PROTEINS: Structure, Function, and Genetics 39:56–67 (2000)

Molecular Dynamics Simulation of Solvated Azurin: Correlation Between Surface Solvent Accessibility and Water Residence Times

A. Luise, M. Falconi, and A. Desideri*
INFM and Department of Biology, University of Rome "Tor Vergata," Rome, Italy

ABSTRACT **A system containing the globular protein azurin and 3,658 water molecules has been simulated to investigate the influence on water dynamics exerted by a protein surface. Evaluation of water mean residence time for elements having different secondary structure did not show any correlation. Identically, comparison of solvent residence time for atoms having different charge and polarity did not show any clear trend. The main factor influencing water residence time in proximity to a specific site was found to be its solvent accessibility. In detail for atoms belonging to lateral chains and having solvent-accessible surface lower than ~16 Å^2 a relation is found for which charged and polar atoms are surrounded by water molecules characterized by residence times longer than the non polar ones. The involvement of the low accessible protein atom in an intraprotein hydrogen bond further modulates the length of the water residence time. On the other hand for surfaces having high solvent accessibility, all atoms, independently of their character, are surrounded by water molecules which rapidly exchange with the bulk solvent. Proteins 2000;39:56–67.** © 2000 Wiley-Liss, Inc.

Key words: hydration sites; protein-solvent interactions; GROMOS force field; surface shape; hydrophilicity; solvent exposure

INTRODUCTION

The full understanding of the structural and dynamical properties of the solvent–protein interaction is an important task to the comprehension of the protein functionality. Interactions between amino acid residues with their aqueous and protein environment, determine protein folding and mediate intermolecular interactions. On the other hand not only water influences protein mobility folding and function but protein as well can modify water structure and dynamics.[1]

Information on the water position around a protein is provided by X-ray or neutron diffraction experiments on protein crystals.[2–4] Both techniques reveal the favored average positions occupied by water molecules, i.e., the sites in which a water molecule is almost always present although likely rapidly exchanging with bulk water. Several studies at high resolution have now provided a detailed structural analysis of water locations, however surface water are not conserved between different crystal forms introducing some ambiguity in defining the preferred hydrated site and some contradiction has been observed between the X-ray and neutron scattering description.[5,6]

Dynamical information on water behavior is provided by inelastic neutron scattering and NMR spectroscopy.[7–9] Inelastic neutron scattering studies of H_2O-hydrated powders of fully deuterated C-phycocyanin has indicated that water undergoes jump diffusion on the protein surface.[10,11] NMR studies have shown that it is possible to distinguish the dynamics of surface and internal waters in a protein.[7,8] Moreover the residence time of surface water of bovine trypsin pancreatic inhibitor (BPTI) has been measured and a subsequent molecular dynamics (MD) simulation was able to reproduce the order of magnitude of the experimental results; however a comparison between the measured and the simulated residence times didn't show a clear correlation.[8]

MD simulation is a powerful tool for examining protein solvation since it can provide a description at the atomic level and at the appropriate time scale, of the water-protein interaction mechanism. Solvent dynamics and structure has been analyzed by several MD studies on various proteins.[12–14] The diffusion of water has been shown to decrease in proximity to proteins and a similar decrease has been found in proximity to DNA indicating that the mobility of the solvent in such systems is governed by a few universal physical principles.[15] On the other hand more conflicting results are reported concerning the length of water residence time on specific atom types.[8,16–21] This analysis has been carried out on various proteins, polypeptides, and DNA sequences but not unequivocal responses have been obtained on the time needed by water molecules to relax.

In this work we present an analysis of the dependence of water residence times on both the atomic character and the surface solvent accessibility carried out on an MD simulation of hydrated azurin a β-barrel protein containing 128 amino acids. The simulation has been performed on the NPT ensemble and by treating the electrostatics with the Ewald summation method[22] to avoid possible

*Correspondence to: A. Desideri, Department of Biology, University of Rome "Tor Vergata", Via della Ricerca Scientifica, 00133 - Rome, Italy. E-mail: desideri@uniroma2.it

Received 29 June 1999; Accepted 25 October 1999

artifacts on the solvent behavior caused by truncation of the electrostatic forces.[13,23]

The results indicate that the main factor influencing the water residence time is the solvent accessibility of the site and that only for values of solvent-accessible surface (SAS) lower than ~16 Å^2, corresponding to a "re-entrant" surface, it is possible to detect chemical differences and a relation is found for which charged and polar atoms display water residence times longer than the nonpolar ones. In the case of polar or charged atoms the involvement in an intraprotein hydrogen bond lowers the water residence time around them. On the other hand, for surfaces having a high solvent accessibility, all the atoms, independently of their character, influence the solvent in a similar way levelling the water residence times to similar values.

MATERIALS AND METHODS

MD Simulation

The MD simulation of solvated azurin has been performed on a Origin 200 SGI workstation by using the DLPROTEIN simulation package released by Daresbury Laboratory.[24,25] We used the GROMOS force field[26] with the set of parameters denoted "*37c*,"[26,27]and the water molecules were represented by means of the SPC/E model.[28]

The starting coordinates were obtained from the X-ray data of one monomer of Pseudomonas aeruginosa oxidized azurin at pH 5.0[29] (entry 4AZU of the Brookhaven Protein Data Bank[30]), which in the crystals appear in a tetrameric form because of the occurrence of specific protein-protein interactions. A schematic picture of the protein is reported in Figure 1. The crystallographic water molecules were not included in the simulation but the protein has been simulated at high hydration level (5.0 g H_2O/g protein) in order to thoroughly investigate the solute-solvent interactions, filling a truncated octahedral boundary box with 3,658 water molecules. During the addition of the solvent a water molecule was trapped in a buried protein cavity and its behavior was investigated in order to estimate the relaxation time of an internal water molecule. Three water molecules, chosen far enough from opposite charges, were replaced by sodium ions to make the system electroneutral. The equilibrium properties of solvated azurin were sampled in the isobaric-isothermal ensemble by using the Nose'-Hoover technique.[31] The temperature was kept at 300 K, and the pressure at 1.0 atm. A cutoff distance of 10.0 Å has been chosen for the short-range van der Waals interactions, while the electrostatic forces have been determined by using the Ewald sum method,[22] imposing an accuracy of 10^{-5} for the electrostatic energy. The high frequency motions due to the chemical bond potential have been constrained by the SHAKE algorithm,[32] choosing a time step of 1.0 fs and keeping the relative error in the conserved quantity of the NPT ensemble of the order of 10^{-4}. The system has been equilibrated in the NPT ensemble starting from the crystal structure, relaxing the system to 0 K by a steepest descent quenching procedure, and successively increasing temperature up to 300 K. After 100 ps of thermalization, when the time dependence of the potential energy has shown a stationary behavior, a trajectory of 500 ps has been fulfilled collecting one configuration every 0.1 ps.

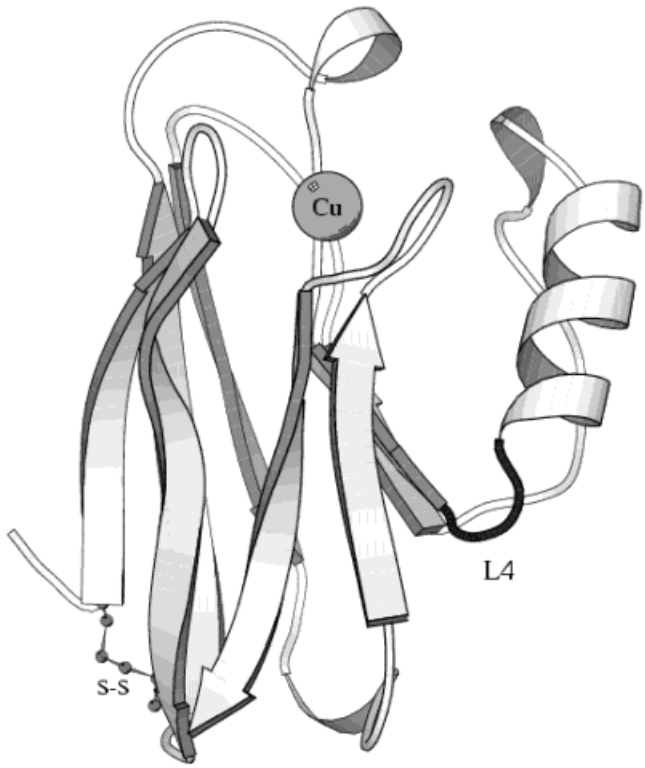

Fig. 1. A graphic representation of the protein azurin produced by using the MolScript v1.4 program.[41] The eight twisted arrows indicate the β-strands forming the barrel, while the unique α-helix of the protein is clearly visible on the right side. The sphere at the top end of the barrel represents the copper atom. The unique disulfide bridge has been evidenced by the S-S label. The black segment on the right side represents the loop L4, which connects the strand S4 with the α-helix.

Structural Analysis

During the MD simulation the secondary structure, the gyration radius and other structural parameters, have been monitored, by using the DSSP program,[33] in order to check the stability of the macromolecule.

The time dependence of the Root Mean Square Deviation (RMSD) from the starting structure, as indication of phase-space accessibility, has been calculated by:

$$RMSD(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} |\mathbf{r}_i^{\min}(t) - \mathbf{r}_i(t_0)|^2} \qquad (1)$$

where the coordinates $\{\mathbf{r}_i^{\min}(t)\}$ are obtained optimally superimposing the instantaneous configurations at time t with the starting structure ($t = t_0$) by removing the global translations and rotations.[34]

The atomic Root Mean Square Fluctuations (RMSF) have been computed by using the following definition:

$$RMSF_i = \sqrt{\sum_{\alpha=1}^{3} \langle (r_{i,\alpha}^{\min}(t) - \bar{r}_{i,\alpha})^2 \rangle_{\mathrm{MD}}} \qquad (2)$$

where the averages have been computed over the equilibrated MD trajectory. This quantity can be directly compared to the temperature factor obtained from X-ray diffraction being proportional to the Debye-Waller factor ($B_i = \frac{8\pi}{3} RMSF_i$).

The standard solvent-accessible surface, defined as the area traced out by the center of a probe sphere (representing a solvent molecule) rolled over the van der Waals surface of the molecule, has been calculated for each protein atom according to the Lee and Richards algorithm.[35]

Water Mean Residence Time

The analysis of water residence time in the first coordination shell of protein atoms has been performed by using the definition given by Impey et al. to describe solvated ions,[36] that has been already applied to protein-water systems by several authors.[8,16,19,20]

The first coordination shell around a protein atom is defined as a sphere of radius $\mathbf{r}_{sh}$:

$$\mathbf{r}_{sh} = \mathbf{r}_{ex} + \mathbf{r}_{OH} + \Delta\mathbf{r}_{RT} \tag{3}$$

where $\mathbf{r}_{ex}$ as determined from the radial distribution function, is the minimum exclusion distance between the solute and the water oxygen atoms, $\mathbf{r}_{OH} = 1.0$ Å is the O-H bond length in a water molecule, and $\Delta\mathbf{r}_{RT} = 0.5$ Å is a correction required to take into account positional fluctuations.

The mean residence time of water around a protein atom, is represented by the mean time that a water molecule spends within the first coordination shell of the selected atom. The water mean residence time for a given atom a is obtained from the "*survival probability function*", defined as follows:

$$P_\alpha(t) = \sum_{j=1}^{N_W} \frac{1}{N - m + 1} \sum_{n=1}^{m} p_{\alpha,j}(t_0, t_0 + t', \Delta t) \tag{4}$$

where $t = m\Delta t$ and $t' = n\Delta t$; the binary function $p_{\alpha,j}$ (t_0, $t_0 + t'$, Δt) takes the value of 1 when the water molecule j resides in the shell α at both times t_0 and $t_0 + t'$, and didn't leave the shell during the time interval t' for a time longer than Δt, otherwise the value assumed by the function is zero; $\Delta t = 0.1$ ps is the configurational data dumping interval, and N is the total number of configurations saved along the MD trajectory, i.e., N = T/Δt; N_w is the number of water molecules in the system.

P_α (t) is a time correlation function representing the average number of water molecules that reside in the coordination shell for a time longer then (or equal to) t. The value assumed by this function for $t = 0$ represents the average number of water molecules residing in the first atomic shell, i.e., it is an estimate of the "*coordination number*" of the atomic site, $N_{co} = P(t = 0)$. This calculation was accomplished through an efficient implementation of the algorithm given by Garcia.[16]

In the case of water molecules solvating simple ions[36] the survival function can be well approximated by a single exponential function:

$$P_\alpha(t) \approx P_0 e^{-t/\tau} \tag{5}$$

where the relaxation time t represents the water mean residence time of the atom. In the case of atomic shells on the protein surface, due to the strong anisotropy of the protein environment, the functional form of $P_\alpha(t)$ is better approximated by a multi-exponential function, defining more than one relaxation time.[16,19] Since such an approach, introducing many degrees of freedom, may lead to an overinterpretation, we have estimated, according to other authors,[16,19] the average residence time by fitting the survival function with the single relaxation time function provided by equation.[5]

RESULTS AND DISCUSSION

Protein Dynamics

To assess the reliability of the simulation, and as measure of structural stability, we performed a series of standard tests on our trajectory. Of the total 600 ps of MD simulation only the last 500 were used for the analysis. The protein all atoms RMSD is stable and lower than 2.5 ~ 3.0 Å over all the trajectory. The RMSF calculated on the same trajectory and averaged over each residue has been reported in Figure 2, in comparison with the RMSF derived from crystallographic B-factors. The two plots hold a similar trend and show the main peaks in the same position. The atomic fluctuations derived from the MD simulation are slightly larger than those evaluated from the crystal structure, likely because of the different hydration degree and/or of crystal packing effects.

A geometrical analysis was carried out by measuring the accessible surface area, the gyration radius and the secondary structure percentage. The first two structural parameters (Fig. 3a,b) are fairly stable and their average values are strictly close to the ones evaluated from the X-ray structure (Table I). The percentage of secondary structure, plotted as a function of time, is also stable (Fig. 3c) and shows a stationary content of β structure, indicating the conservation of the β-barrel motif of azurin and then of the overall 3D-fold of the protein. Comparison of the averaged secondary structure values with those observed in the crystal structure (see Table I) indicate that, during the simulation, an identical percentage of helix is maintained while a slight reduction of β structure, coupled to a concomitant increase of the random coil, is observed. Such a structural variation is however compatible with the maintaining of the β-barrel motif.[27]

Water Residence Time Analysis and Secondary Structure

The distribution of water residence times within the first coordination shell of each azurin atom is shown in Figure 4. The majority of the atomic sites is characterized by water residence time lower than 20 ps, while the overall distribution is two order of magnitude larger. There are few atoms (less than 20, data not shown) characterized by a residence time longer than 100 ps. They are atoms for which the survival function has not yet dropped to zero at

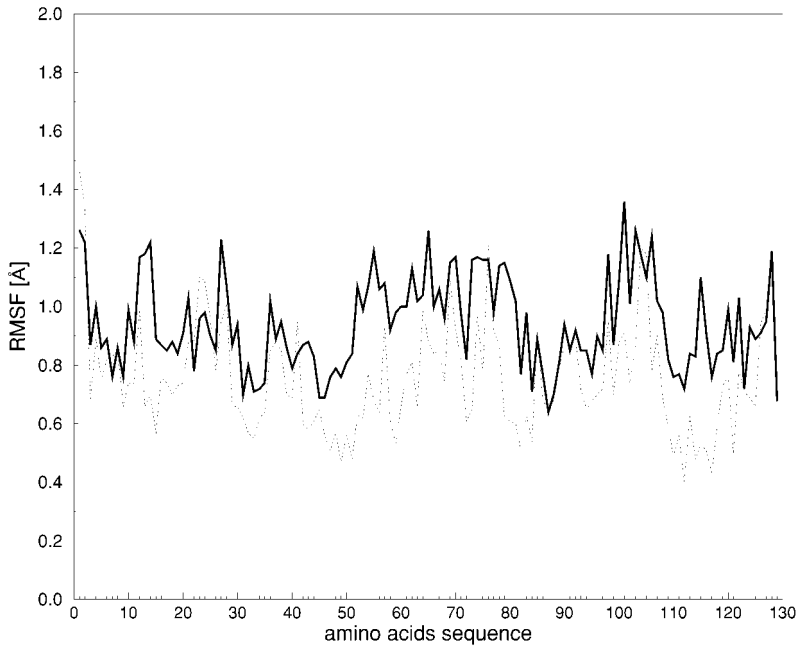

Fig. 2. Root mean square fluctuations of each azurin amino acid calculated from the simulation (solid line) and obtained from the X-ray Debye-Waller factors (dotted line).

t = 100 ps, indicating that, during this period of time, at least one water molecule has never exchanged with bulk solvent. These water molecules are sterically trapped into holes or deep crevices of the protein surface which slow down the ability of the water molecules to exchange with the bulk solvent. It is interesting to notice that the water molecule trapped in a buried protein cavity during the water box preparation (see computational methods), remains there, hydrogen-bonded to the hydroxyl group of Ser51 almost for all the trajectory, getting out of the protein core only after 480 ps.

The mean residence times of both the side chains and backbone atoms, averaged over the secondary structure elements of azurin, are reported in Fig. 5. The two plots show a similar trend and have the main peak in the same position. The residence time values calculated for the secondary structural elements of azurin are similar, except for the large values observed in the region included between residue Val49 and Asp55, that accommodates β-strand 4 (S4) and the loop 4 (L4), connecting the carboxyl-terminal of S4 with the amino-terminal of the unique α-helix. A high contribution to the residence time

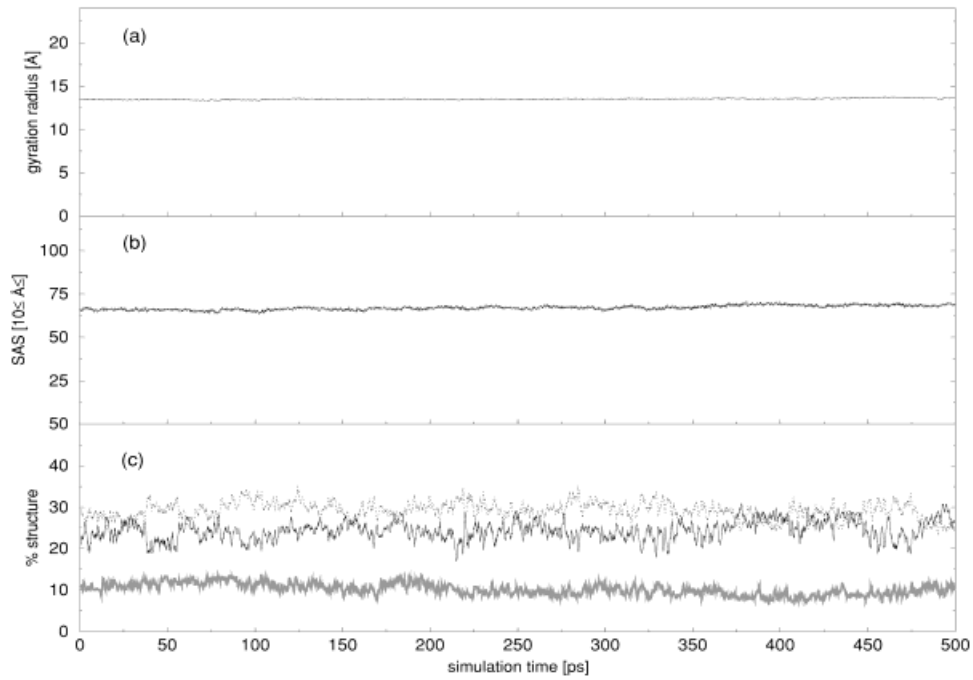

Fig. 3. Time behavior of some geometrical properties during the entire simulation. **(a)** Gyration radius. **(b)** Total solvent-accessible surface. **(c)** Secondary structure percentages: alpha helix (grey line), β-strands (black line), and random coil (dotted line).

TABLE I. MD Averages of Some Geometrical Properties of Azurin Compared with the Corresponding X-Ray Values†

	MD	X-Ray
SAS [Å^2]	67 (13)	67
Gyration radius [Å]	13.53 (0.08)	13.8
β-structure percentage	24.4 (3.6)	34.0
a-helix percentage	10.2 (2.0)	12.3
Random coil percentage	29.1 (3.1)	22.1

†Standard deviations in parentheses.

value of strand S4 is given by the internal water molecule that permanently contact the hydroxyl group of Ser51. On the other hand the flexibility of the region including L4, as evidenced by the RMSF plot reported in Figure 2, contributes to the formation of cavities where some water molecules remain trapped for long times. The region from Thr52 to Val80, including S4, L4 and the α-helix, represents the unique large protrusion of the polypeptide chain from the β-barrel of azurin. In this protrusion the guanidinium group of Arg79, bound through salt bridges to Asp62 and Asp77, act as an hydrophilic cap that regulates with its motions the entrance of water in the protein pocket formed by L4 and the helix from one side and the β-barrel from the other.

Some differences in the water residence times is observed between the backbone and side chain atoms belonging to the β-strands (see Fig. 5 and Table II). In this case the main chain atoms exhibit residence times shorter than those of the side chains. This behavior is likely due to the involvement of the backbone atoms of the antiparallel β-strands in inter-strands hydrogen bonds which reduce interaction of these atoms with the external water molecules. Notwithstanding, as may be inferred by looking at the similar values reported in Table II, the secondary structure does not help us in classifying the water residence times. This result is in agreement with previous findings[20] indicating that the values of residence times found for well-defined secondary structure segments are approximately equal to those found for polypeptide segments with non-repetitive structure.

Water Residence Time and Atomic Species

An analysis of the residence times taking in consideration different atom types, has been also carried out. The atoms belonging to the lateral chains and to the backbone have been divided in five and two classes respectively. The following seven groups have been considered:

1) polar oxygen (OH) from serine, tyrosine and threonine;
2) polar nitrogen (NH_2) from asparagine and glutamine;
3) charged oxygen (COO^-) from aspartic and glutammic acids;
4) charged nitrogen from lysine (NH_3^+);
5) side chain carbon atoms (CH_2 and CH_3);
6) polar oxygen (CO) from backbone;
7) polar nitrogen (NH) from backbone.

The average values of the water mean residence time and the coordination number, obtained for each group, have been reported in Table III. In the lateral chains, oxygen atoms show residence times longer than nitrogen ones while the presence of a charge or a dipole moment does not seem to play a dominant role in determining the values of the residence times. On the other hand nitrogen atoms have coordination numbers higher than oxygen ones and the presence of a charge increases the coordination number in both nitrogen and oxygen atoms. An opposite relation is found for backbone atoms (Table III).

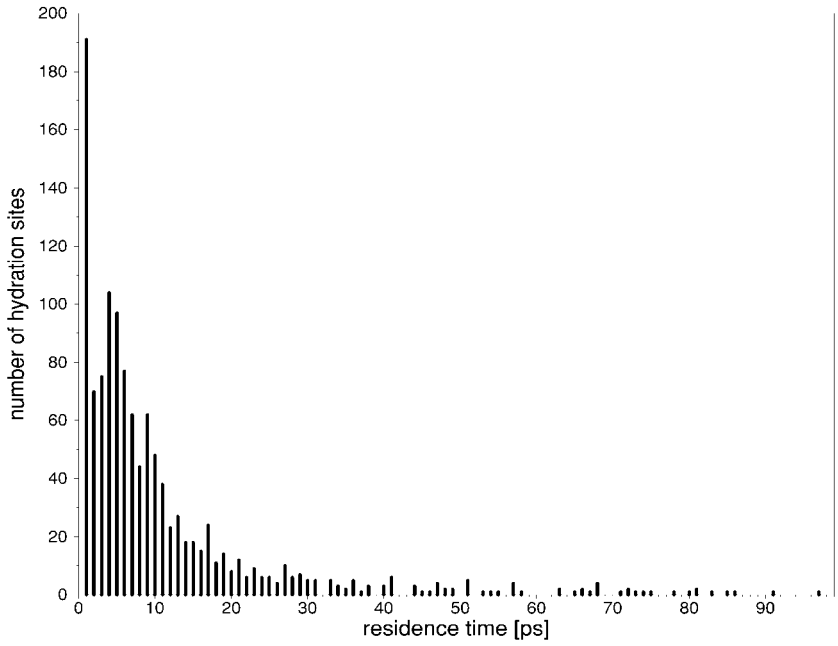

Fig. 4. Distribution of water mean residence times within the first solvation shell of each azurin atom.

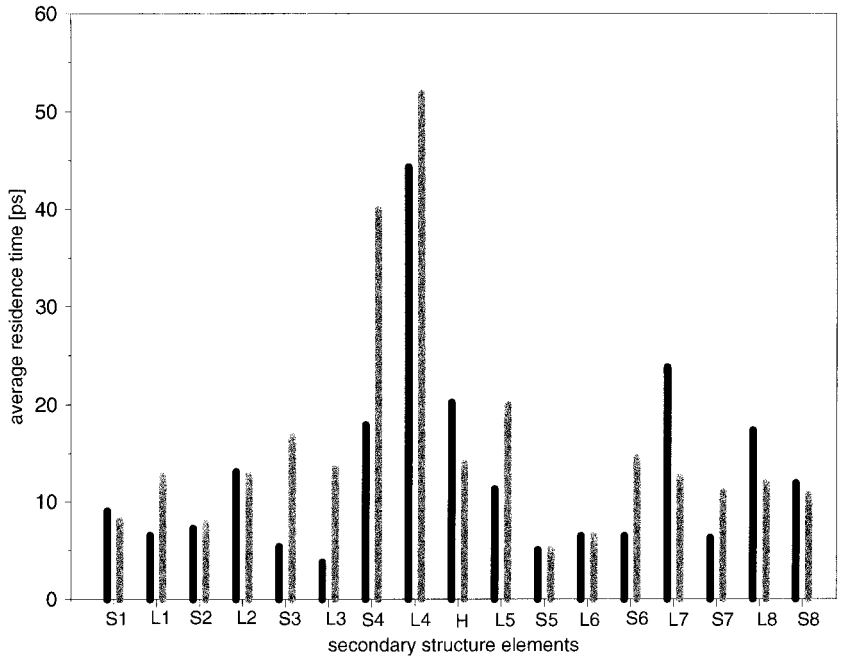

Fig. 5. Water mean residence times of the side-chains (grey bars) and backbone atoms (black bars) averaged over each secondary structure element. β-strands, α-helix and loops are indicated with the letters S, H, L respectively.

TABLE II. Mean Residence Times (τ) Averaged Over Side Chain and Backbone Atoms Belonging to Different Secondary Structure Elements[†]

	β-Strands	a-Helix	Loops
τ Backbone [ps]	10 (13)	20 (40)	16 (28)
τ Side chain [ps]	15 (45)	14 (12)	15 (26)

[†]Standard deviation are reported in parentheses.

TABLE III. Mean Residence Times (τ) and Coordination Numbers (N_{co}), With Their Standard Deviations, Averaged Over Atoms Belonging to the Atomic Groups Indicated in Parentheses

	τ [ps]	(σ_τ)	N_{co}	(σ_{Nco})
Side chain polar O (OH)	33	(28)	3.6	(0.9)
Side chain charged O (COO^-)	22	(21)	4.4	(1.0)
Side chain charged N (NH_3^+)	18	(21)	6.1	(0.9)
Backbone polar N (NH)	18	(16)	2.4	(0.6)
Side chain apolar C (CH_2, CH_3)	14	(26)	5.4	(2.0)
Backbone polar O (CO)	12	(18)	3.1	(1.1)
Side chain polar N (NH_2)	9	(12)	4.4	(1.3)

As far as the average residence times are concerned the following ranking relation is observed:

$$\tau_{OH} > \tau_{O^-} > \tau_{N^+} = \tau_{NH} > \tau_{CH_2/CH_3} > \tau_{CO} > \tau_{NH_2}$$

Such an order is similar to the hydrophilicity scale derived by analyzing the closest protein-water contacts in 56 high-resolution crystallographic structures.[37] It must be recalled however that crystallographic well-ordered water molecules are not necessarily residing for a long time in a specific region, but they are molecules whose potential of mean force has a local minimum at a given point, which permits a defined peak in the electron density map. Comparison of the location of water sites having a long residence time with the azurin crystallographic waters having a low B-factor is not straightforward since in the tetrameric crystal form each azurin molecule is partially masked from the solvent by the other subunits.[29] However we have observed that for atomic sites located in cavities or crevices and having average solvent accessibilities similar to those observed in the crystal structure, there is a positive correlation between crystallographic water sites with low B-factor and long residence times. For example the crystallographic water bridging the backbone oxygen of Phe114 and the carboxyl oxygen of Asp77, shows a low B-factor (12 $Å^2$) and the residence time calculated for these atoms is of the order of 100 ps. The same correlation is observed for a crystallographic water (B = 13 $Å^2$) that bridges the carboxyl oxygen of Asp77 with the backbone oxygen of Val59, both the atoms showing long residence times (τ = 47 ps and τ = 250 ps respectively).

A comparison of our relation with previous works evaluating water residence times in different proteins[8,16–19] is not straightforward because of the difference in determining such a values (i.e., in the definition of mean residence time or in the choice of parameters like the "resolution" time interval Δt and the coordination shell radius, and in the definition of the class of atoms). As a further complication some authors[19] have reported their analysis dividing the amino acids according to their electrostatic character and then averaging the residence time values over the side-chain rendering the comparison even more difficult. Taking in consideration these cautions the relation found in this work for the atoms belonging to the lateral chains is in quite good agreement with the results reported for crambin:[16]

$$\tau_{charged} > \tau_{polar} > \tau_{non\text{-}polar}$$

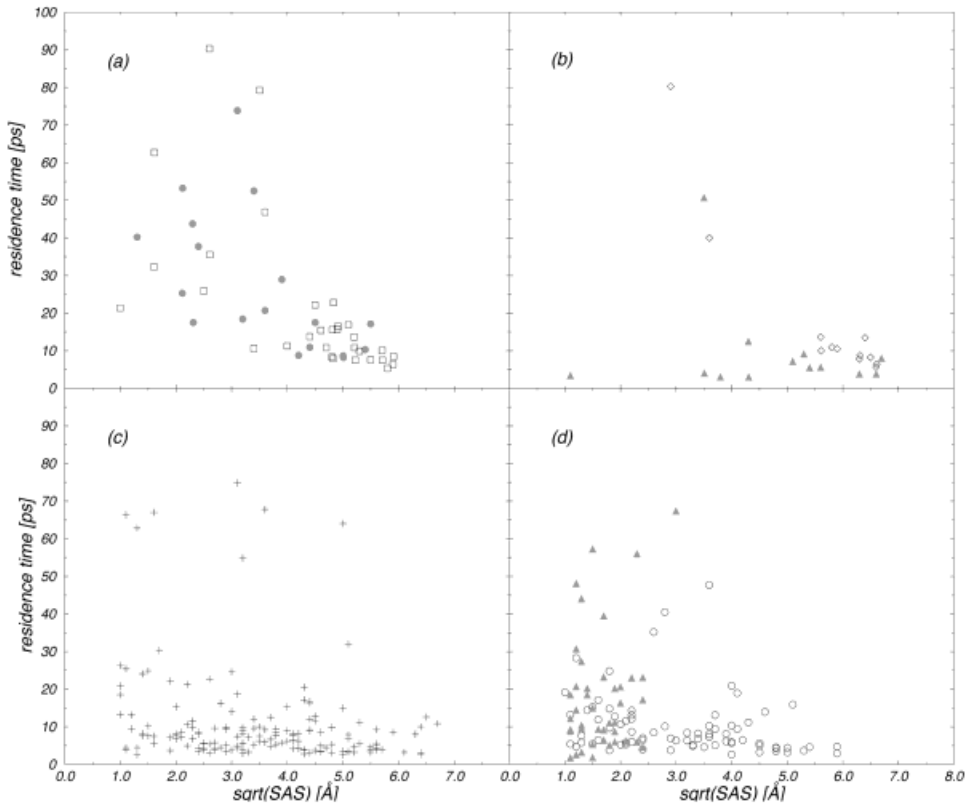

Fig. 6. Distribution of water mean residence times (ps) of seven selected atom types as a function of the square root of their average solvent accessible surface (Å). **(a)** Side chain oxygen atoms: grey filled circles and empty squares represent polar hydroxyl oxygen and charged carboxyl oxygen respectively. **(b)** Side chain nitrogen atoms: grey filled triangles and empty diamonds represent polar amidic nitrogen and charged aminic nitrogen respectively. **(c)** Side chain non-polar carbon atoms. **(d)** Backbone polar atoms: grey filled triangles and empty circles represent nitrogen and oxygen respectively.

and differ from the one reported for BPTI:[8]

$$\tau_{polar} > \tau_{non\text{-}polar} > \tau_{charged}$$

which however is somewhat in contrast with the interfacial water self-diffusion coefficient D evaluated in the same work.

The large variances found for the average residence time for different atom types (see Table III) suggests that other factors like chemical environment, surface shape, or solvent-accessibility influence the local dynamics of water. This is in agreement with a previous analysis of water residence times on two peptides having hydrophobic and hydrophilic character, which has proposed that not the atomic type but the local environment has a primary role in determining the solvent residence time.[20]

Water Residence Time and Solvent Accessibility

In order to check how solvent accessibility of an atomic site influences its hydration, the water residence time for each atom belonging to the seven atomic groups previously defined has been evaluated and reported in a plot as a function of the square root of their average solvent-accessible surface (Fig. 6a–d). The water residence time distribution for side-chain oxygen atoms (groups 1 and 3), side-chain nitrogen atoms (groups 2 and 4), side-chain carbon atoms (group 5) and backbone atoms (groups 6 and 7) is reported in Figure 6a–d. A general feature can be extracted from these figures: sites having solvent-accessible surface $\geq$ 16 Å^2 have in their coordination sphere, independently of their polarity, water molecules characterized by a residence time $<$ 20 ps. This implies that for solvent-accessible surface $\geq$ 16 Å^2 the various atom types are indistinguishable as far as the length of their residence time is concerned.

The longest residence times are observed for solvent-accessible surface between 1 and 16 Å^2. This is particularly evident for the atoms belonging to the lateral chains and in particular for the charged nitrogen (Fig. 6b) and for both the polar and the charged oxygen (Fig. 6a). A low accessible atom corresponds to a site located at the bottom of holes or crevices on the protein surface, hence the shape of the protein in proximity to a partially-accessible site protects the water molecule that is in contact with this atom, shielding the site and reducing the rate of exchange with the bulk solvent.[38]

This relation is not fully respected for the atoms belonging to the backbone. The residence times of water near these atomic groups for SAS values $<$ 16 Å^2 are much more scattered and a unique relation cannot be extracted, indicating that others factors influence the water residence time. It is anyway comfortable that for the backbone CO group, the water residence time at sites with solvent-accessible surface $>$ 16 Å^2 is short, as found for the atoms belonging to the lateral chains. This relation cannot be verified for the backbone NH group, since all these sites have a solvent accessibility below 10 Å^2.

Water in the proximity to an apolar carbon atom is characterized, even when the carbon atom has a low

accessible surface, by short residence times, i.e., faster motions, when compared to water close to a predominantly charged or polar (oxygen or nitrogen) low accessible surface atom. This result indicates that a hydrophobic surface offers a low friction to the water, independently of its solvent-accessible surface.

An analysis of 56 high-resolution structures has related the distribution of surface waters to protein surface topography.[39] The authors indicate that ordered waters are three times more likely to be in surface grooves than elsewhere on the surface. Moreover intrinsic chemical differences between binding sites were observed only in the grooves, where water molecules have a preference for hydrophilic location, while non-groove waters show less discrimination between polar and non-polar groups. It is interesting to notice that X-ray and MD, which observe different aspect of the same phenomenon, provide the same conclusion. In particular a peak in the electron density map indicates the overall probability of finding a water molecule in a given position, while only MD (or NMR) may estimate the exchange rate with the bulk. The agreement of MD and X-ray diffraction results, at least for the atoms belonging to the lateral chains, indicates that water molecules have a preference in both position and residence time for polar and charged atoms in crevices, which are then the preferred hydration sites in a protein.

Analysis of Water Residence Time Around Specific Atoms

The analysis carried out in the previous paragraph has indicated that sites having a large accessible surface have in their coordination shell water molecules characterized by short residence times. The description of sites having low solvent-accessible surface is more complex. Charged nitrogen and both polar and charged oxygens belonging to the lateral chains display a general trend (low SAS $\rightarrow$ long τ) at variance on what found, beside some exceptions, for the apolar carbons which display a short τ independently of their accessible surface (see Fig. 6). However the values of the water residence times are widely scattered for the backbone oxygen and nitrogen atoms having low SAS values. To give a rationale for this behavior we have singly analyzed these sites. Such an analysis permits us to state that another factor strongly influences the water residence time in proximity to a low accessible group, i.e., the involvement of this site in an intraprotein hydrogen bond. To make this point clear we describe some sites, as observed in a single frame, belonging to the groups represented in Figure 6a-d, having solvent-accessible surface in the range 1–16 Å^2 and characterized by long and short residence times.

OH

As it is shown in Figure 6a, hydroxyl oxygens having a low SAS are surrounded by water molecule residing in their shell for a long time. However the hydroxyl oxygen of Thr124 have one of the shortest residence time in this SAS range (τ = 18 ps; SAS = 7 Å^2). This atom is involved in an intraprotein hydrogen bond with the lateral carbonyl oxygen of Gln107, that partially screens it from the solvent (see Fig. 7). The water closest to this site points toward the solvent and does not interact with any other protein atom. On the other hand the hydroxyl oxygen of Tyr108 which is characterized by τ = 52 ps and SAS = 7 Å^2, lies in a small crevice and is not involved in any intraprotein hydrogen bond, as represented in Figure 8. Two water binding sites are discernible, in one of them the water molecule forms a bridge between this atom and the backbone nitrogen of Lys103 (τ = 31 ps).

COO$^-$

We describe two examples of carboxyl sites, those of Asp11 and Asp71, whose oxygens show a short (τ = 15 ps; SAS = 3 Å^2) and a long (τ = 90 ps; SAS = 7 Å^2) residence time respectively. The two oxygens of the carboxyl group of Asp11 are hydrogen-bonded to the backbone nitrogen atoms of Asn38 and Leu39 respectively. Because of this bond only one oxygen weakly interacts with a solvent water molecule which is then characterized by a short residence time. On the other hand the oxygens of the carboxyl group of Asp71 are not involved in any intraprotein hydrogen bond and reside in the center of a cavity where a long-lived water molecule bridges them with the backbone carbonyl group of the same residue.

NH$_2$

There are few amidic sites with low solvent-exposed surface in the protein (Fig. 6b). Here we describe two of them showing an opposite behavior as far as the interaction with the water molecules is concerned. The amidic nitrogen of Asn47 (τ = 4 ps; SAS = 1 Å^2) is hidden in a cavity, and is hydrogen-bonded to the hydroxyl group of both Thr84 and Thr113. As already noticed for the other atoms, the occurrence of intraprotein hydrogen bonds makes the group only partly available for an interaction with the solvent molecules. As a consequence the average residence time of the water molecules around it is short. On the other hand only one hydrogen of the amidic group of Gln57 (τ = 51 ps; SAS = 13 Å^2) is bonded with the hydroxyl oxygen of Thr61, while the other is available for a long lasting interaction with a water molecule that bridges the amidic group and the sulfur of Met64.

NH3$^+$

There are only two aminic sites with solvent accessible surface $<$ 16 Å^2, both characterized by long water residence times. Here we report the example of the aminic group of Lys24 (τ = 80 ps; SAS = 8 Å^2) which points toward the interior of a small cavity of the protein trapping a single water molecule inside of it. This water is hydrogen-bonded to the aminic nitrogen of Lys24 and to the backbone oxygens of both Val22 and Val99, which are the only non-hydrophobic atoms building the cavity. At the top end of this cavity another water molecule is hydrogen-bonded to the aminic group of Lys24 for a long time, bridging this atom with the carbonyl group of Lys128.

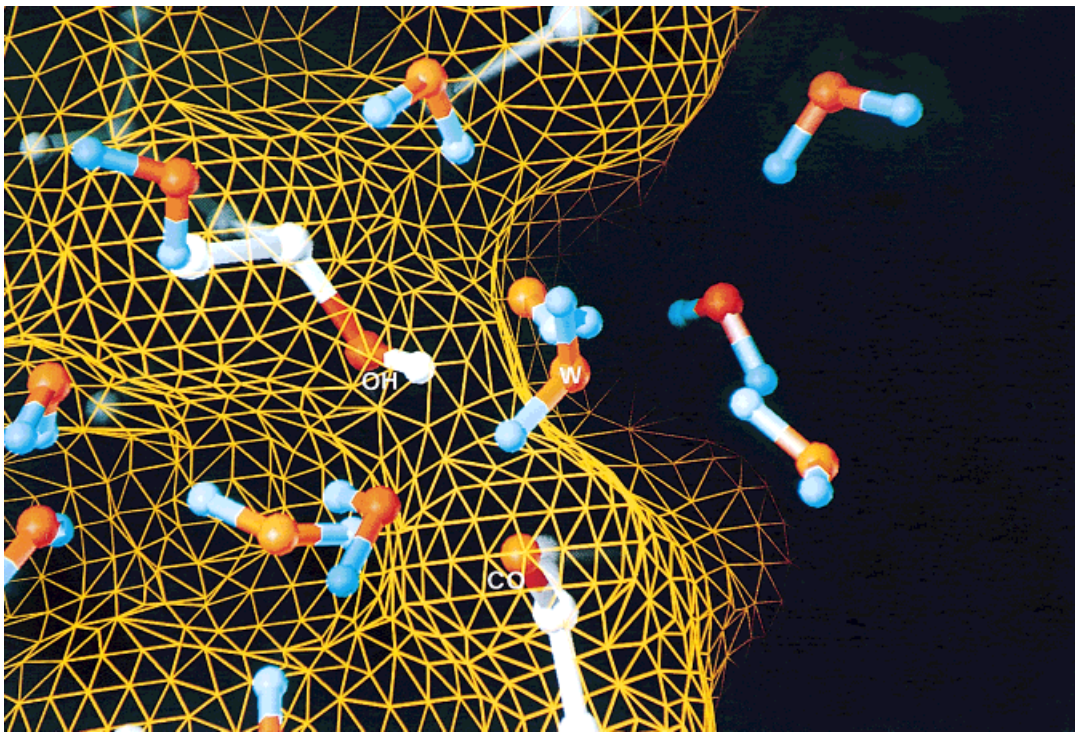

Fig. 7. Enlarged view of the environment including the hydroxyl group of Thr124, characterized by a short water residence time. The hydroxyl group of Thr124, labeled with OH, is hydrogen-bonded to the carbonyl oxygen of Gln107, labeled with CO. A water molecule, labeled with W, is hydrogen-bonded to the hydroxyl group and points toward the bulk. The oxygens of these groups are displayed in red and the other protein atoms in white. Oxygen and hydrogen atoms of water molecules are represented in red and cyan respectively.

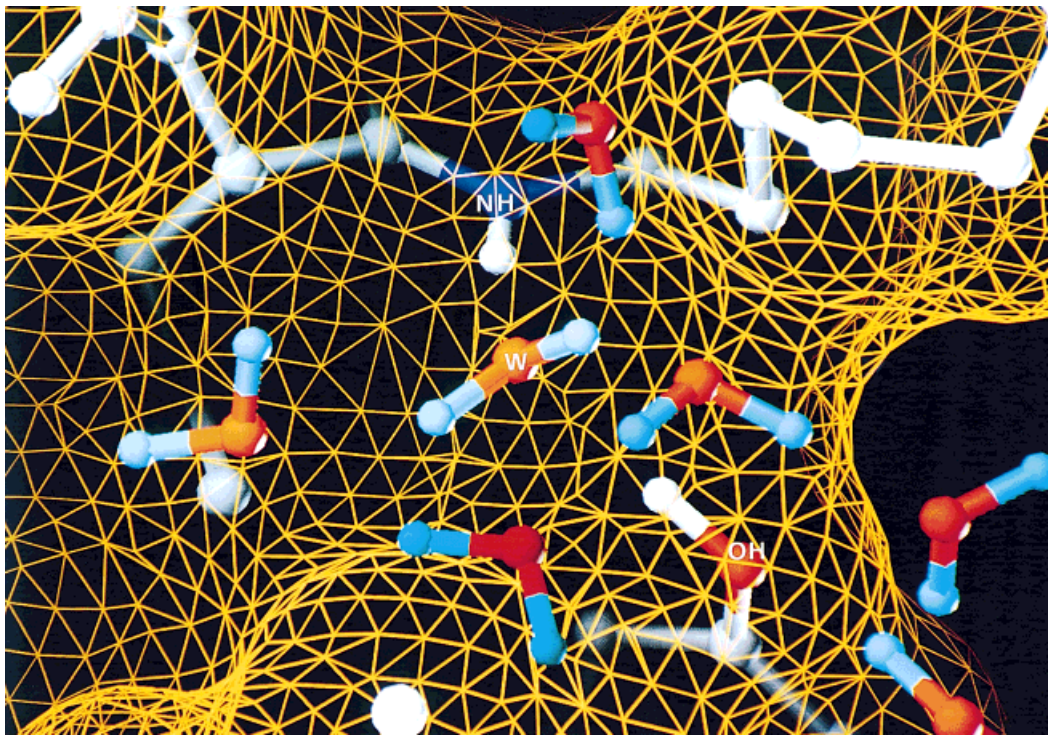

Fig. 8. Enlarged view of the environment including the hydroxyl group of Tyr108, characterized by a long water residence time. The hydroxyl group, labeled with OH, is hydrogen-bonded to a water molecule, labeled with W, that bridges this group to the backbone nitrogen of Lys103, labeled with NH. The oxygens of the hydroxyl group is displayed in red, the backbone nitrogen in blue and the other protein atoms in white. Oxygen and hydrogen atoms of water molecules are represented in red and cyan respectively.

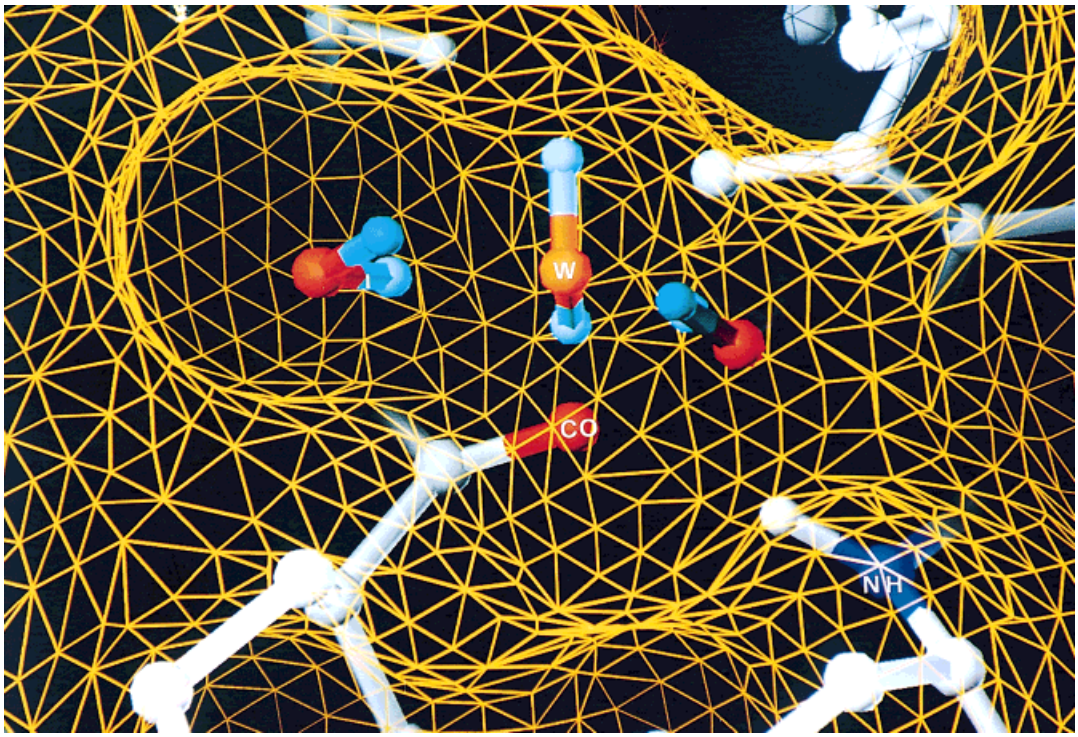

Fig. 9. Enlarged view of the environment including the carbonyl group of Gln8, characterized by a short water residence time. The carbonyl group, labeled with CO, is hydrogen-bonded to the backbone nitrogen of Asn16, labeled NH. A water molecule, labeled with W, is hydrogen-bonded to the carbonyl group and points toward the bulk. The oxygens of these groups are displayed in red, the backbone nitrogen in blue and the other protein atoms in white. Oxygen and hydrogen atoms of water molecules are represented in red and cyan respectively.

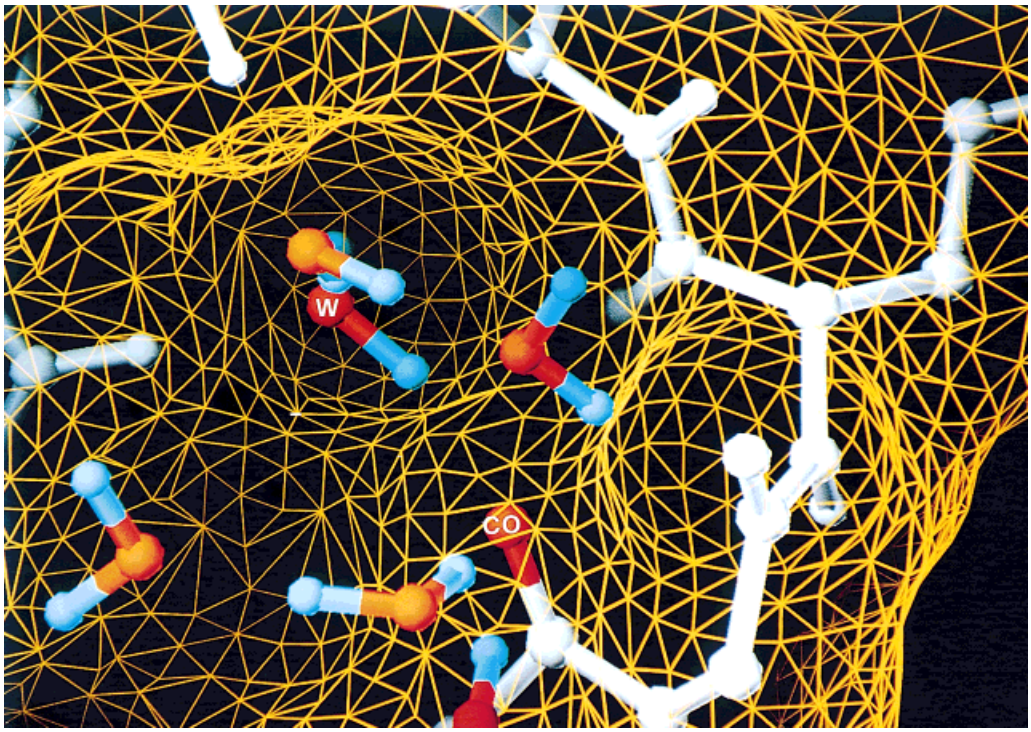

Fig. 10. Enlarged view of the environment including the carbonyl group of Glu104, characterized by a long water residence time. The carbonyl group, labeled with CO, is hydrogen-bonded for a long time to a water molecule, labeled with W. The oxygens of the hydroxyl group is displayed in red and the other protein atoms in white. Oxygen and hydrogen atoms of water molecules are represented in red and cyan respectively.

CO

The backbone carbonyl oxygens are characterized by having in their coordination shell water molecules with residence time values scattered over a wide range. Here we describe two carbonyl groups, that of Gln8 and that of Glu104, having comparable SAS (SAS = 6 Å^2 and SAS = 13 Å^2 respectively) but very different residence times (τ = 7 ps and τ = 48 ps respectively). The reason for this different behavior must be found in the presence or absence of an intraprotein hydrogen bond. The carbonyl of Gln8 which is hydrogen-bonded with the NH group of Asn16 has a residence time τ = 7 ps (see Fig. 9). On the other hand the carbonyl group of Glu104, not involved in intraprotein hydrogen bonds, is situated in a cavity where a water molecule, bridging it to the backbone NH group of Leu127, resides for a long time (see Fig. 10).

NH

The reason for the wide scattered values of water residence time around the NH group is again correlated with the fact that part of them participate in intraprotein hydrogen bonds that prevent them from the interaction with water molecules. As an example, the NH group of Lys85 that is hydrogen-bonded to the carboxyl group of Asp93 shows a very short residence time (τ = 2 ps; SAS = 2 Å^2), while the backbone nitrogen of Gly116 that is not involved in any intraprotein hydrogen bond shows a long residence time (τ = 70 ps; SAS = 9 Å^2). In this case a water molecule bridges this backbone nitrogen with the backbone oxygen of Cys112 (τ = 80 ps; SAS = 4 Å^2).

CH_2-CH_3

As we have already described the apolar CH_2 group are surrounded by water molecules having short residence times independently of their SAS. Here we describe a non-polar site which does not obey to this rule. In fact the Cβ of Pro40, although quite accessible to the solvent (SAS = 17 Å^2), has in its coordination shell water molecules showing an average residence time of 70 ps. This is due to the fact that this atom, lying at the top of a small crevice built by polar atoms, is in contact with a water molecule that forms stable hydrogen bonds with both carboxyl group of Glu91 and backbone nitrogen of Lys41, residing at the top and at the bottom of this crevice respectively.

CONCLUSIONS

The dynamical behavior of water close to the protein surface has been investigated through an MD simulation of hydrated azurin as a function of several parameters. The analysis indicates that non-polar atomic sites display short water residence time independently of their solvent accessibility. Water residence time for atom located on an exposed surface is short and, as a first approximation, does not depend on the surface polarity. On the other hand for atoms having a reduced accessibility, i.e., atoms in grooves or crevices, the residence time values depend on the polarity of the atom itself and/or on its involvement in intraprotein hydrogen bonds.

As long ago as 1988, MD simulations indicated that water molecules close to polar atoms have a residence time longer than the ones close to any non-polar atom.[17] On the other hand not always the same order was found in more recent works.[8,16,19,20] Our finding that intrinsic chemical differences modulate water residence time only in grooves may likely reconcile the conflicting order proposed in previous works since this order may be influenced by the number of grooves and cavities present in the various proteins. In agreement with our results, a thermodynamic perturbation theory calculation of the affinity of water for cavities has indicated favorable and unfavorable interactions when they are surrounded by polar and apolar atoms respectively.[40] Water residence time depends, as previously stated,[38] on the free energy of the transition state, which is related with both the strength of the binding (entalpic contribution) and the geometry of the site (entropic contribution). In a wide pocket or a flat surface a bulk water may easily replace a protein-interacting water independently of the polarity of the surface. On the other hand in a narrow pocket the exchange is more difficult and such difficulty increases with the strength of the water-protein interaction, i.e., with the degree of the surface polarity and with the number of hydrogen bonds the atom could form. NMR measurements of water residence times on proteins or model materials, having a different degree of polarity and roughness, are needed to confirm such a description of the solute-solvent interaction.

REFERENCES

1. Phillips Jr. G, Pettitt BM. Structure and dynamics of water around myoglobin. Protein Sci 1995;4:149–158.
2. Teeter MM. Water-protein interactions: theory and experiment. Annu Rev Biophys Biophys Chem 1991;20:577–600.
3. Karplus PA, Faerman C. Ordered water in macromolecular structure. Curr Opin Struct Biol 1994;4:770–776.
4. Burling FT, Weis WI, Flaherty KM, Brunger AT. Direct observation of protein solvation and discrete disorder with experimental crystallographic phases. Science 1996;271:72–77.
5. Kossiakoff AA, Sintchak MD, Shpungin J, Presta LG. Analysis of solvent structure in proteins using neutron D2O-H2O solvent maps: Pattern of primary and secondary hydration of trypsin. Proteins 1992;12:223–236.
6. Finer-Moore JS, Kossiakoff AA, Hurley JH, Earnest T, Stroud RM. Solvent structure in crystals of trypsin determined by X-ray and neutron diffraction. Proteins 1992;12:203–222.
7. Otting G, Liepinsh E, Wuthrich J. Protein hydration in aqueous solution. Science 1991;254:974–980.
8. Brunne RM, Liepinsh E, Otting G, Wutrich K, van Gunsteren WF. Hydration of proteins. A comparison of experimental residence times of water molecules solvating the bovine pancreatic trypsin inhibitor with theoretical model calculations. J Mol Biol 1993;231:1040–1048.
9. Zanotti JM, Bellissent-Funel MC, Parello J. Hydration-coupled dynamics in proteins studied by neutron scattering and NMR: The case of the typical EF-hand calcium-binding parvalbumin. Biophys J 1999;76:2390–2411.
10. Bellissent-Funel MC, Lal J, Bradley KF, Chen SH. Neutron structure factors of in vivo deuterated amorphous protein C-phycocyanin. Biophys J 1993;64:1542–1549.
11. Bellissent-Funel MC, Zanotti JM, Chen SH. Slow dynamics of water molecules on the surface of a globular protein. Faraday Discuss 1996;103:281–294.
12. Komeiji Y, Uebayasi M, Someya J, Yamato I. A molecular dynamics study of solvent behavior around a protein. Proteins 1993;16:268–277.
13. Abseher R, Schreiber H, Steinhauser O. The influence of a protein

on water dynamics in its vicinity investigated by molecular dynamics simulation. Proteins 1996;25:366–378.
14. Bizzarri AR, Cannistraro S. Molecular dynamics simulation evidence of anomalous diffusion of plastocyanin hydration water. Phys Rev E 1996;53:3040–3043.
15. Makarov VA, Feig M, Andrews BK, Pettitt BM. Diffusion of solvent around biomolecular solutes: A molecular dynamics simulation study. Biophys J 1998;75:150–158.
16. Garcia AE, Stiller L. Computation of the mean residence time of water in the hydration shells of biomolecules. J Comput Chem 1993;14:1396–1406.
17. Levitt M, Sharon R. Accurate simulation of protein dynamics in solution. Proc Natl Acad Sci USA 1988;85:7557–7561.
18. Muegge I, Knapp EW. Residence times and lateral diffusion of water at protein surfaces: application to BPTI. J Phys Chem 1995;99:1371–1374.
19. Rocchi C, Bizzarri AR, Cannistraro S. Water residence times around copper plastocyanin: a molecular dynamics simulation approach. Chem Phys 1997;214:261–276.
20. Kovacs H, Mark AE, van Gunsteren WF. Solvent structure at a hydrophobic protein surface. Proteins 1997;27:395–404.
21. Bonvin AMJJ, Sunnerhagen M, Otting G, van Gunsteren WF. Water molecules in DNA recognition II: a molecular dynamics view of the structure and hydration of the trp operator. J Mol Biol 1998;282:859–873.
22. Frenkel D, Smit B. Understanding molecular simulation. From algorithms to applications. San Diego: Academic Press; 1996. 347 p.
23. Alper HE, Bassolino-Klimas D, Stouch TR. The limiting behavior of water hydrating phospholipid monolayer: A computer simulation study. J Chem Phys 1993;99:5547–5559.
24. Smith W, Forester TR. DL_POLY_2.0: A general-purpose parallel molecular dynamics simulation package. J Mol Graph 1996;14: 136–141.
25. Melchionna S, Luise A, Venturoli M, Cozzini S. DLPROTEIN: A molecular dynamics package to simulate biomolecules. In: Voli M, editor. Science and supercomputing at CINECA - 1997 report. Supercomputing Group, CINECA; 1998. p 496–505.
26. van Gunsteren WF, Berendsen HJC. Groningen Molecular Simulation (GROMOS) Library Manual. Groningen: Biomos; 1987.
27. Melchionna S, Falconi M, Desideri A. Effect of temperature and hydration on protein fluctuations: molecular dynamics simulation of Cu, Zn superoxide dismutase at six different temperatures. Comparison with neutron scattering data. J Chem Phys 1998;108: 6033–6041.
28. Berendsen HJC, Grigera JR, Straatsma TP. The missing term in effective pair potentials. J Phys Chem 1987;91:6269–6271.
29. Nar H, Messerschmidt A, Huber R, Van De Kamp M, Canters GW. Crystal structure analysis of oxidized pseudomonas aeruginosa azurin at ph 5.5 and pH 9.0. A pH-induced conformational transition involves a peptide bond flip. J Mol Biol 1991;221:765–772.
30. Bernstein F, Koetzle T, Williams G, et al. The protein data bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
31. Melchionna S, Ciccotti G. Atomic stress isobaric scaling for systems subjected to holonomic constraints. J Chem Phys 1997;106: 195–200.
32. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motions of a system with constraints: molecular dynamics of N-alkanes. J Comp Phys 1977;23:327–341.
33. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
34. Kneller GR. Superposition of molecular structures using quaternions. Mol Sim 1991;7:113–119.
35. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. J Mol Biol 1971;55:379–400.
36. Impey RW, Madden PA, McDonald IR. Hydration and mobility of ions in solution. J Phys Chem 1983:87:5071–5083.
37. Kuhn LA, Swanson CA, Pique ME, Tainer JA, Getzoff ED. Atomic and residue hydrophylicity in the context of folded protein structure. Proteins 1995;23:536–547.
38. Levitt M, Park BH. Water: now you see it, now you don't. Structure 1993;15:223–226.
39. Kuhn LA, Siani MA, Pique ME, Fisher CL, Getzoff ED, Tainer JA. The interdependence of protein surface topography and bound water molecules revealed by surface accessibility and fractal density measures. J Mol Biol 1992;228:13–22.
40. Wade RC, Mazor MH, McCammon JA, Quiocho FA. A molecular dynamics study of thermodynamic and structural aspects of the hydration of cavities in proteins. Biopolymers 1991;1:919–931.
41. Kraulis J. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. J Appl Cryst 1991;24:946–950.