Residence Times of Water Molecules in the Hydration Sites of Myoglobin

Vladimir A. Makarov,*† B. Kim Andrews,† Paul E. Smith,‡ and B. Montgomery Pettitt†

*Program in Structural and Computational Biology and Molecular Biophysics, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030; †Department of Chemistry, University of Houston, Houston, Texas 77204-5641; and ‡Department of Biochemistry, Kansas State University, Manhattan, Kansas 66506-3702 USA

ABSTRACT Hydration sites are high-density regions in the three-dimensional time-averaged solvent structure in molecular dynamics simulations and diffraction experiments. In a simulation of sperm whale myoglobin, we found 294 such high-density regions. Their positions appear to agree reasonably well with the distributions of waters of hydration found in 38 x-ray and 1 neutron high-resolution structures of this protein. The hydration sites are characterized by an average occupancy and a combination of residence time parameters designed to approximate a distribution of residence times. It appears that although the occupancy and residence times of the majority of sites are rather bulk-like, the residence time distribution is shifted toward the longer components, relative to bulk. The sites with particularly long residence times are located only in the cavities and clefts of the protein. This indicates that other factors, such as hydrogen bonds and hydrophobicity of underlying protein residues, play a lesser role in determining the residence times of the longest-lived sites.

INTRODUCTION

There exists ample experimental and theoretical evidence that a protein in an aqueous solution modifies both structure and dynamics of water around it (Teeter, 1991; Brunne et al., 1993; Lounnas and Pettitt, 1994a,b; Schoenborn et al., 1995; Phillips and Pettitt, 1995; Denisov and Halle, 1996; Burling et al., 1996). Previously, we explored the relationship between the solvent structure and dynamics at the protein-water interface in terms of diffusion rates and radial distribution functions (Makarov et al., 1998b).

Here we intend to study the protein-water interface from a different point of view, concentrating on those water molecules that are closely associated with the protein so that their motion can no longer be described simply as diffusive. The concept of hydration sites (Lounnas and Pettitt, 1994a,b; Hummer et al., 1995), as opposed to crystal waters of hydration, appears to be useful in this respect.

Hydration sites appear as local maxima in the time-averaged singlet solvent density (Lounnas and Pettitt, 1994a,b; Scanlon and Eisenberg, 1975, 1981). This definition immediately invites a comparison to the solvent structure observed in x-ray (Scanlon and Eisenberg, 1975, 1981; Lounnas and Pettitt, 1994a,b; Hummer et al., 1995) or neutron diffraction (Gu and Schoenborn, 1995) experiments. Although this comparison is useful for verification of the overall structure of the hydration shell observed in the simulation, a direct comparison can be problematic. There are two reasons for this. First, the simulated hydration sites, unlike experimental hydration sites, are not subject to normal refinement constraints, as they are simply points in space and not individual molecules. The conditions used to define the simulated sites may be chosen so as to best reproduce the complicated three-dimensional solvent distribution by a set of local maxima, rather than to comply with steric restrictions and hydrogen bonding requirements of individual molecules. Second, there are no crystal packing effects in the simulation, because we simulate a solution, albeit a very concentrated one. Therefore, we intend to use the hydration site concept not just to provide direct comparisons to crystal structures, but also as a tool for the analysis of solvent dynamics at the interface with the protein. Temporal ordering of solvent in the hydration shell of the protein may be characterized by the population analysis of hydration sites (Lounnas and Pettitt, 1994b; Gu and Schoenborn, 1995), from which the average water residence times may be derived (Brunne et al., 1993; Lounnas and Pettitt, 1994b; Schoenborn et al., 1995; Abseher et al., 1996; Rocchi et al., 1997). Hydration sites are ideal as volume elements for the molecular residence time calculation. Using the hydration sites, as opposed to shells (Abseher et al., 1996) or irregular volume elements proximal to specific solute residues (Brunne et al., 1993; Garcia and Stiller, 1993; Kovacs et al., 1997; Rocchi et al., 1997), allows us to relate the temporal order parameters (residence times) to the spatial structure of solvent in the most natural way.

In turn, this comparison enables us to compare our analysis to that of x-ray crystallography or nuclear magnetic resonance (NMR), even though the precise definitions of our densities and residence time parameters may differ in detail from those used in these experimental fields. Last, but not least, hydration sites offer a significant technical advantage over other definitions of volumes for the residence time calculation. All sites can be forced to be of the same shape and size, which makes all occupancy and temporal parameters consistent and enables a direct comparison of different sites.

Specifically, we seek to understand the correlation between the spatial and temporal structure of the protein hydration shells. We wish to quantify the extent to which it

Received for publication 25 May 1999 and in final form 11 July 2000.

Address reprint requests to B. M. Pettitt, Department of Chemistry, University of Houston, Houston, Texas 77204-5641. E-mail: pettitt@uh.edu.

0006-3495/00/12/2966/09 $2.00

is possible to characterize the mobility of the interfacial water in terms of specific residence times and occupancies, as is commonly done both in simulation (Lounnas and Pettitt, 1994a,b; Gu and Schoenborn, 1995; Brunne et al., 1993; Schoenborn et al., 1995; Abseher et al., 1996; Rocchi et al., 1997) and experimental studies (Brunne et al., 1993; Schoenborn et al., 1995). In this respect, we wish to determine the factors that affect the lifetimes of water molecules on the protein surface.

METHODS

Molecular dynamics (MD) simulation

The analysis presented in this paper is based upon a MD simulation of sperm whale myoglobin that has been reported in detail in a separate publication (Andrews et al., 1998). Therefore, only the brief account of the simulation setup will be presented here. The all-atom CHARMM-23 parameter set (MacKerell et al., 1992) was used to model a single molecule of carbomonoxy myoglobin (Protein Data Bank entry 2mgk, Bernstein et al., 1977; crystal structure by Quillin et al., 1993) solvated by 3717 flexible TIP3P (Jorgensen et al., 1983) water molecules in a 60.4 Å × 54.7 Å × 40.7 Å box under periodic boundary conditions. The system was prepared through a series of energy minimization and heating steps. The heating stage was followed by equilibration at 294 K for 200 ps. Equations of motion were integrated using a 0.5-fs time step without constraints. Electrostatic interactions were treated with a 13 Å cutoff and a potential-based switching function beginning at 10 Å. The trajectory was continued for a total of 1.1 ns, of which the last 900 ps were chosen for analysis. The trajectory was recorded with 0.1-ps intervals.

Calculation of solvent density distribution

This part of our analysis followed procedures previously developed in this lab (Lounnas and Pettitt, 1994a,b; Lounnas et al., 1994; Rudnicki and Pettitt, 1997; Makarov et al., 1998a). For each step of the MD trajectory, the protein was fitted to a consistent frame of reference. Next, the same transformation was applied to the water molecule coordinates, taking the periodic boundaries into account. The coordinates of the water oxygen atoms were then mapped onto the three-dimensional rectangular grid with a 0.5 Å grid step, producing an average three-dimensional number density distribution. The particular choice of the grid step is a compromise between the uncertainty in location of the density features and the statistical error in the local density value that arises due to a lower number of counts in each grid cell. At the chosen grid step every cell in the regions corresponding to bulk solvent would have at least 50 counts over the entire trajectory. The density map was smoothed by averaging the value of each cell with six of its nearest neighbors before further manipulations.

Localization of hydration sites

For the purposes of this work, protein hydration sites are defined as local maxima in the water oxygen number density map that satisfy certain conditions. They should be no farther than 5 Å from any protein atom, no closer than at least 1 Å (two grid steps) from each other, and have a maximum density value no lower than 150% of the bulk water density in the simulation. For the normal water density of 1 g/mL, the last cutoff is equivalent to 0.05 particles/Å^3. The maxima were determined by comparison of the density in the current grid cell to each of its six nearest neighbors (the 7-point rule). If close peaks were found, their coordinates were weighted by their respective maximum density values and averaged together, resulting in a single peak.

Comparison with crystal water positions

To compare the hydration peaks from the above simulation with the positions of the crystal waters of hydration, we have selected a number of sperm whale myoglobin structures, all at the resolution of 2 Å or better, from the Protein Data Bank (Bernstein et al., 1977). There were 73 sperm whale myoglobin structures in the database as of March 1998. Every protein was oriented in space to fit the reference structure, and the positions of any hydration water molecules were then saved and compared with those from all other structures on the list. Sometimes this comparison would indicate a trivial dependency between a group of structures. In such cases, only one structure from the group was taken into consideration. The purpose of this elimination procedure was to avoid potential bias due to identical water positions that might have been copied from a different myoglobin structure in a molecular replacement during crystallographic refinement. This procedure resulted in the following selection of structure files: 1abs, 1bvc, 1bvd, 1fcs, 1mbc, 1mbd, 1mbi, 1mcy, 1mgn, 1mlf, 1moa, 1mti, 1mym, 1spe, 1swm, 1vxa, 1vxb, 1vxc, 1vxf, 1yog, 1yoi, 2cmm, 2mb5, 2mbw, 2mgb, 2mgc, 2mgd, 2mge, 2mgf, 2mgg, 2mgh, 2mgi, 2mgj, 2mgk, 2mgl, 2mgm, 2mya, 2spl, and 4mbn. We then compared the simulated hydration sites to a collection of crystallographic water molecules contained in the above list. Quantitatively, the agreement between the theoretical and experimental water positions is measured by the proximal radial distribution of the crystal waters around the simulated hydration sites (Hummer et al., 1995).

Calculation of water residence times

We note that the definition of a "residence time" varies widely in the available simulation and NMR literature. Some of these definitions are discussed in a review by Schoenborn et al. (1995). The method used here is that of the time correlation function (Brunne et al., 1993), although the correlation function itself is modified and we offer a different interpretation. The correlation function in question is simply

$$P(\tau) = \int_0^\infty \delta(N(t) - N(t+\tau))dt \qquad (1)$$

where the function $\delta(x - y)$ takes the value of 1, when $x = y$ ($x \neq 0$ and $y \neq 0$), or 0, when either $x \neq y$ or both x and $y = 0$, meaning that the site is not occupied. $N(t)$ is the index of the water molecule found in the hydration site at time t. The site was considered occupied if a water molecule was found in a spherical volume with radius of 1.5 Å (roughly equal to the van der Waals radius of water) centered on the coordinates of this site. The resulting time correlation function is usually fitted by a single exponential (Brunne et al., 1993), but we have noticed that the single exponential fit is often poor. Using a double exponential model resulted in a more accurate fit in all cases. The two new rate constants in case of solvent diffusion correspond to solvent molecules that stay in the hydration shell for prolonged periods of time or enter and then immediately leave. Both of these types of water behavior are observed in the simulation, and it is important to distinguish between them. In fact, it has been suggested (Garcia and Stiller, 1993; Rocchi et al., 1997) that in some cases, groups of water molecules may show different residence times with respect to the same atomic site. The process of solvent diffusion is clearly much more complex than that described by either single or double exponential model and probably should be characterized by a distribution of rate constants. Although a Kohlrausch-Williams-Watts stretched exponential is often used to describe phenomena governed by multiple relaxation rates (Abseher et al., 1996), its interpretation is not so straightforward as that of a simple biexponential model. The model used here is

$$P(\tau) = W_0[(1 - w)\exp(-k_1\tau) + w\exp(-k_2\tau)] \qquad (2)$$

where W_0 is the average site occupancy, defined as the fraction of time the site was occupied by any water molecule, and k_1, $(1 - w)$ and k_2, w are the rate constants and the weights for the fast and the slow components, respectively. The short and the long residence times are then defined simply as $\tau_1 = 1/k_1$ and $\tau_2 = 1/k_2$.

For comparison, we have also repeated this calculation for sites in the bulk region of solvent in the simulation. There are 8 such sites located in the corners of the rectangular simulation box. The occupancy and residence time parameters for these bulk sites were then averaged together.

RESULTS AND DISCUSSION

Hydration site location and comparison with crystal data

Using the procedures described above, we have identified a total of 294 hydration sites around myoglobin. The number of hydration sites and their position in relation to the protein appear to be rather insensitive to the parameters of the localization procedure. For instance, in a lower resolution water oxygen number density map (at 1 Å grid step) the algorithm yields 288 sites at approximately same positions (there is, of course, a small error associated with the larger grid step). Thus, we conclude that our solvation sites are sufficiently reliable. For convenience, the sites are numbered in descending order depending on their maximum density value.

The spatial distribution of hydration sites is both highly anisotropic and asymmetric. In general, the hydration shell follows the shape of the protein, but there are regions where the hydration sites are tightly clustered and regions that are almost devoid of sites. In Fig. 1, the high density region is located on top of the protein, and the low density areas without hydration sites are to the right. The protein surface residues in the areas without high density sites (highlighted in yellow in Fig. 1) do not fall into any one particular residue class. Both charged and polar residues are found in that region along with the hydrophobic residues (Table 1). In contrast, the distribution of crystal hydration waters (Fig. 2) is much more uniform and contains few excessively populated or empty regions. To provide a quantitative measure for the comparison of crystal water positions to the MD hydration sites, we have constructed and displayed in Fig. 3 a cumulative distribution of nearest-neighbor distances between crystal waters and hydration sites (Hummer et al., 1995). The histogram achieves saturation after approximately 3 Å, meaning that even the worst positioned MD hydration sites would have at least one corresponding crystal water molecule at this distance. From this point of view, agreement of our data with the experiment is better than that found in the simulation used by Lounnas and Pettitt (1994a). On the other hand, even better reproduction of the experimental data is achieved in simulations of small hydrated molecules, such as nucleic acid fragments (Hummer et al., 1995; Feig and Pettitt, manuscript in preparation). The most probable reason for the discrepancy between the simulated and crystal hydration sites is the absence of the crystal packing effects in the simulation of the protein in solution.

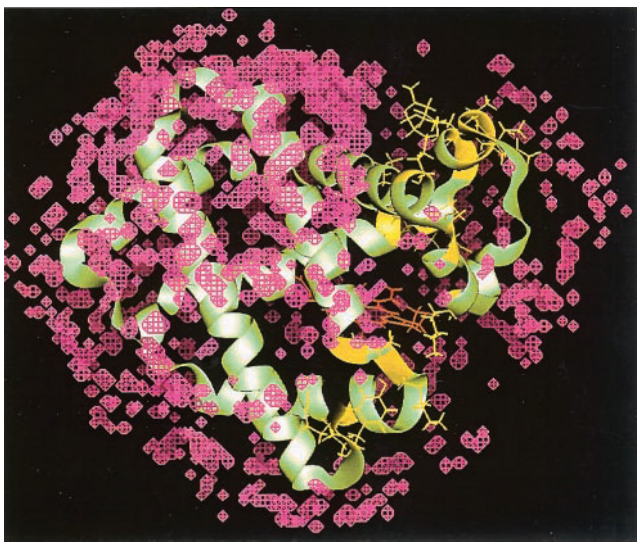

FIGURE 1 Three-dimensional distribution of high-density solvent peaks around myoglobin. Exposed protein residues that do not have solvent peaks associated with them are colored yellow. These residues form the bald spots on the surface of the protein. The heme is in red.

The simulated hydration sites form an elaborate hydrogen bonding network that encircles the protein and involves participation of many of its surface-accessible residues. A small fragment of this network is shown in Fig. 4. This polygonal structure is located in the high density region and is actually a part of an even more complex network of sites. Similar arrangements have been observed in the earlier study of myoglobin hydration (Lounnas and Pettitt, 1994a,b) and in high-resolution crystal structures (Teeter, 1991). Fig. 5 shows the site-site radial distribution function (RDF). It has a broad peak centered at 2.6 Å (the histogram step is 0.2 Å), which corresponds to water-to-water hydrogen bonding distance. The peaks at 1 Å and 1.4 Å are induced by the grid. In the previous study of myoglobin hydration sites (Lounnas and Pettitt, 1994a), the site-site RDF had a peak at 2.1 Å. Although the hydration sites do not represent individual water molecules and thus are not subject to usual steric constraints, our present site-site RDF is more consistent with chemical intuition and data from protein crystallography (Schoenborn et al., 1995).

TABLE 1 Amino acid composition of the bald spots in the hydration network of myoglobin

Location of "bald spot" (in Fig. 1)	Residues
Lower big spot	Lys-43, Gln-92, Lys-99, Pro-101, Tyr-152, Gln-153
Lower small spot	Ala-85, Glu-86
Upper spot	Pro-38, Ala-54, Glu-55, Lys-57, Asp-61

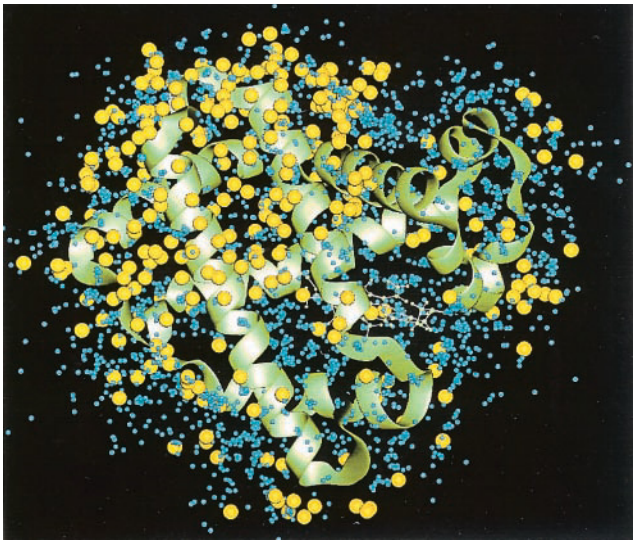

FIGURE 2 A view of the myoglobin hydration sites generated in the MD simulation (*yellow dots*) compared to the collection of the hydration sites contained in the Protein Data Bank (*blue dots*).

Temporal ordering of water around the protein

Population analysis of the hydration sites shows that all of them are never occupied simultaneously. In fact, the number of sites occupied at the same time never exceeds 164 out of 294, and there are no cases of double occupancy. Fig. 6 shows the population time correlation functions for two distinct hydration sites (numbers 20 and 36) along with that for the bulk solvent. We begin with the description of these bulk sites to establish a baseline for future comparisons. Average occupancy (W_0) in the bulk sites, which are positioned in the corners of the simulation box as described above, is 49%, and the residence time parameters are t_1 = 0.34 ps, t_2 = 4.1 ps, and w = 0.39. We note that both bulk residence times are short and the weight of the slow component (w) is low.

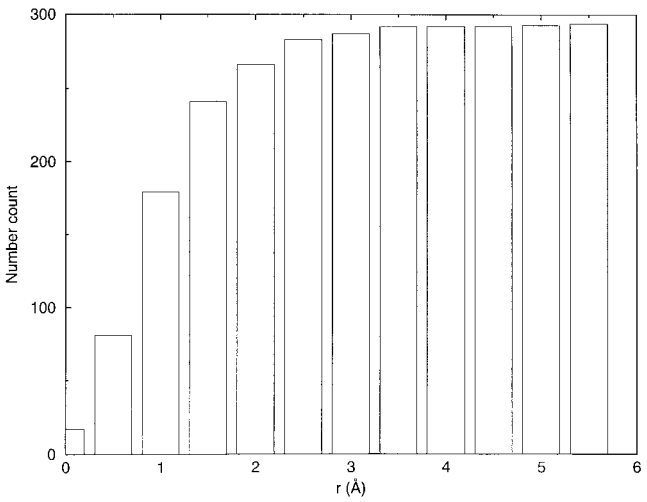

FIGURE 3 Cumulative histogram of nearest neighbor distances between the MD and the Protein Data Bank water sites.

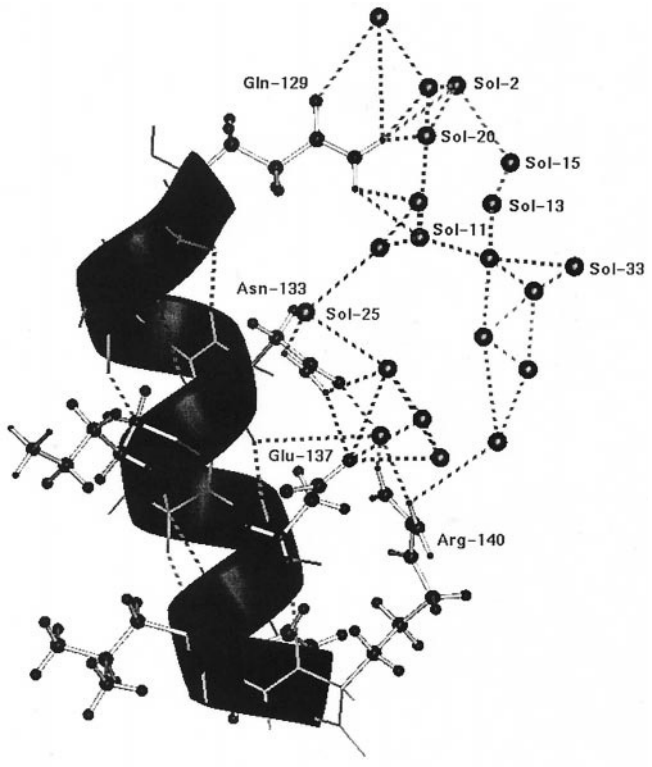

FIGURE 4 A fragment of the hydrogen-bonding network of hydration sites in the vicinity of site number 2 and Gln-129. Only the participating protein residues and the highest density solvent sites are labeled. The numbers of lower-density sites are omitted for clarity. Hydration sites were treated as atoms that may participate in a hydrogen bond as both a donor and an acceptor at the same time. The definition of a hydrogen bond followed a general convention that requires a distance between donor and acceptor no more than 3.3 Å.

Distributions of the site occupancy, t_1, t_2, and w for all 294 sites are shown in Fig. 7. The distributions of the residence times and occupancy peak at their respective bulk values, which indicates that many of the hydration sites are rather bulk-like. Typical residence times (t_2) vary between 5 and 35 ps, and the average site occupancy varies from 15% (site 113) to 59% (site 2). Long tails on the right hand side of the residence time distributions indicate that the protein may prolong the life of the bound water molecules by up to one order of magnitude, but such sites are few. A residence time on the order of hundreds of picoseconds in a 1-ns simulation is likely to be caused by a single visit of a water molecule to the particular site, occurring only once over the course of the simulation, and this may make the calculated residence time statistically unreliable. Nonetheless, examples of these long-lived sites are listed in Table 2. The influence of the protein on the surrounding solvent is manifested to an even greater degree in the distribution of the

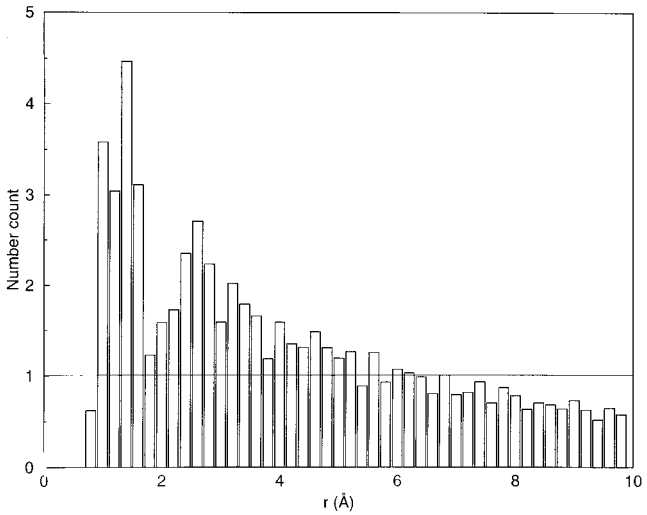

FIGURE 5 Site-site radial distribution function.

slow component weight parameter, w (Fig. 7 *d*). It is shifted to the right of its respective bulk value in agreement with the fast component dominating the motion in the bulk. In terms of water mobility this means that there are two alternative scenarios possible for every water molecule that enters a site: one may stay there for a time period of t_2, on the average, or enter and immediately exit with a characteristic short residence time t_1. The w parameter may be thought of as simply a fraction of water molecules that follow one of these two alternative scenarios. Thus, the distribution of residence times at the interface as opposed to bulk appears to be skewed toward the longer components.

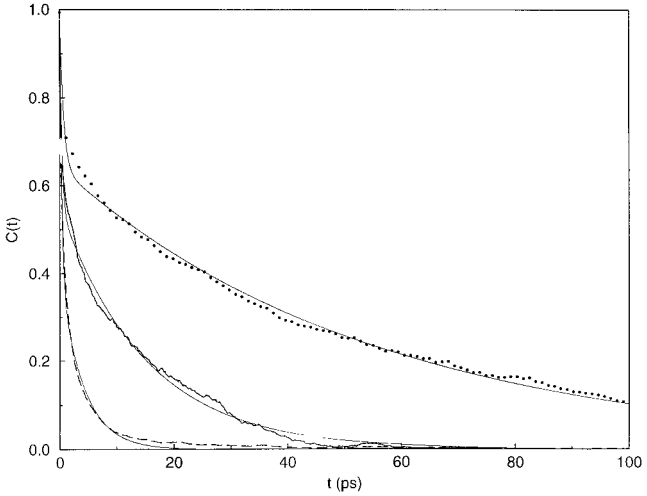

FIGURE 6 Examples of the population time correlation functions for different hydration sites: *solid line*, site number 20; *dotted line*, site 36; *dashed line*, bulk water. *Thin solid lines* represent the double exponential fits (Eq. 2) to the correlation functions. The corresponding fit parameters are $t_1(20) = 0.35$ ps, $t_2(20) = 15.0$ ps, $w(20) = 0.55$ and $t_1(36) = 0.6$ ps, $t_2(36) = 54.6$ ps, $w(36) = 0.64$.

Our present observations with regard to the water residence times agree very well with available experimental and simulation data. Short residence times for water molecules in protein hydration shells were observed in many other simulations, without respect to the nature of the protein studied or to the exact method for the calculation of the residence times (Abseher et al., 1996; Brunne et al., 1993; Garcia and Stiller, 1993; Kovacs et al., 1997; Lounnas and Pettitt, 1994b; Muegge and Knapp, 1995; Rocchi et al., 1997). Water residence times on a 10- to 50-ps time scale were also observed in protein NMR experiments (Brunne et al., 1993; Denisov and Halle, 1996). Long-tail residence time distributions with a peak below 10 ps were found in the analysis of a MD simulation of hydrated ubiquitin (Abseher et al., 1996). These distributions are of essentially the same shape as those computed here. It has been suggested (Abseher et al., 1996) that the existence of multiple relaxation rates and residence time distributions in the protein hydration shell is due to the structural and chemical heterogeneity of the protein surface.

In this regard, it is important to note that there is no simple correlation between the maximum site density, occupancy, and residence times. Table 3 illustrates the absence of a relationship between the site maximum density and the temporal order parameters. Further, lack of correlation between density and residence time is evident from Table 2. The sites are numbered depending on their maximum density, starting with the highest density site (number 1), but the sites with longest residence times are 42, 1, 214, and 6.

This lack of correlation between the spatial (density) and temporal (residence times) order in the surrounding solvent is hardly surprising, given the observations in NMR experiments. A vast majority of water molecules that appear ordered in protein crystal structures cannot be resolved in NMR, partly because of the lack of crystal packing in solution, but also due to their rapid exchange with bulk solvent (Denisov and Halle, 1996). On the other hand, many protein crystal structures appear to have rather large empty internal cavities, where the presence of positionally disordered water is indicated only by NMR dispersion experiments (Denisov and Halle, 1996).

Yet another comparison is possible to the xenon binding sites observed experimentally in the structure of metmyoglobin (Tilton et al., 1984). Of the four xenon binding sites, we observe water in three, namely the two xenon sites in the proximal heme pocket (overlapping hydration sites 113, 115, and 19) and the xenon site in the EF/H helix corner. The latter xenon binding site overlaps with many hydration sites, of which the site number 34 has the longest residence time. The last xenon binding site, on the distal side of the myoglobin molecule, apparently does not contain spatially or temporally ordered water. This may be correlated to its low occupancy by xenon in experiment, which is the lowest of all four xenon sites at 0.46 (Tilton et al., 1984).

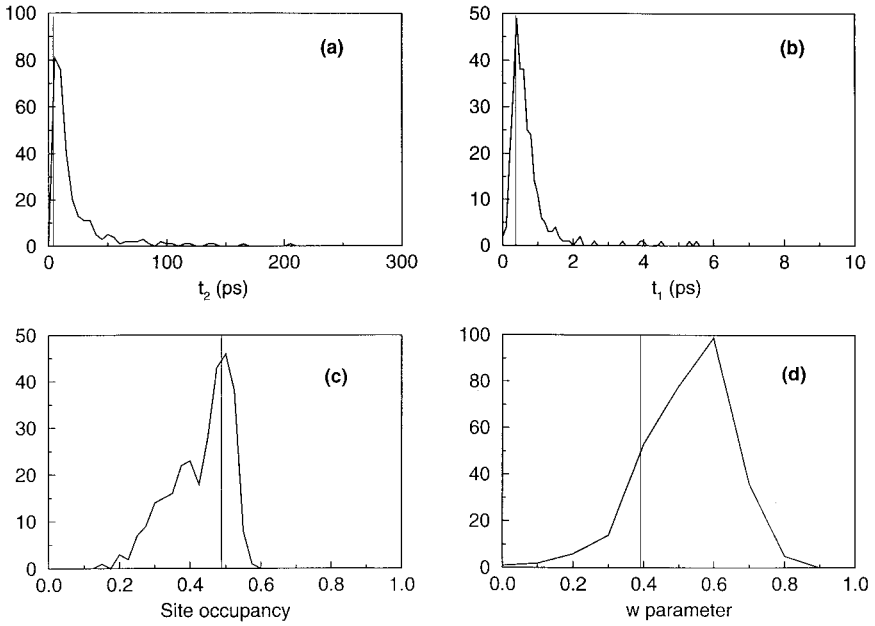

FIGURE 7 Distributions of the various parameters of the population time correlation functions fitted with the double exponential. (*a*) Long residence time t_2. (*b*) Short residence time t_1. (*c*) Site occupancy. (*d*) *w* parameter. In each plot, the thin vertical line denotes the corresponding bulk value.

Because there is no obvious relation between the number density in the hydration shell and the water residence times, one may rightfully question what factors determine the residence times. At least three potential answers have been offered in the literature: (i) the chemical nature of the protein residues close to the hydration shells for which the residence times are measured (Schoenborn et al., 1995); (ii) number of hydrogen bonding opportunities available at a given site (Denisov and Halle, 1996); and (iii) the local geometry of the protein surface, i.e., whether the site is buried or exposed to bulk (Denisov and Halle, 1996).

The first hypothesis has been investigated in many protein simulation studies (Brunne et al., 1993; Garcia and Stiller, 1993; Muegge and Knapp, 1995; Kovacs et al., 1997; Rocchi et al., 1997) that attempted to produce ranking relations for the residence times of water molecules in hydration shells of charged, polar and non-polar protein residues. The results of these studies have been inconclusive or conflicting (Schoenborn et al., 1995; Denisov and Halle, 1996). Our own attempt in this direction was also unsuccessful.

The second hypothesis has been dismissed by Denisov and Halle (1996) upon examination of the hydration waters in various protein crystal structures. While ordered water molecules make, on average, 2 to 4 hydrogen bonds to the protein, this number is comparable to the average number of hydrogen bonds per water molecule in the bulk (3 to 4), and nearly all of this ordered water is short-lived (Denisov and Halle, 1996).

The third hypothesis may be tested immediately by mapping the sites with different residence times onto the protein

TABLE 2 Hydration sites with unusually long residence times (t_2)

Site number	Residence time (ps)
42	456.5
1	454.0
214	205.2
6	168.3
115	141.0
32	139.0
4	120.0
139	115.2
72	106.8
113	101.5
34	98.2
71	97.8
19	89.2
266	84.5
87	83.0
30	81.4

TABLE 3 Density, occupancy and residence time parameters of the first 10 hydration sites

Site number	Maximum density	Average occupancy	t_1 (ps)	t_2 (ps)	*w*
1	2.61	0.50	1.32	453.92	0.69
2	2.52	0.59	0.34	15.17	0.64
3	2.34	0.41	0.30	29.09	0.72
4	2.34	0.54	0.89	120.02	0.64
5	2.31	0.49	0.90	17.00	0.56
6	2.28	0.30	0.55	168.26	0.73
7	2.25	0.56	1.56	43.30	0.30
8	2.22	0.24	4.57	56.61	0.57
9	2.16	0.43	0.35	13.08	0.50
10	2.16	0.31	1.03	39.33	0.38

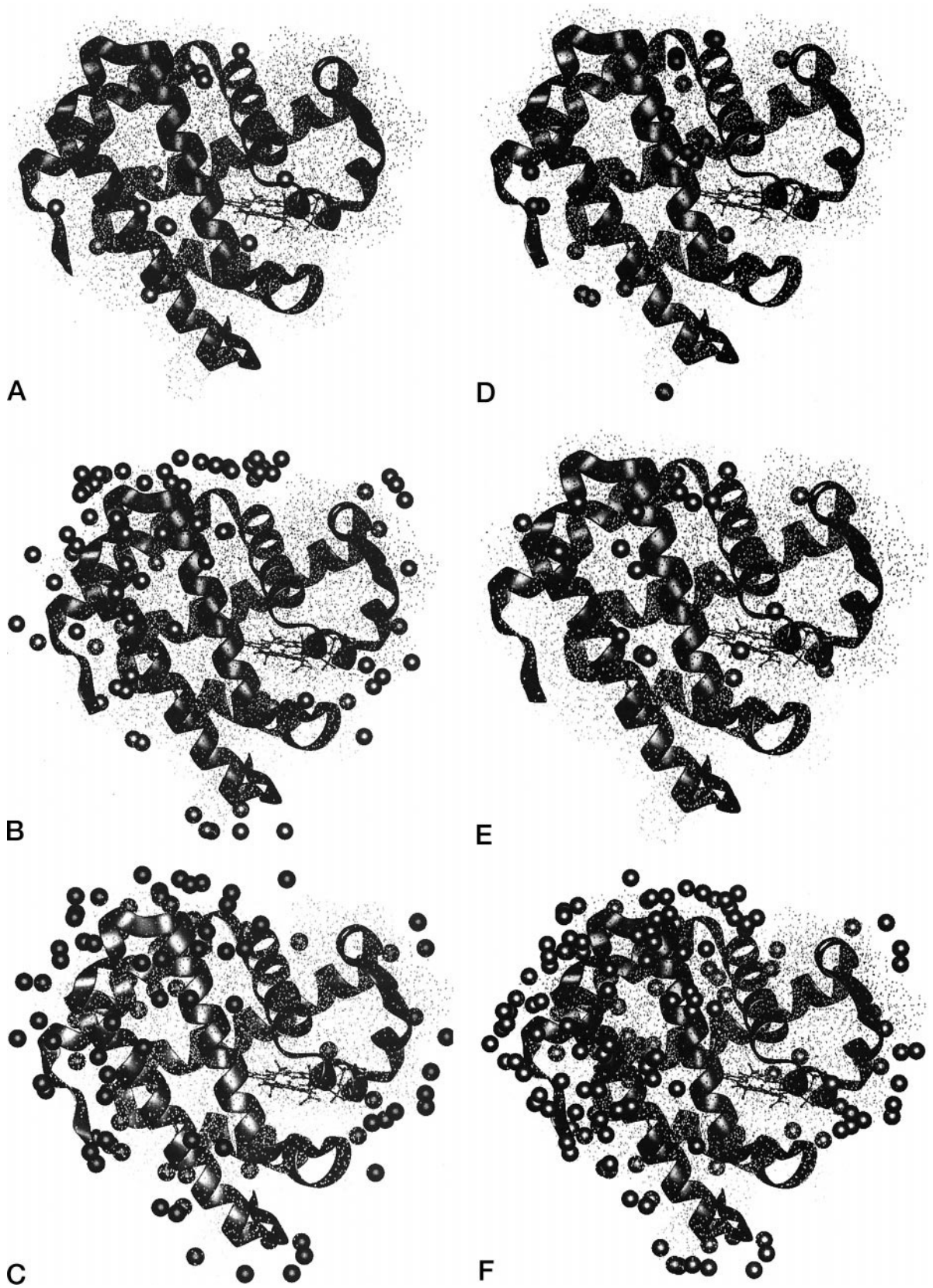

FIGURE 8 Spatial distributions of myoglobin hydration sites. (*a*) Sites with high residence time t_2. (*b*) Sites with low residence time t_2. (*c*) Sites with high occupancy. (*d*) Sites with low occupancy. (*e*) Sites with high *w* parameter (*f*) Sites with low *w* parameter. The gray dotted area represents the molecular surface of the protein.

structure. Without exception, all sites with very long residence times (defined here as $t_2 \geq 80$ ps) are trapped either in the cavities inside the protein or in the grooves and concave regions, whereas the low residence time sites ($t_2 \leq 10$ ps) are in convex areas and appear to be only loosely associated with the protein (Fig. 8). The opposite is true for the occupancy: high occupancy sites ($W_0 \geq 0.5$) are exposed, and low occupancy sites ($W_0 \leq 0.3$) are more buried. Interestingly, all sites with a high slow exponent weight ($w \geq 0.6$) are also buried, and all sites with a low w ($w \leq 0.4$) are exposed. Although the three classes of buried sites (high t_2, low W_0, and high w) do overlap, they are not identical, with only 3 sites in common.

Therefore, following numerous experimental observations (reviewed by Schoenborn et al., 1995; Denisov and Halle, 1996), we find that the water residence times are determined almost exclusively by the position of the water molecules on the protein surface: cleft and buried waters have significantly longer lifetimes than those that are exposed. Although other factors mentioned above may also affect the residence times, they appear to be secondary to the geometry with respect to local topography. We hypothesize that these factors do influence the residence time of the interfacial water, but on different time scale. For instance, an ability to form hydrogen bonds to the protein or to the neighboring sites may determine the length of the water residence in a surface site with $t_2 \ll 50$ ps, but it will not be as important in a cleft site with t_2 on the order of 500 ps. This would explain the failure of the ranking relation analysis mentioned above. Additional analysis, preferably using even longer and more accurate computer simulations, should shed more light on this issue.

CONCLUSIONS

In conclusion, we confirm that there is no direct correlation between the spatial (number density) and temporal (residence time) order of solvent in the protein hydration shell. We find that there is no single specific residence time for any particular location in the protein-solvent interface. Instead, there exists a distribution of residence times, that may have both extremely short and extremely long components. In our study, the long and the short residence times t_2 and t_1 play this role. Close to the protein surface the residence time distribution is biased toward the longer components. The residence times of the interfacial water molecules appear to depend most strongly upon the degree of exposure of the hydration site to bulk solvent. Buried and concave sites have much longer residence times than those that are convex and exposed to bulk. In addition, hydrogen bonding opportunities and the hydrophobic character of the underlying protein residues may influence the water residence times, but these factors only appear to be important for the exposed sites with inherently short residence times.

Supplementary material

Hydration site coordinates and their occupancy and residence time parameters are available in electronic form on request from the authors.

Financial support from The Robert A. Welch Foundation, the National Institutes of Health, and the National Science Foundation is gratefully acknowledged. B. K. A. was supported by the Keck Center for Computational Biology when this project was started. The myoglobin simulation was done with a grant of time from the Center for Research in Parallel Computing at the California Institute of Technology. P. E. S. is partially supported by the Kansas Agricultural Experimental Station.

REFERENCES

Abseher, R., H. Schreiber, and O. Steinhauser. 1996. The influence of a protein on water dynamics in its vicinity investigated by molecular dynamics simulation. *Proteins Struct. Funct. Genet.* 25:366–378.

Andrews, B. K., T. Romo, J. B. Clarage, B. M. Pettitt, and G. N. Phillips. 1998. Characterizing global substates of myoglobin. *Structure.* (in press)

Bernstein, F. C., T. F. Koetzle, G. J. B. Williams, E. F. J. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. 1977. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112:535–542.

Bizzarri, A. R., C. X. Wang, W. Z. Chen, and S. Cannistraro. 1995. Hydrogen bond analysis by MD simulation of copper plastocyanin at different hydration levels. *Chem. Phys.* 201:463–472.

Brunne, R. M., E. Liepinsh, G. Otting, K. Wüthrich, and W. F. van Gunsteren. 1993. 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.* 231: 1040–1048.

Burling, F. T., W. I. Weis, K. M. Flaherty, and A. T. Brünger. 1996. Direct observation of protein solvation and discrete disorder with experimental crystallographic phases. *Science.* 271:72–77.

Denisov, V. P., and B. Halle. 1996. Protein hydration dynamics in aqueous solution. *Faraday Discuss.* 103:227–244.

Garcia, A. E., and L. Stiller. 1993. Computation of the mean residence time of water in the hydration shells of biomolecules. *J. Comp. Chem.* 14:1396–1406.

Gu, W., and B. P. Schoenborn. 1995. Molecular dynamics simulation of hydration in myoglobin. *Proteins.* 22:20–26.

Hummer, G., A. E. Garcia, and D. M. Soumpasis. 1995. Hydration of nucleic acid fragments: comparison of theory and experiment for high-resolution crystal structures of RNA, DNA, and DNA-drug complexes. *Biophys. J.* 68:1639–1652.

Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

Kovacs, H., A. E. Mark, and W. F. van Gunsteren. 1997. Solvent structure at a hydrophobic protein surface. *Proteins.* 27:395–404.

Lounnas, V., and B. M. Pettitt. 1994a. A connected-cluster of hydration around myoglobin: correlation between molecular dynamics simulations and experiment. *Proteins.* 18:133–147.

Lounnas, V., and B. M. Pettitt. 1994b. Distribution function implied dynamics versus residence times and correlations: solvation shells of myoglobin. *Proteins.* 18:148–160.

MacKerell, A. D., D. Bashford, M. Bellott, R. L. Dunbrack, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B.

Prodhom, B. Roux, M. Schlenkrich, J. C. Smith., R. Stote, J. Straub, J. Wiorkiewicz-Kuczera, and M. Karplus. 1992. Self-consistent parameterization of biomolecules for molecular modeling and condensed phase simulations. *FASEB J.* 6A:143

Makarov, V. A., B. K. Andrews, and B. M. Pettitt. 1998a. Reconstructing the protein-water interface. *Biopolymers.* 10:469–478.

Makarov, V. A., M. Feig, B. K. Andrews, and B. M. Pettitt. 1998b. Diffusion of solvent around biomolecular solutes: a molecular dynamics simulation study. *Biophys. J.* 75:150–158.

Muegge, I., and E. W. Knapp. 1995. Residence times and lateral diffusion of water at protein surfaces: application to BPTI. *J. Phys. Chem.* 99: 1371–1374.

Phillips, G. N., and B. M. Pettitt. 1995. Structure and dynamics of the water around myoglobin. *Protein Sci.* 4:149–158.

Quillin, M. L., R. M. Arduini, J. S. Olson, and G. N. Phillips. 1993. High-resolution crystal structures of distal histidine mutants of sperm whale myoglobin. *J. Mol. Biol.* 234:140–155.

Rocchi, C., A. R. Bizzarri, and S. Cannistraro. 1997. Water residence times around copper plastocyanin: a molecular dynamics simulation approach. *Chem. Phys.* 214:261–276.

Rudnicki, W. R., and B. M. Pettitt. 1997. Modeling the DNA-solvent interface. *Biopolymers.* 41:107–119.

Scanlon, W. J., and D. Eisenberg. 1975. Solvation of crystalline proteins: theory and its application to available data. *J. Mol. Biol.* 98:485–502.

Scanlon, W. J., and D. Eisenberg. 1981. Solvation of crystalline proteins. Solvent bound in sperm whale metmyoglobin type A crystals at 6.1 and 23.5°C. *J. Phys. Chem.* 85:3251–3256.

Schoenborn, B. P., A. Garcia, and R. Knott. 1995. Hydration in protein crystallography. *Prog. Biophys. Mol. Biol.* 64:105–119.

Tilton, R. F., I. D. Kuntz, and G. A. Petsko. 1984. Cavities in proteins: structure of a metmyoglobin–xenon complex solved to 1.9 Å. *Biochemistry*. 23:2849–2857.

Teeter, M. M. 1991. Water-protein interactions: theory and experiment. *Ann. Rev. Biophys. Biophys. Chem.* 20:577–600.