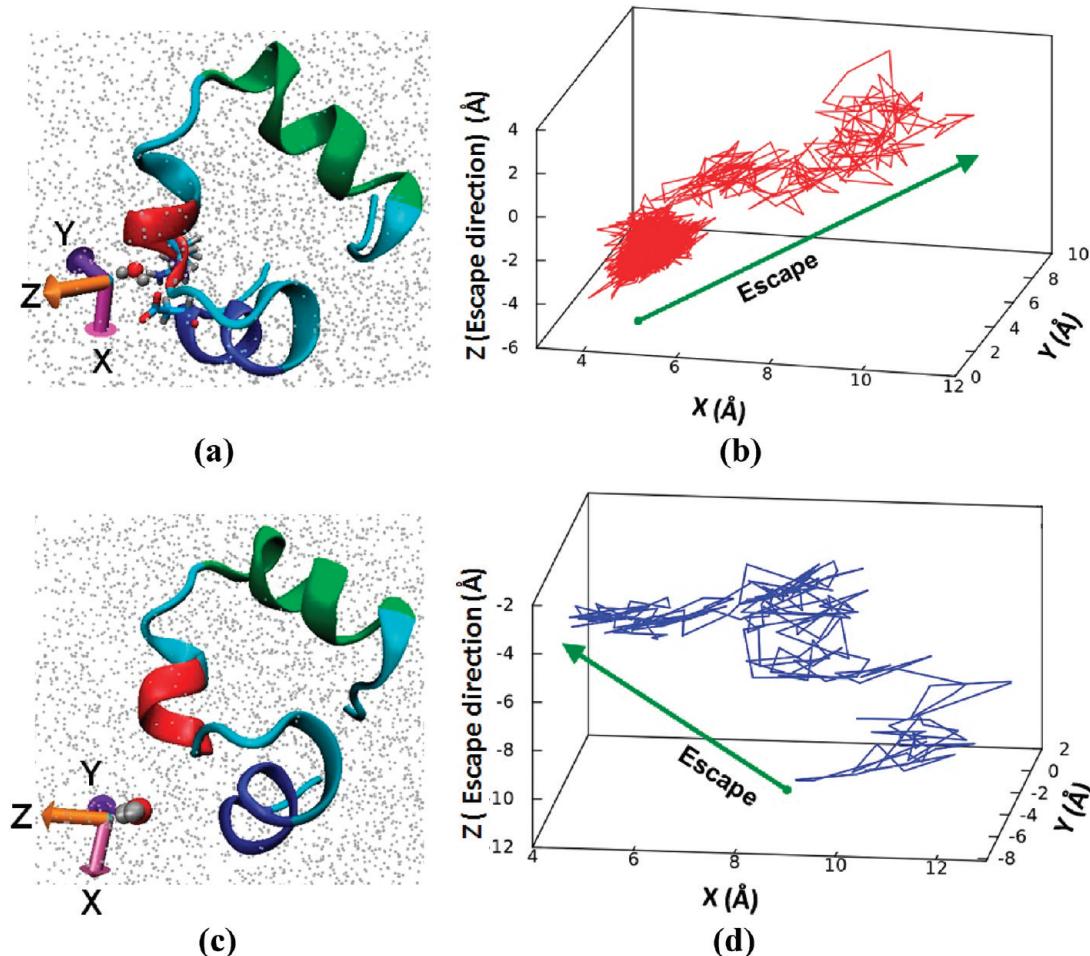
accessibility to the polar probe is low near helix-2 (44.47%) compared to that of helix-1 (62%). In fact we have observed that the exposed side chains of helix-1 assist to bind the water molecules near helix-2.

**4.4. Conformational Fluctuations of Secondary Structures.** To understand the dynamical coupling between the conformational fluctuation of protein and surrounding water molecules, we have monitored the root-mean-square fluctuation (RMSF) of position of all non-hydrogen atoms of each protein residue. This is shown in Figure 9. RMSF can provide important information about the side chain mobility of



**Figure 9.** Root-mean-square fluctuation (RMSF) of C-alpha atoms containing each amino acid residue of HP-36. The RMSF of three helix residues are highlighted in the inset which shows remarkable lowering of fluctuation near helix-2. The higher RMSF values for helix-3 residues indicate an immense flexible environment around helix-3.

the residues which can influence and also be influenced by the hydration layer dynamics. An important feature to note from this plot is the significantly lower value of RMSF for each second helix residue which assists to build up a stable hydration layer in the second helix premises. It is also interesting to note the RMSF values of helix-3 residues that exhibit large conformational fluctuation. The higher conformational fluctuation of helix-1 and helix-3 is expected because of their closeness to the C-terminus which is found to be more flexible in nature. Such dynamical flexibility around helix-3 can facilitate its biological activity because fluctuations around the active site are necessary to soften its reaction energy barrier.<sup>36,37</sup>



**Figure 10.** Characterizing  $X$ -,  $Y$ -, and  $Z$ -coordinates of the (a) interfacial quasi-bound water molecule and (c) interfacial free water molecule. The  $Z$ -coordinate represents the direction of escape of the water molecule from protein interface to the bulk region where the  $X$ - $Y$  plane corresponds to the protein interface. The projection of the trajectory of the water molecule is shown along the escape direction ( $Z$ ) and along the protein surface ( $X$ - $Y$ ) for (b) interfacial quasi-bound and (d) interfacial free water. The trajectories are collected during their stay within 6 Å. Note the escape along  $Z$ -direction for both of the cases.

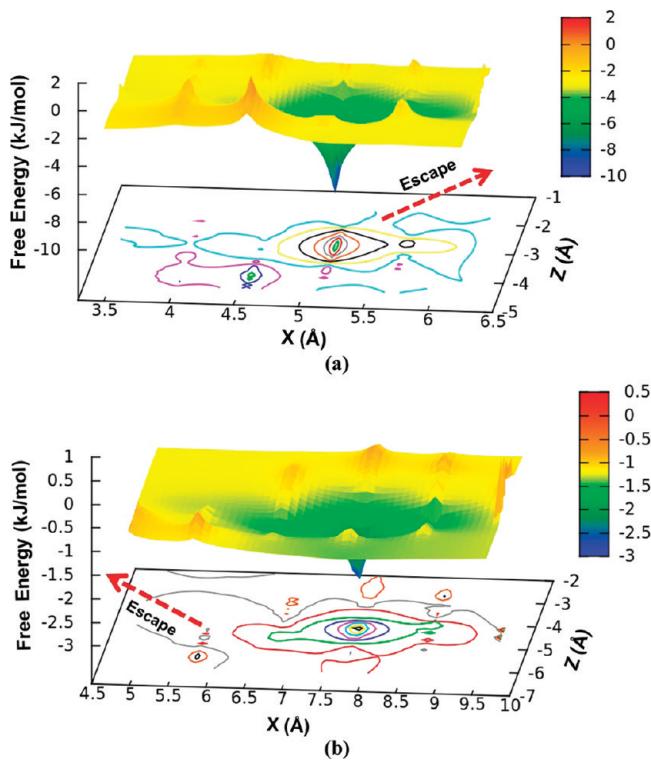
## 5. WATER MOTION IN THE PROTEIN HYDRATION SHELL

In this section, we trace the movement of a strongly hydrogen bonded water molecule near helix-1 and helix-2 (Figure 10a). We also notice the motion of interfacial free water molecule located in the same region (Figure 10c). Water motion on the protein surface can be decomposed into two types of the movement, that in the tangent (to the protein surface) direction and along the normal direction. We investigate both the in and out motion of the two types of interfacial water molecules (interfacial free and quasi-bound) along the escape direction ( $Z$ -axis) and the lateral or parallel motion along the protein interface ( $X$ - $Y$  plane). We find from the trajectories that the water molecules move mostly along the  $Z$ -direction to escape from its hydration layer, and hence the  $Z$ -direction is designated as the escape direction. Both of the three-dimensional trajectories are cropped during its stay within 6 Å (corresponding to the second minima of the radial distribution function of oxygen atom of water) from the protein helix interface. The dense part of the trajectory shown in Figure 10b indicates the restricted movement of quasi-bound water molecule within the hydration layer. During this time interval, the quasi-bound water molecule visits some stable region once and again returns back to the region from where it starts and carries out the dangling motion. Afterward this water molecule

moves laterally as well as in the perpendicular direction over a distance corresponding to the escape from the hydration layer. The trajectory of interfacial free water is somewhat different from the quasi-bound water molecules (Figure 10d). The disordered motion during its residence time signifies lower stability due to the weak interaction with the protein residues.

## 6. FREE ENERGY SURFACE FROM TRAJECTORY ANALYSIS

We evaluate free energy surfaces of the tagged interfacial water molecules from the histogram of the above trajectories. The contour map as well as free energy surface for strongly hydrogen bonded water molecule shows the presence of two distinct minima along the protein interfacial plane bearing its doubly hydrogen bonded character. The deep minimum arises in Figure 11a probably due to the breaking of one strong hydrogen bond and the comparatively shallow minimum appears due to loss of another hydrogen bond formed with the protein residues. However, along the  $Z$ -axis, (which is characterized as the direction of escape), its escape to the bulk is evident from the plot. On the other hand, in the case of interfacial free water molecules, we find only one shallow minimum (Figure 11b) which may be due to the presence of one weak hydrogen bond interaction.



**Figure 11.** Two-dimensional free energy surface along with its contour map for (a) strongly hydrogen bonded quasi-bound water and (b) interfacial free water. The color code of the free energy landscape has been so chosen that the closely spaced regions can be distinguished clearly. The presence of two minima in part a corresponds to two hydrogen bond breaking events, whereas in part b the existence of a single minimum indicates only one weak hydrogen bond rapture. For both cases, the escape along the Z-direction is evident from the contour.

## 7. THEORETICAL ESTIMATE OF THE RESIDENCE TIME FROM FREE ENERGY BARRIER

It is evident from Figures 3 and 4 that for a few water molecules (whose number is small) the barrier for escape from the hydration layer is quite high. One can therefore apply theory of activated barrier crossing dynamics, such as Kramers' theory<sup>38</sup> to estimate the residence time, as was proposed originally by Nandi and Bagchi.<sup>4</sup> Kramers' theory of reaction rate in the high friction limit is given by

$$k_R = \frac{1}{\omega_b} \left( -\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + \omega_b^2} \right) \left\{ \frac{\omega_R}{2\pi} \exp\left(-\frac{E_b}{k_B T}\right) \right\} \quad (1)$$

where  $\omega_R$  and  $\omega_b$  are the frequencies that confine the harmonic reactant well and barrier top, respectively. In this expression, the part between the curly brackets gives the transition state theory (TST) result for the reaction rate and the expression in front of the curly brackets approximates the correction to the TST result, for the moderate-to-strong friction ( $\gamma$ ) regime.  $E_b$  corresponds to the activation energy required to cross the free energy barrier.

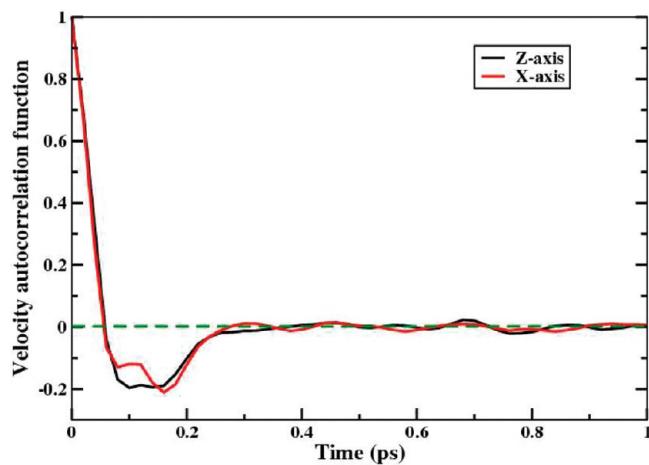
The friction along the reaction coordinate has been obtained by using Einstein's relation (DSE) between diffusion and friction,

$$\gamma = \frac{k_B T}{D} \quad (2)$$

The diffusion coefficient has been obtained from velocity autocorrelation function by using the well-known relation.

$$D = \frac{1}{3} \int_0^\infty \langle v(0)v(t) \rangle dt \quad (3)$$

In Figure 12, we show the computed velocity time correlation function along the escape (reaction) coordinate (Z-axis)



**Figure 12.** Velocity autocorrelation function along the escape direction (termed as Z-axis as defined earlier) and its perpendicular direction. Note the increased oscillation along the escape direction than along its perpendicular direction.

and along its perpendicular direction (X-axis). We observe the existence of more oscillations in the velocity time correlation function along the escape direction, reflecting the quasi-bound state. We have calculated the diffusion coefficient both for the escape direction and its perpendicular direction from the above velocity autocorrelation function within its residence time in the hydration layer. The motion of this water molecule at the interface is highly anisotropic because of the interaction with protein side chain residues and the hydration layer water network.<sup>39</sup> In the present study, the velocity autocorrelation function along the escape direction satisfactorily goes to zero by 1 ps which allows us to define a "local" diffusion coefficient in those directions comprehensively. On comparing these diffusion constant values, it is clear that these quasi-bound water molecules are translationally more constrained along the X-direction within this time limit. The motion along the escape direction has a diffusion constant value  $0.71 \times 10^{-5} \text{ cm}^2/\text{s}$  indicating significantly restricted shifting compared to the bulk. The calculated value correlates well with the previously obtained diffusion constant value observed for interfacial water around myoglobin.<sup>24</sup> The solvent friction is evaluated following eq 2. The free energy well and the barrier top in plot of Figure 3b are fitted to a harmonic well, and then, we obtain the harmonic reactant well frequency ( $\omega_R$ ) and barrier top frequency ( $\omega_b$ ), respectively. These evaluations make possible to assess the rate of escape of the interfacial quasi-bound water obeying eq 1. The residence time calculated from this theoretical rate estimate contrasts favorably well with our simulation results. All the parameters calculated are enlisted in Table 3.

**Table 3. Estimation of Properties of Quasi-bound Water Molecule in the Protein Interface at 300 K**

properties	estimation
diffusion along Z ( $D_Z$ ) ( $\times 10^{-5}$ cm $^2$ /s)	$0.71 \pm 0.003$
diffusion along X ( $D_X$ ) ( $\times 10^{-5}$ cm $^2$ /s)	$1.00 \pm 0.005$
friction ( $\gamma$ ) ( $\times 10^{14}$ s $^{-1}$ )	$2.14 \pm 0.002$
well frequency ( $\omega_R$ ) ( $\times 10^{13}$ s $^{-1}$ )	$1.37 \pm 0.001$
barrier frequency ( $\omega_b$ ) ( $\times 10^{13}$ s $^{-1}$ )	$2.58 \pm 0.001$
rate of escape ( $k_R$ ) ( $\times 10^9$ s $^{-1}$ )	2.47
residence time (ps) (theoretical)	404
residence time (ps) (simulation)	437

## 8. HETEROGENEOUS SOLVATION DYNAMICS AT PROTEIN SURFACE

Experiments have reported vastly different time constants for solvation dynamics of dyes placed near the surface of a protein. The experimental values of solvation time range from a few picoseconds to a few nanoseconds. Detailed experiments by Zewail and co-workers have attributed the larger than bulk water solvation time to the existence of a slow component in the solvation time correlation function which they attributed to the slow dynamics of quasi-bound water molecules.<sup>6</sup>

The present study shows that such solvation dynamics studies can lead to different time constants depending on the location of the probe. For example, if the probe is placed by a covalent bonding to third helix, then the dynamics would reflect fast dynamics. On the other hand, if the probe is placed near the junction of first and second helix where the cluster of slow water molecules is placed, then may find slower dynamics.

Pal et al. established a simple relation between the solvation time and residence time. Although that relation may provide an upper bound for the solvation time correlation function, it is expected to reflect a trend in the long time component of the solvation time correlation function.<sup>7</sup>

This heterogeneity in water dynamics at protein surface is of course is a consequence of the transient structure of the protein-water combined system. But one important point to note here is that the presence of strongly interactive water molecules is mainly responsible for the emergence of region wise distinct dynamical behavior. The doubly hydrogen bonded species are relatively rare because their formation requires constrained arrangement that is entropically demanding. Nevertheless, their long-lived nature is largely responsible for the dynamical heterogeneity evolved in hydration layer.

## 9. CONCLUSIONS

Let us now summarize the main findings of this paper. In this article, we have presented results from free energy calculation of interfacial water molecules around HP-36 to understand the structural and dynamical heterogeneity in the protein campus. The evaluated free energy change of escaping a water molecule from hydration layer to the bulk is consistent with that of the previous estimated values.<sup>16,24–26</sup> The lowest thermodynamic stability of the hydration layer water molecules near helix-3 correlates well with their flexible nature as found in the previous simulation studies.<sup>13,14</sup> We locate such a region near helix-2 which allows staying a cluster of water molecules which are long-lived as well as strongly hydrogen bonded with protein residues. The remarkably higher stability of these quasi-bound water molecules with respect to the bulk water is evident from the depth of their free energy minima as well as from their

single particle energy distribution. They have substantially lower interaction potential due to favorable interaction with protein. As the free water molecules do not form such interaction with the protein residues, they behave similarly to the bulk energetically.<sup>33,36</sup> But in terms of free energy, they are less stable than bulk.

Residence time distributions demonstrate the long-lived nature of these quasi-bound water molecules which again correlates well with the sluggish dynamical behavior near second helix of HP-36. Moreover, after quantification of the ratio of different types of water molecules in different helical regions, the relative abundance of doubly hydrogen bonded species near the second helix premises accompanied by a comparatively high free energy barrier of such quasi-bound water molecules firmly establish the microscopic origin of the dynamical slowing down near the second helix region.

To establish a relation between the relative stability of interfacial waters and exposure of the polar probe residues near each helix, we have calculated relative polar SASA. The higher polar SASA offers higher stability to the neighboring water molecules. The coupled performance of conformational flexibility of protein and the dynamics of water is explored monitoring the RMSF values of all residues of HP-36. Helix-1 and helix-3 residues exhibit large scale conformational fluctuation and heterogeneous dynamics. This heterogeneity becomes more prominent near helix-3. But the residues in helix-2 are exposed to almost identical environment. The exposure of the side chains of the polar residues (ASP-15 and Ser-16) to the solvent is mainly responsible for the existence of well-organized water clusters. Interestingly some residues like Asp-4, Thr-14 are also exposed to create a location in between helix-1 and helix-2 where maximum quasi-bound water can hang about for a long time. During this residence time, we trace the movement of a few interfacial water molecules and notice their escape from the hydration shell. The two-dimensional free energy surfaces obtained from the trajectory analyses clearly indicate the higher stability of the quasi-bound water molecules than the interfacial free water molecules. Hydrogen bond lifetime kinetics of water molecules surrounding HP-36 has already been studied in detail.<sup>13</sup> Though in that particular work no distinctions were made between singly and doubly hydrogen bonded water molecules, it is in good agreement with our present study.

Thus here we contribute our effort to establish a correlation between the thermodynamics and dynamics of hydration layer water molecules influenced by the conformational fluctuation of protein. This investigation offers important information about the location created by the protein residues of lower flexibility. The exposed polar side chains in this area allow some strongly quasi-bound water molecules that are significantly stable and their existence slows down the dynamics near second helix. The noisy third helix region is responsible for the faster dynamics of hydration layer water molecules of helix-3. Most importantly the dynamics is firmly validated by the thermodynamics of interfacial water molecules.

While earlier studies have explored dynamics of water near the protein surfaces largely by employing molecular dynamics simulations, no free energy calculations correlated with the dynamics and structure were reported, to the best of our knowledge. Clearly, the quasi-bound water molecules present at the junction of helix-2 and helix-3 act as a “clamp” to stabilize the structure of the protein in solution. This allows helix-3 to

execute larger amplitude motions that facilitate its biological function, without destabilizing the structure.

We conjecture that this stabilizing effect of quasi-bound (on the time scale of motions of other) water molecules could be a general phenomenon in the functions of various water-soluble proteins and enzymes. However, this needs to be verified.

## 10. METHODOLOGY

**10.1. Molecular Dynamics Simulation Details.** As mentioned earlier several studies have been done to investigate the free energy of hydration water, both experimentally and computationally. Here we correlate for the first time the thermodynamics with dynamics of those hydration waters. To assess the dynamics we performed molecular dynamics (MD) simulations of the protein in water by using the GROMACS Package. The simulation began with the crystal structure of the HP-36 obtained from the NMR structure of the villin headpiece subdomain as reported by McKnight et al.<sup>27,28</sup> The initial coordinate collected from the Protein Data Bank (PDB ID 1VII). The two ends (Met-1 and Phe-36) of the protein were capped properly. Protein is centered in a cubic box of 5.14 Å. All atom topologies were generated with the help of pdb2gmx and OPLS set of parameters available in GROMACS. The protein was solvated with pre-equilibrated SPC/E water model using genbox.<sup>40</sup> A total of 4487 water molecules were added. After steepest descent energy minimization, each trajectory was propagated in a *NVT* ensemble and equilibrated for 2 ns. All the simulations in this study were done at 300 K and 1 bar pressure. The temperature was kept constant using the Nose–Hoover thermostat.<sup>41,42</sup> It was followed by an *NPT* equilibration for 10 ns using the Parrinello–Rahman barostat.<sup>43</sup> Finally, production runs were performed for each system in an *NPT* ensemble. Each simulation used a time-step of 2 fs. All the analyses were executed from the 20 ns trajectory with each configuration saved after 1 ps. Periodic boundary conditions were applied and nonbonded force calculations employed a grid system for neighbor searching. Neighbor list generation was performed after every 5 steps. A cutoff radius of 1.2 nm was used both for neighbor list and van der Waal's interaction. To calculate the electrostatic interactions, we used PME<sup>44</sup> with a grid spacing of 0.12 nm and an interpolation order of 4.

**10.2. Free Energy Calculation Method: Umbrella Sampling.** The free energy profile of escape of water from the hydration layer to bulk was evaluated using the umbrella sampling method.<sup>45</sup> Free energy is related to the probability by the following expression:

$$F(q_B) - F(q_A) = -k_B T \ln \left[ \frac{\langle \rho(q_B) \rangle}{\langle \rho(q_A) \rangle} \right]$$

where  $\langle \rho(q_A) \rangle$  is the probability to find the system in state *A* at reaction coordinate  $q_A$ , (and same for  $\langle \rho(q_B) \rangle$  and  $q_B$ ). This definition does not allow the evaluation of free energy (in conventional molecular dynamics simulations) if the two states are separated by high energy barrier. A way out to this problem is to use umbrella sampling which has been employed here.

The umbrella sampling technique was first developed by Torrie and Valleau.<sup>46</sup> In this method, the microscopic system of interest is simulated in the presence of an artificial biasing window potential,  $V(q)$  introduced to enhance the sampling in the neighborhood of a chosen value  $q$ . Typically, the biasing potential serves to confine the variations of the coordinate  $q$  within a small interval around some prescribed value, helping to

achieve a more efficient configurational sampling in this region. For example, a reasonable choice to produce the biased ensembles is to use harmonic functions of the form  $V_i(q) = 1/2K(q - q_i)^2$ , centered on successive values of  $q_i$ . For further technical details of this method see ref 46.

By varying the distance between center of mass point of specific region of the protein and the oxygen atom of the tagged water, a number of MD simulations have been performed. A force constant of 3000 kJ/mol·nm<sup>2</sup> was used to restrain the water at respective distance in the hydration layer from protein interface. We have performed 1 ns long simulations for each umbrella window. The corresponding probabilities of each simulation have been computed and WHAM (weighted histogram analysis) has been applied to estimate the unbiased free energy. Umbrella sampling simulations were performed using pull code from GROMACS version 4.0. In order to calculate the average free energy profile, each water molecule was pulled in both the tangent and normal directions from the protein surface, respectively.<sup>26,47</sup> In the beginning, we selected 12–20 such water molecules around each helix which are not strongly bound to the protein residues. We then computed the free energy profile for each of them and averaged over them to obtain the final free energy surface that we reported here. Subsequently we repeated the same procedure for the doubly hydrogen bonded water molecules which are few in number. We therefore carefully repeated calculations for them with different initial configurations. Two representative profiles among them we have demonstrated in the article.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: bbagchi@sscu.iisc.ernet.in.

## ■ ACKNOWLEDGMENTS

We thank Dr. Prabal Maiti, Dr. Biman Jana, and Mr. Bharat V. Adkar for many useful discussions on this problem. This work was supported in parts by grants from DST and CSIR, India. B.B. acknowledges support from the JC Bose fellowship from DST, India.

## ■ REFERENCES

- (1) Vajda, S.; Jimenez, R.; Rosenthal, S. J.; Fidler, V.; Fleming, G. R.; Castner, E. W. Jr. *J. Chem. Soc., Faraday Trans.* **1995**, 91, 867.
- (2) Bagchi, B. *Chem. Rev.* **2005**, 105, 3197; *Chem. Phys. Lett.* **2011**, doi: 10.1016/j.cplett.2011.12.065.
- (3) Bagchi, B.; Jana, B. *Chem. Soc. Rev.* **2010**, 39, 1936.
- (4) Nandi, N.; Bhattacharyya, K.; Bagchi, B. *Chem. Rev.* **2000**, 100, 2013.
- (5) Bhattacharyya, K. *Acc. Chem. Res.* **2003**, 36, 95.
- (6) Pal, S. K.; Zewail, A. H. *Chem. Rev.* **2004**, 104, 2099–2123.
- (7) Pal, S. K.; Peon, J.; Bagchi, B.; Zewail, A. H. *J. Phys. Chem. B* **2002**, 106, 12376.
- (8) Das, S.; Datta, A.; Bhattacharyya, K. *J. Phys. Chem. A* **1997**, 101, 3299–3304.
- (9) Castrillón, S. R. V.; Giovambattista, N.; Aksay, I. A.; Debenedetti, P. G. *J. Phys. Chem. B* **2009**, 113, 1438.
- (10) Pizzitutti, F.; Marchi, M.; Sterpone, F.; Rossky, P. J. *J. Phys. Chem. B* **2007**, 111, 7584.
- (11) Abbyad, P.; Shi, X.; Childs, W.; McAnaney, T. B.; Cohen, B. E.; Boxer, S. G. *J. Phys. Chem. B* **2007**, 111, 8269–8276.
- (12) Bandyopadhyay, S.; Chakraborty, S.; Balasubramanian, S.; Bagchi, B. *J. Am. Chem. Soc.* **2005**, 127, 4071–4075.
- (13) Bandyopadhyay, S.; Chakraborty, S.; Bagchi, B. *J. Am. Chem. Soc.* **2005**, 127, 16660–16667.

- (14) Bandyopadhyay, S.; Chakraborty, S.; Balasubramanian, S.; Pal, S.; Bagchi, B. *J. Phys. Chem. B* **2004**, *108*, 12608.
- (15) Adkar, B. V.; Jana, B.; Bagchi, B. *J. Phys. Chem. A* **2011**, *115*, 3691–3697.
- (16) Hamelberg, D.; McCammon, J. A. *J. Am. Chem. Soc.* **2004**, *126*, 7683–7689.
- (17) Bhattacharyya, K. *Chem. Commun.* **2008**, 2848.
- (18) Dennis, S.; Camacho, C. J.; Vajda, S. *Proteins: Struct., Funct., Genet.* **2000**, *38*, 176–188.
- (19) Dennis, S.; Camacho, C. J.; Vajda, S. *Optim. Comput. Chem. Molec. Biol.* **2000**, *243*–261.
- (20) Zhang, L.; Hermans, J. *Proteins: Struct., Funct., Genet.* **1996**, *24*, 433–438.
- (21) Roux, B.; Nina, M.; Pomes, R.; Smith, J. C. *Biophys. J.* **1996**, *71*, 670–681.
- (22) Zheng, L.; Lazaridis, T. *Phys. Chem. Chem. Phys.* **2007**, *9*, 573–581.
- (23) Jana, B.; Pal, S.; Maiti, P. K.; Lin, S. T.; Hynes, J. T.; Bagchi, B. *J. Phys. Chem. B* **2006**, *110*, 19611–19618.
- (24) Peon, J.; Pal, S. K.; Zewail, A. H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 10964.
- (25) Pal, S. K.; Peon, J.; Zewail, A. H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 15297.
- (26) Sheu, S. Y.; Yang, D. Y. *J. Phys. Chem. B* **2010**, *114*, 16558–16566.
- (27) McKnight, C. J.; Matsudaira, P. T.; Kim, P. S. *Nat. Struct. Biol.* **1997**, *4*, 180–184.
- (28) McKnight, C. J.; Doering, D. S.; Matsudaira, P. T.; Kim, P. S. *J. Mol. Biol.* **1996**, *260*, 126–134.
- (29) Doering, D. S.; Matsudaira, P. *Biochemistry* **1996**, *35*, 12677–12685.
- (30) Pope, B.; Way, M.; Matsudaira, P. T.; Weeds, A. *FEBS Lett.* **1994**, *338*, 58–62.
- (31) Duan, Y.; Kollman, P. A. *Science* **1998**, *282*, 740–744.
- (32) Duan, Y.; Wang, L.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9897–9902.
- (33) Pal, S.; Balasubramanian, S.; Bagchi, B. *J. Phys. Chem. B* **2003**, *107*, 5194.
- (34) Hummer, G.; Rasaiah, J. C.; Noworyta, J. P. *Nature* **2001**, *414*, 188–190.
- (35) Thirumalai, D.; Klimov, D. K.; Dima, R. I. *Curr. Opin. Struct. Biol.* **2003**, *13*, 1–14.
- (36) Jana, B.; Pal, S.; Bagchi, B. *J. Phys. Chem. B* **2010**, *114*, 3633.
- (37) Bhabha, G.; Lee, J.; Ekiert, D. C.; Gam, J.; Wilson, I. A.; Dyson, H. J.; Benkovic, S. J.; Wright, P. E. *Science* **2011**, *332*, 234–238.
- (38) Kramers, H. A. *Physica* **1940**, *7*, 284.
- (39) Mukherjee, A.; Bagchi, B. *PhysChemComm.* **2003**, *6*, 28.
- (40) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *89*, 6269.
- (41) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.
- (42) Nose, S. *J. Chem. Phys.* **1984**, *81*, 511.
- (43) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182.
- (44) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
- (45) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Academic Press: San Diego, CA, 2002.
- (46) Torrie, G. M.; Valleau, J. P. *Chem. Phys. Lett.* **1974**, *28*, 578–581.
- (47) Godawat, R.; Jamadagni, S. N.; Garde, S. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 15119–15124.