PROTEINS: Structure, Function, and Genetics 18:148–160 (1994)

Distribution Function Implied Dynamics Versus Residence Times and Correlations: Solvation Shells of Myoglobin

Valère Lounnas and B. Montgomery Pettitt
Department of Chemistry, University of Houston, Houston, Texas 77204-5641

ABSTRACT **The dynamics of water at the protein–solvent interface is investigated through the analysis of a molecular dynamics simulation of metmyoglobin in explicit aqueous environment. Distribution implied dynamics, harmonic and quasiharmonic, are compared with the simulated macroscopic dynamics. The distinction between distinguishable solvent molecules and hydration sites developed in the previous paper is used. The simulated hydration region within 7 Å from the protein surface is analyzed using a set of 551 hydration sites characterized by occupancy weights and temperature *B*-factors determined from the simulation trajectory. The precision of the isotropic harmonic and anisotropic harmonic models for the description of proximal solvent fluctuations is examined. Residence times and dipole reorientation times of water around the protein surface are compared with NMR and ESR results. A correlation between diffraction experiment quantities such as the occupancy weights and temperature factors and the residence and correlation times resulting from magnetic resonance experiments is found via comparison with simulation.** © 1994 Wiley-Liss, Inc.

Key words: myoglobin, solvation, dynamics

INTRODUCTION

Many studies have considered the relation between the Debye–Waller factors refined from protein crystallographic data and the implied protein dynamics.[1] Static disorder plays a role in the refinement of *B*-factors and is not directly part of the dynamics though it may be an equilibrium consequence of dynamics during or after crystallization.[2,3] To help identify and control for such effects the configurations of a molecular dynamics computer simulation were used to generate an average protein density, subject it to refinement, and compare it with the dynamically averaged coordinates from the simulation.

In the previous paper an analysis was performed on the surrounding solvent distribution in a spirit similar to the previous protein analyses.[1,4] However, due to the mobile, fluid nature of solvent compared to native globular protein, different techniques were necessary in the analysis and subsequent mock refinement of the solvent in the first few solvation layers. The main distinction made use of the difference between individual solvent molecules and hydration sites or local maxima in the average solvent density near the protein.

Here we use that distinction to extend the previous analysis to dynamic properties and consequences. We examine the refined distribution implied dynamics in terms of the harmonic model (plus higher order corrections) and the quasiharmonic model[2] of hydration sites versus the underlying microscopic dynamics of the distinguishable waters of the simulation. Residence times and orientational relaxation times of the water molecules associated with the hydration sites are compared with relevant experimental dynamic data in an effort to reveal the mechanism underlying the correlations with *B*-factors and occupancy weights from crystallographic studies.

METHOD

The present study is the continuation of a previous analysis involving a 170-psec molecular dynamics trajectory of metmyoglobin in explicit aqueous surrounding. Specifically, the preceeding paper contains the method and details concerning the computation of the three-dimensional solvent density distribution maps and the cartesian locations of the 551 hydration sites that are used in the following analysis of the dynamics of the myoglobin hydration network. Choosing a convention, hydration sites are named from W154 to W705, the numbers 1 to 153 being used for the protein amino acid residues.

Isotropic Harmonic Model

The shape of a hydration site is related to the local dynamics of the protein polar or charged groups to which the water molecules are attached by H-bonds.

Received June 3, 1993; revision accepted October 8, 1993.
Address reprint requests to Dr. B. Montgomery Pettitt, Department of Chemistry, University of Houston, Houston, Texas 77204-5641.

Although intrinsically anisotropic and somewhat anharmonic, the local protein motions are usually well fitted by harmonic or quasiharmonic oscillations of atoms around their mean positions.[1–4] For this reason, anisotropic Debye–Waller tensors can be used in crystallographic refinement to account for the protein as well as for the solvent motions.[5,6] In this case, a set of six parameters $B_j^{xx} = < \Delta x_j^2 >$, $B_j^{yy} = < \Delta y_j^2 >$, $B_j^{zz} = < \Delta z_j^2 >$, $B_j^{xy} = < \Delta x_j \Delta y_j >$, $B_j^{yz} = < \Delta y_j \Delta z_j >$, and $B_j^{zx} = < \Delta x_j \Delta z_j >$ is required to characterize the second moments of harmonic projections of the fluctuation of each atom j around its average position. In the simplest case where the atomic motions are assumed to be perfectly isotropic then

$$B_j^{xx} = B_j^{yy} = B_j^{zz} \tag{1}$$

and

$$B_j^{xy} = B_j^{yz} = B_j^{xz} = 0. \tag{2}$$

Another common assumption of X-ray refinement procedures concerns the nature of the atom motions around their average locations. These motions are assumed to occur in harmonic potential wells which restrains their effective dynamics to a reduced number of variables. In the case of isotropic motions, the variance of the positional fluctuations

$$\sigma_j^2 = (B_j^{xx} + B_j^{yy} + B_j^{zz}) \tag{3}$$

represents the dynamics of the thermal motions.

This quantity is directly proportional to the usual temperature or B-factor of X-ray crystallography. B-factors or isotropic mean-square displacements are quantities easy to determine from an MD trajectory. We have computed the B-factor characterizing each hydration site from the positional fluctuations of the water oxygen atoms within a 1.5-Å sphere centered at the site locus $\mathbf{r}_s$.

$$B = \frac{8\pi^2}{3} < |\mathbf{r}_w(t) - \mathbf{r}_s|^2 >_{|\mathbf{r}_w - \mathbf{r}_s| < 1.5\text{Å}}. \tag{4}$$

The isotropic probability distribution function, $P_s(r)$, of the distance from the cartesian location of each hydration site is computed as

$$P_s(r) = \frac{1}{N_t \times \Delta\tau(r)} \sum_{j=1}^{N_w} \sum_{i=1}^{N_t} \delta[r_j(t_i) - r] \tag{5}$$

where $\mathbf{r}_j(t_i)$ stands for the coordinates of the water oxygen centers, with t_i and j indicating the i^{th} time point and water residue number, respectively. N_t *and* N_w represent, respectively, the total numbers of snapshots and water molecules. $\Delta\tau(r) = 4\pi r^2 \Delta r$ is the usual volume normalization factor. Notice also that

$$W = \sum_{r<1.50\text{Å}} P_s(r) \times \Delta\tau(r) \tag{6}$$

represents the site occupancy weight computed in a volume within 1.5 Å of a particular site. In the case of isotropic and harmonic motions, the $P_s(r)$ distribution is defined with a gaussian probability distribution,

$$P_g(r) = \frac{1}{W(2\pi\sigma^2)^{1/2}} exp\left(\frac{-r^2}{2\sigma^2}\right) \tag{7}$$

centered at a site location which is characterized by a variance σ^2 and an occupancy weight W.

Anisotropic Quasiharmonic Model

The deviation from isotropy in the thermal motions of the atoms around their average location is reflected by unequal variances,

$$B_j^{xx} \neq B_j^{yy} \neq B_j^{zz} \tag{8}$$

and often nonvanishing covariances,

$$B_j^{xy} \neq 0 \quad B_j^{yz} \neq 0 \quad B_j^{xz} \neq 0. \tag{9}$$

In this case, the anisotropy of the distribution can be parameterized according to

$$A_1 = \left[\frac{B_j^{xx}}{1/2(B_j^{yy} + B_j^{zz})}\right] - 1 \tag{10}$$

where $B_j^{xx} > B_j^{yy}$, and $B_j^{xx} > B_j^{zz}$. Other cases lead to a similar formula by permutation of X, Y, and Z.[7]

The analyses of the third and fourth moments of the positional distribution are indicative of the degree of anharmonicity of the atomic motions.[8] The asymmetry is measured by the coefficients of skewness identically defined in the X, Y, or Z directions by

$$\alpha_3^x = \frac{<(\Delta x)^3>}{<(\Delta x)^2>^{3/2}}. \tag{11}$$

The peakedness (flatness) is measured by the coefficients of excess kurtosis which is equivalently defined in the X, Y, or Z directions according to

$$\alpha_4^x = \frac{<(\Delta x)^4>}{<(\Delta x)^2>^2} - 3. \tag{12}$$

If the potential is harmonic in a given direction i = X, Y, or Z then $\alpha_3^i = \alpha_4^i = 0$.

When the probability density distribution of the atom centers has or is modeled as having an ellipsoidal character, the directions and amplitudes of the positional fluctuation in the principal axis system can be obtained by diagonalization of the symmetric matrix formed with the B-set of variances and covariances.

The actual thermal fluctuation can be fit by an ellipsoidal gaussian distribution P_e, which is actu-

ally a product of one-dimensional gaussians along the principal axes of the distribution. This is the quasiharmonic distribution

$$P_e = P_g(X_a) \times P_g(Y_b) \times P_g(Z_c) \quad (13)$$

or

$$P_e(r) = \frac{1}{(8\pi^3\sigma_a^2\sigma_b^2\sigma_c^2)^{1/2}} exp\left(-\frac{X_a^2}{2\sigma_a^2} - \frac{Y_b^2}{2\sigma_b^2} - \frac{Z_c^2}{2\sigma_c^2}\right) \quad (14)$$

where σ_a^2, σ_b^2, σ_c^2, denote the variances along the principal axis directions X_a, Y_b, and Z_c.

THREE-DIMENSIONAL DENSITY RECONSTRUCTION

We have used the occupancy weights and temperature parameters characterizing each hydration site to reconstruct the complete network of hydration according to both isotropic harmonic and anisotropic harmonic models. To this end, a grid with a mesh size of 1 Å, identical to the one used to compute the solvent density map in the previous paper, is generated for the theoretical models. The grid points $\mathbf{r}_{ijk}$ were initially set to the simulation average density value $\rho^1(\mathbf{r}_{ijk}) = \rho^0$ except those corresponding to simulation values smaller than $0.7 \times \rho^0$ which were set to zero. This was done to account for the region of the simulation box where the protein presence excludes the solvent to a high degree. In a second step, the density at the grid points around each hydration site was computed as a function of their distance from a site location according to the gaussian distribution characterized by the isotropic site temperature factor and occupancy weight. Cases of grid points being in the vicinity of several closely overlapping sites were treated by adding the density resulting from the contributions of the different neighbors.

The real space R-factors between the three-dimensional density distributions resulting from the simulation ρ_{ijk} and the reconstructed gaussian models ρ^g_{ijk} were calculated as follows:

$$R = \frac{\rho(\mathbf{r}_{ijk}) - \rho^g(\mathbf{r}_{ijk})}{\rho(\mathbf{r}_{ijk})}; \qquad |\mathbf{r}_{ijk} - \mathbf{r}_s| < 1.5\ \text{Å} \quad (15a)$$

for results concerning examples of particular sites, and

$$R_{tot} = \frac{\sum_s \sum_{i,j,k} [\rho(\mathbf{r}_{ijk}) - \rho^g(\mathbf{r}_{ijk})]}{\sum_s \sum_{i,j,k} \rho(\mathbf{r}_{ijk})}; \quad |\mathbf{r}_{ijk} - \mathbf{r}_s| < 1.5\text{Å} \quad (15b)$$

for the solvation region defined by the 551 hydration sites. The densities $\rho(\mathbf{r}_{ijk})$ and $\rho^g(\mathbf{r}_{ijk})$ calculated in terms of volume occupancy at the grid point $\mathbf{r}_{ijk}$ result, respectively, from the solvent trajectory and from either the isotropic or anisotropic gaussian model defined in Eqs. (7) and (13).

RESIDENCE TIMES AND WATER REORIENTATION

Residence times T_s^r were first determined for each hydration site by averaging the time periods during which a distinct solvent molecule oxygen center was continuously found within a 1.5 Å from a location site.

$$T_s^r = \frac{1}{N_{ex}} \sum_{j=1}^{N_w} \sum_{i=1}^{N_t} \delta\left(E\left[\frac{|\mathbf{r}_j(t_i) - \mathbf{r}_s|}{R_s}\right]\right) \times \Delta t \quad (16)$$

where R_s = 1.5 Å, and

$$N_{ex} = \sum_{j=1}^{N_t} \delta[I_w(j) - I_w(j+1)] \quad (17)$$

is the number of times water molecules in a given site are exchanged, N_w is the total number of water molecules, N_t is the total number of configurations, and $I_w(j)$ is the time-ordered series of the residue numbers of the water molecules populating each hydration site.

The determination of the reorientational correlation function is a useful tool to describe the water motions at the interface with proteins. It can be related to features that are observed in NMR, Raman, IR, and neutron spectroscopy spectra.[9,10] Single exponent relaxation times often characterizing the orientational correlation functions are usually interpreted as the actual residence times of water molecules inside different hydration shells around proteins.[11,12] Reorientational time correlation functions for water molecules in each hydration site, $C_s(t)$, have been computed as follows:

$$C_s(t) = \frac{1}{\mu^2} \sum_{t'=0}^{T} \sum_{\Delta t=0}^{T} [\boldsymbol{\mu}(t') \cdot \boldsymbol{\mu}(t' + \Delta t)]; \; t' + \Delta t < T \quad (18)$$

where T = 100 psec is the length of the equilibrated part of the simulation trajectory,[13] and $\boldsymbol{\mu}$ is dipole vector for a water molecule. The reorientation times were determined as the single exponent relaxation time, τ, best fitting the water dipole correlation functions in the [0.25–1.5] psec interval of time.

RESULTS

Correlation Between Thermal Motions and Occupancy

The B-factors were found to vary between 16 and 35 Å^2 as illustrated in Figure 1. Most of the points are located beneath and along the diagonal region thus indicating, as a general trend, the linear decrease of thermal fluctuations with occupancy increase. Hence, thermal fluctuations of water molecules in a given site appear to be correlated on average to its population. It is noteworthy that no

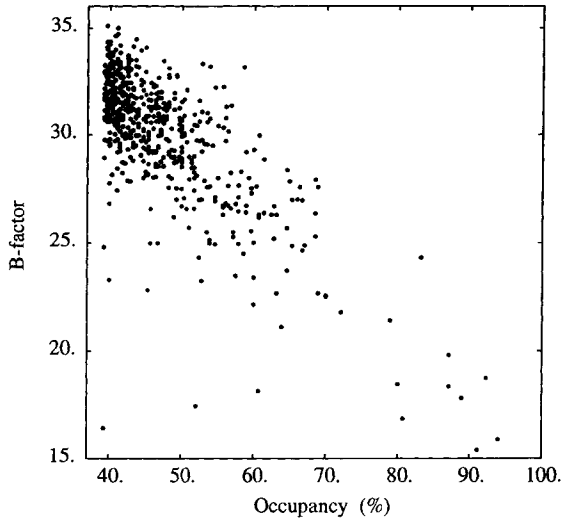

Fig. 1. Distribution of hydration site B-factors (Å^2) versus their occupancy weight in (%). B-factors are computed as the isotropic mean-square displacement of the solvent oxygen atoms within a 1.5 Å sphere centered at each hydration site.

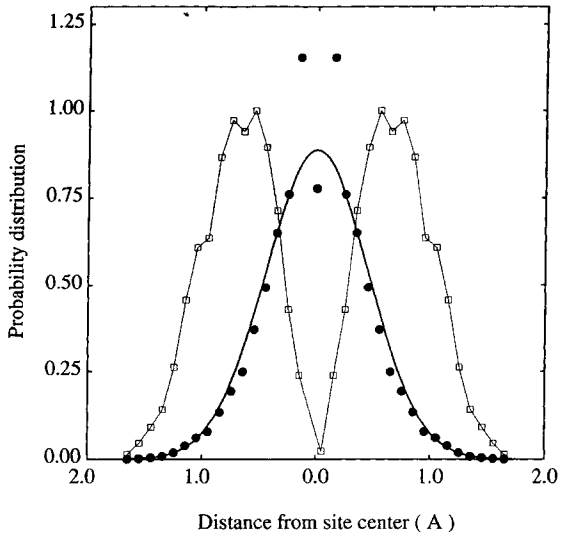

Fig. 2. Volume normalized probability distribution (black circles) of the water oxygens in the vicinity of site W154, computed for incremental values of the distance to the site center. The white squares represent the relative sampling for each point of the distribution. The thick solid line is the gaussian distribution reconstructed from the width and weight of the water oxygen distribution according to the procedure in the text. The data have been symmetrized.

point appears in the upper right part of this diagram excluding the possibility of high occupancy (strongly bound) water molecules with a large B-factor. However, the possibility of low B-factor for less occupied sites is not excluded and several examples were found.

The Isotropic Harmonic Model

Detailed analysis of the probability distribution of the water oxygen coordinates for the most resolved hydration sites exhibiting density contours of spheroidal shape has been made. Small values of $A_1 < 0.15$ characterize such sites indicating that the water oxygen distribution is reasonably isotropic. The dynamics of water molecules in this type of site is well represented as quasiharmonic with $|\alpha_3|$ and α_4 coefficients, respectively, in the [0.05–0.50] and [0.15–0.70] intervals, respectively.[1] Figure 2 shows the probability distribution of water for the particular example of site W154 whose density distribution is depicted in Figure 3a. This site W154 is characterized by a high occupancy weight W = 91.%, a low temperature B-factor of 16 Å^2, and a small anisotropy A_1 = 0.11. It is populated by water molecules sharing on average four simultaneous H-bonds at the respective average distances of 3.39, 3.28, 3.35, and 3.07 Å from Leu-2:N, Lys-133:O, Ala-134:O, and a neighboring water site which has a lower occupancy weight of only 42% (see Fig. 3a). A site such as W154 essentially becomes part of the protein structure itself by ensuring an effective linkage between the A and H helices which are the two ends of the polypeptide chain.[14] Figure 3b shows how the partial volume occupancy in the neighborhood of site W154 is remarkably fitted by the gaussian model with an R-factor of only 0.06. However, notice that, although very small, an inherent anharmonicity characterizes the water dynamics in site W154 as revealed by the skewness and excess kurtosis of the distribution observed in Figure 2 and reported in Table I where characteristics of sites with occupancy > 80% are gathered. There exist very few such hydration sites for myoglobin; they are buried in the protein surface at locations where water molecules share 2 or 3 or 4 H-bonds with the protein.

The reconstruction of the complete hydration region covered by the 551 resolved sites (see accompanying paper) has led to a global relative error, or R-factor, of 0.17 between the isotropic harmonic model and the actual simulation density distribution for water. Using a constant density $\rho^1(\mathbf{r}_{ijk}) = \rho^0$ for the grid points off the hydration region the R-factor for the entire solvent region was brought to 0.09. Using a step distribution of 2.8 Å spheres and the occupancy weight of each site the hydration network was reconstructed with an R-factor of 0.44. The relative error between the simulation singlet density $\rho^1(\mathbf{r}_{ijk})$ and a flat model at constant density ρ^0 for the protein hydration network region is R = 0.54. Thus the isotropic harmonic description leads to an improvement in R by a factor of 3 with respect to a step-function model.

Despite a reasonable global fit as evidenced by the R-factors, the previous reconstruction fails to reproduce the local three-dimensional details of most of

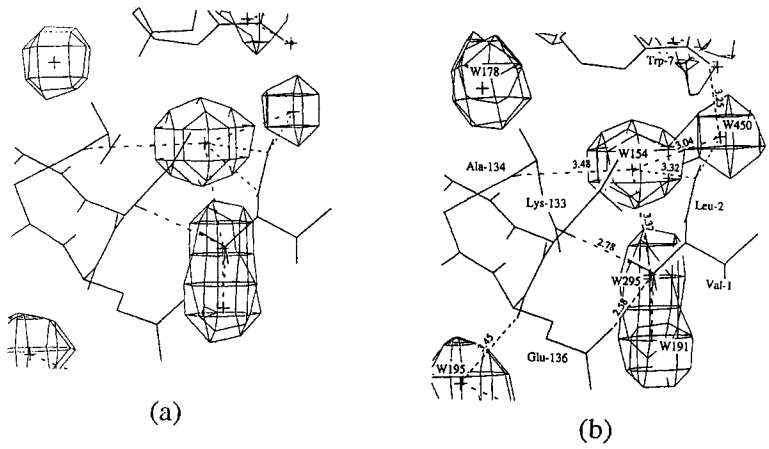

Fig. 3. (**a**) A view of the reconstructed water partial volume occupancy $\rho^1(\mathbf{r})$ in the vicinity of site W154 using an isotropic gaussian distribution with an occupancy weight of 91% and a temperature *B*-factor of 16 Å^2. Continuous solid line represents covalent bonds between the atomic sites and dashes describe hydrogen bonds. Numbers indicate the distances in Å between the bound atomic sites. (**b**) Actual density distribution computed from the trajectory for site W154 for the contour levels $\rho^1(\mathbf{r}) = 1.35\rho^0$.

TABLE I. List of Hydration Site With Occupancy ≥ 80%

Site	Location	d(Å)	W(%)	B_{fac}	$(A_1,\|\alpha_3\|,\|\alpha_4\|)$	R_{iso}/R_{ani}
W154	N:Leu-2	3.39	90	16.5	(0.11, 0.42, 0.22)	0.06/0.07
	O:Lys-133	3.28				
	O:Ala-134	3.35				
	W450	3.07				
W155	OE1:Gly-83	3.03	84	25.1	(0.65, 0.27, 0.69)	0.17/0.17
	OE2:Gly-83	3.04				
	W196	3.03				
	W229	3.36				
	W399	3.26				
	W553	3.01				
W157	O:Ala-74	3.24	91	19.4	(0.46, 0.23, 0.57)	0.17/0.20
	OE2:Glu-85	2.98				
	W183	3.27				
	W386	3.19				
	W500	1.99				
W158	O:Asp-127	3.13	84	21.9	(0.62, 0.20, 0.43)	0.18/0.10
	OD1:Asp-127	3.35				
	NH1:Arg-31	3.04				
	W174	3.37				
	W185	3.23				
W159	OE2:Glu-4	3.08	94	15.8	(0.07, 0.06, 0.14)	0.05/0.05
	NZ:Lys-79	3.06				
	W336	3.36				
W164	O:Leu-115	3.39	87	19.9	(0.11, 0.38, 0.27)	0.17/0.15
	W180	3.01				
W166	O:Thr-70	3.15	80	19.2	(0.35, 0.14, 0.20)	0.14/0.13
	NZ:Lys-77	3.24				
	W309	3.31				
	W665	3.02				
W167	NH2:Arg-31	3.12	80	23.4	(0.15, 0.37, 0.49)	0.14/0.09
W172	O:Leu-11	3.21	88	17.4	(0.41, 0.20, 0.26)	0.14/0.16
W180	OD2:Asp-122	3.15	88	21.5	(0.73, 0.33, 0.54)	0.14/0.11
	W164	3.01				

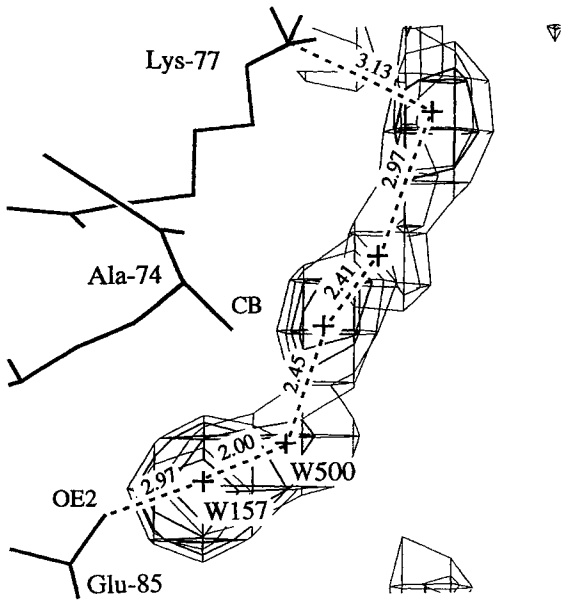

Fig. 4. View of the clustering aspect of the hydration network near site W157 and W500. The contour levels are identical to those defined in Figure 3b.

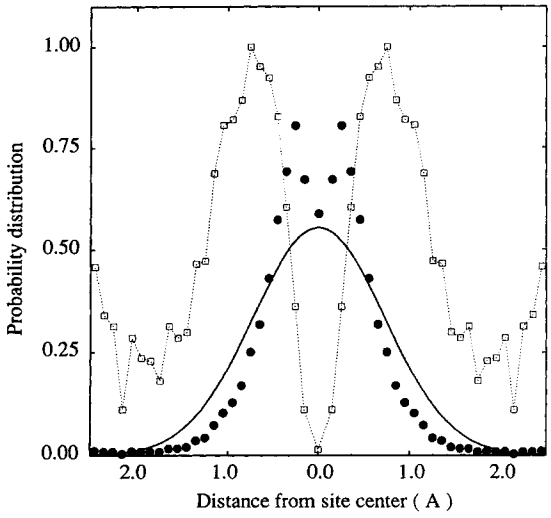

Fig. 6. Volume normalized spherically averaged positional distribution function (pdf) around the site W157 (black circles). The white squares represent the sampling frequency for each point of the pdf. The gaussian curve is obtained from the occupancy weight 91% and isotropic *B*-factor 19 Å^2 characterizing the site W157.

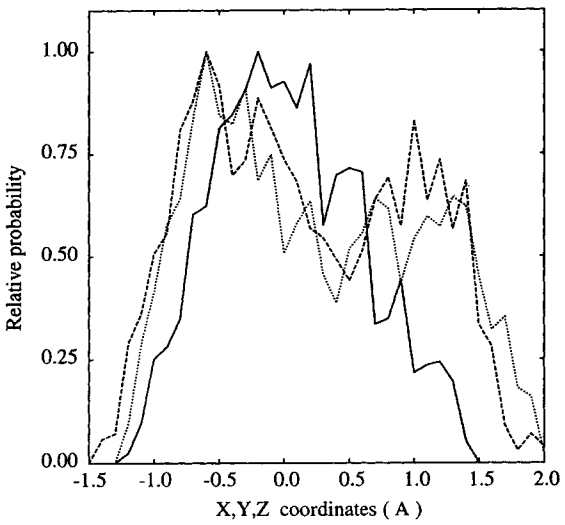

Fig. 5. Relative probability distribution along the *X* (solid line), *Y* (dashes), and *Z* (dots) directions of the simulation box for the positional fluctuations of water oxygens within 2.5 Å from the hydration site W157.

the hydration sites. The local anisotropy and anharmonicity of the solvent density distribution results from fluctuations close to ρ^0. An *R*-factor of 0.17 represents the fraction of the hydration network which cannot be fitted by the isotopic harmonic motions.

Anisotropic and Quasiharmonic Fluctuations

We investigate in this section the utility of the anisotropic harmonic model in characterizing the solvent fluctuations near the protein. To illustrate we have chosen to focus on the details of two distinct hydration sites, W157 and W500. The complexity of the average density distribution around the site W157 in each cartesian direction, which is depicted in Figure 4, can be regarded as a good example of anisotropy. Site W157 is included in a tube of density involved in the hydration of the hydrophobic residues Ala-74 and Ala-84. The equilibrium distances W157–C_β74 and W157–C_β84 are, respectively, 3.4 and 3.6 Å and a strong hydrogen bond is shared between W157 and Glu-85-$O_{\epsilon 1}$ at 2.9 Å. An average population of 91%, an anisotropy $A_1 = 0.46$, average $|\alpha_3| = \frac{1}{3}\Sigma_i|\alpha_3^i|$ and $\alpha_4 = \frac{1}{5}\Sigma_i\alpha_4^i$ of, respectively, 0.51 and 0.57 characterize the positional distribution of water in this hydration site.

The relative probability distributions of water oxygen centers in the *X*, *Y*, and *Z* directions around the site W157 are displayed in Figure 5. The strong anisotropy of the distributions in the *Y* and *Z* direction is due to the delocalization of the water positions over a close neighbor site W500 with a lower occupancy of 47% and at only 2.0 Å from site W157 (see Fig. 4). The spherically averaged unidirectional probability distribution of water is computed according to the method reported in the section on the Isotropic Harmonic Model and displayed in Figure 6. The deviation from the gaussian isotropic model is reflected by isotropic *R*-factors differences $R^{iso}_{w157} = 0.17$ and $R^{iso}_{w500} = 0.27$.

To further refine the reconstruction of sites W157 and W500 we assume that their dynamics can be represented by anisotropic harmonic motions. The

TABLE II. Comparison of Site W157 and W500

Site	W157	W500
Occupancy (%)	91.0	47.0
B-factor (Å^2)	19.4	32.9
Anisotropy A_1	0.46	0.87
Skewness		
$\lvert\alpha_3^x\rvert$	0.28	0.06
$\lvert\alpha_3^y\rvert$	0.30	0.11
$\lvert\alpha_3^z\rvert$	0.12	0.15
Excess kurtosis		
$\lvert\alpha_4^x\rvert$	0.68	0.57
$\lvert\alpha_4^y\rvert$	0.31	0.81
$\lvert\alpha_4^z\rvert$	0.64	1.22
$R_{\text{isotropic}}$	0.17	0.27
$R_{\text{anisotropic}}$	0.20	0.18

sites W157 and W500 were reconstructed using the procedure described in the section on Anisotropic Quasiharmonic Model with anisotropic R-factors $R^{\text{ani}}_{\text{w157}} = 0.20$ and $R^{\text{ani}}_{\text{w500}} = 0.18$, respectively. Surprisingly, the quality of the fit decreases slightly for W157 whereas it increases significantly for W500. The analysis of the moments of positional distribution for the respective sites W157 and W500 (Table II) shows that the origin of this effect is in the high anisotropy and the reduced asymmetry of site W500 as compared with site W157. Only the fourth moments α_4^i substantially deviate from zero indicating that the strong anisotropy $A_1 = 0.87$ of site W157 is the result of its strong kurtosis in one particular direction of space due to the presence of W500. Conversely, the more pronounced skewness and a less marked kurtosis which is equally distributed over two directions of space prevents its fit by an ellipsoidal distribution and leads to additional errors in the density reconstruction. The water positional fluctuations along the principal axis of the ellipsoid which are displayed in Figure 7 show how the water molecules populating the site W157 are also constantly visiting the neighbor site W500 which is located 2 Å away along the ellipsoid principal direction Z_c. Notice also the marked peakedness in the X_a direction. Figure 8a, depicts the hydration region around sites W157 and W500, respectively, as it appears from the anisotropic reconstruction (a) and the actual result from the simulation (b).

The density for the solvent region was reconstructed according to the section on the Anisotropic Quasiharmonic Model using probability distributions modeled by anisotropic gaussians instead of isotropic gaussians. A view of the myoglobin hydration network cut through the protein core resulting from the anisotropic reconstruction and the comparison with the MD simulation map is presented in Figure 9. The size and shape of the majority of sites

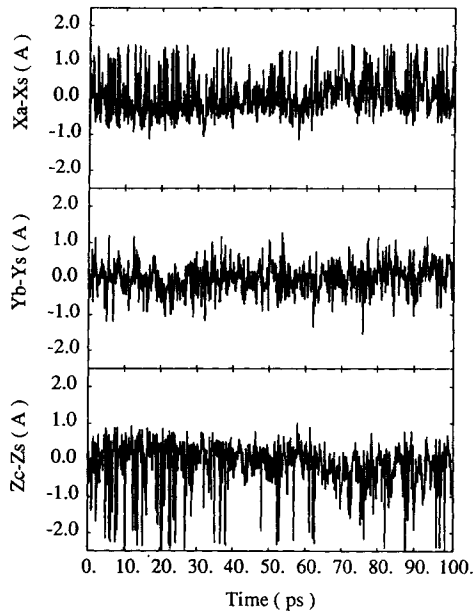

Fig. 7. Oxygen atom fluctuation versus time, for water in the hydration site W157, projected on the principal axis X_a, Y_b, Z_c of the ellipsoid obtained from diagonalization of the symmetric matrix formed with the set of B-parameters.

are fairly well reproduced except for a relatively few cases near an isotropic sites. The R-factor computed from the anisotropic gaussian distribution model is 0.13, which has to be compared with 0.17 found for the isotropic model. The respective distribution of the isotropic and anisotropic R-factors of each hydration site as a function of the occupancy weight is displayed in Figure 10. The major improvement occurs for hydration sites with poor R^{iso}-factors between 0.20 and 0.35 which exhibit significantly reduced R^{ani}-factors ≤ 0.20 for the whole range of occupancy from 40 to 80%. In addition, a noticeable fraction of hydration sites with occupancy between 40 and 50% has anisotropic R-factors below 0.10. Finally, the few sites with occupancy weights > 80% are not qualitatively affected by the passage from isotropic to anisotropic gaussian distributions and present R-factors ranging from 0.05 to 0.18.

Residence Times and Water Reorientation

As depicted in Figure 11, the overall majority of sites have relatively short average residence times varying between 1 and 5 psec which is two orders of magnitude shorter than the NMR results at temperatures near the freezing point.[15] Conversely, similar residence times between 1 and 4 psec were observed in a 210-psec MD simulation of BPTI in an explicit 2607 water surrounding.[16] A few examples correspond to long-lived sites as compared with the previous ones and are frequently associated with more intimately bound water molecules. Those sites have

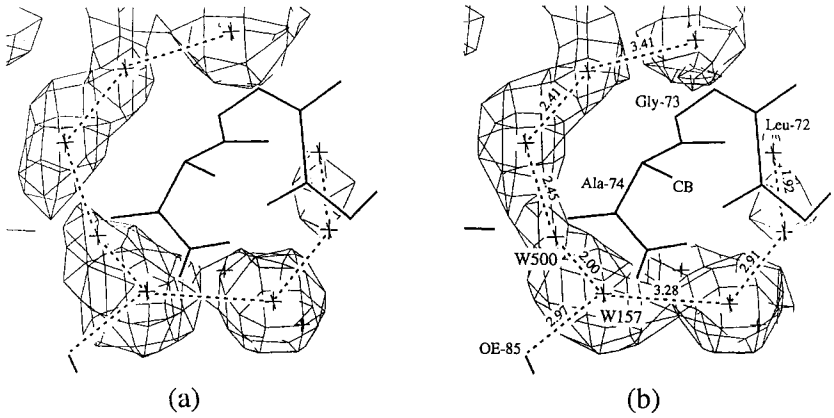

Fig. 8. View of the ring of solvation around the hydrophilic residue Ala-74 as it appears for the contour level $\rho^1(\mathbf{r}) = 1.35\rho^0$ of the density map reconstructed with the anisotropic harmonic model (**a**) and computed from the simulation trajectory (**b**). Crosses represent the hydration site mean positions, dashes symbolize H-bonds, and distances between sites are indicated in Å.

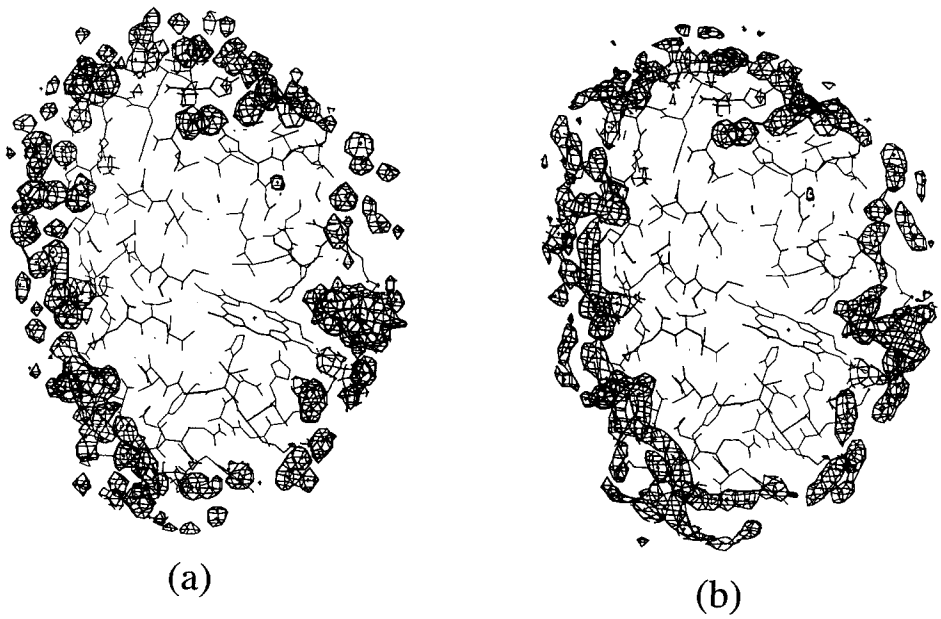

Fig. 9. Global comparison between the hydration network of myoglobin reconstructed with the anisotropic harmonic model (**a**) and the solvent density map obtained directly from the simulation after time averaging the water partial volume occupancy (**b**). The $\rho^1(\mathbf{r}) = 1.35\rho^0$ contour level is drawn for both (a) and (b) and the stick figure represents the protein time averaged atomic structure.

an average residence time greater than 10 and up to 30 psec. They uniformly appear in the lower part of Figure 12 where the number of distinct water molecules versus site occupancies is displayed. As depicted in Figures 11 and 12, the average residence time of water in each site is apparently not simply correlated to the occupancy weight in any case. Residence times found in the 1 to 30 psec range are in agreement with those resulting from computer simulation studies of coordinated water dynamics around small molecules and small protein crystals.[17,18]

Another important aspect of the water dynamics in biological materials is its response to alternating external fields.[19–21] Relaxation times of water surrounding (and in) myoglobin crystals have been studied in the microwave region over a wide range of temperature.[22] Interestingly, at 273 K no dielectric discontinuity was observed indicating that the 400 water molecules present in the crystal were per-

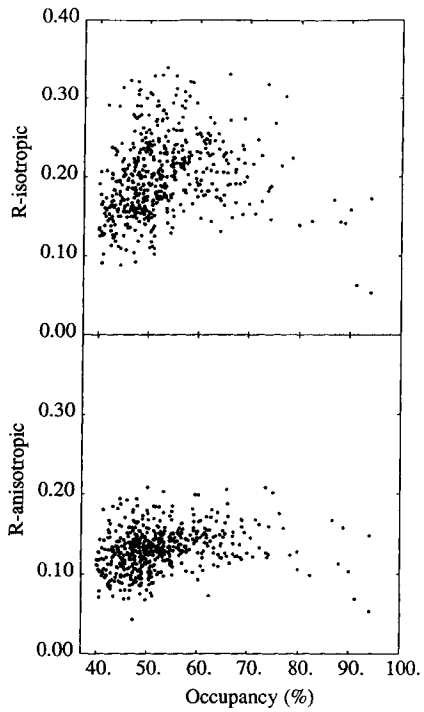

Fig. 10. Comparison between the *R*-factors obtained for each hydration site with both isotropic harmonic model (**a**) and anisotropic harmonic model (**b**) as a function of the occupancy weight (%).

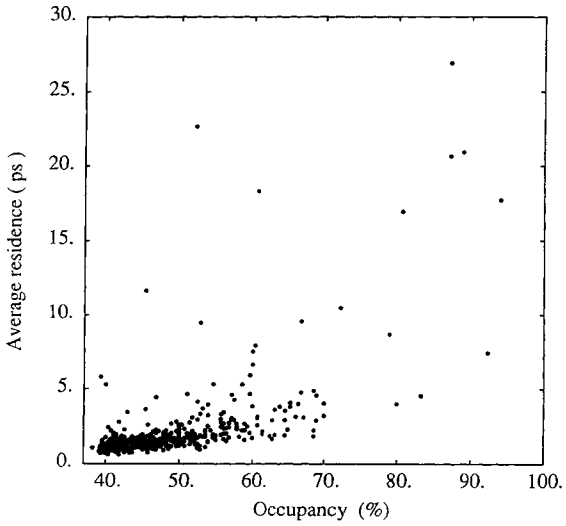

Fig. 11. Distribution of the average residence time (psec) versus the occupancy weight (%) for each of the 551 hydration sites within 7 Å from the myoglobin atomic surface.

turbed by the presence of the protein as expected from the general trend of colligative effects. Figures 13 and 14 show the comparison between the orientational behavior of two different types of hydration

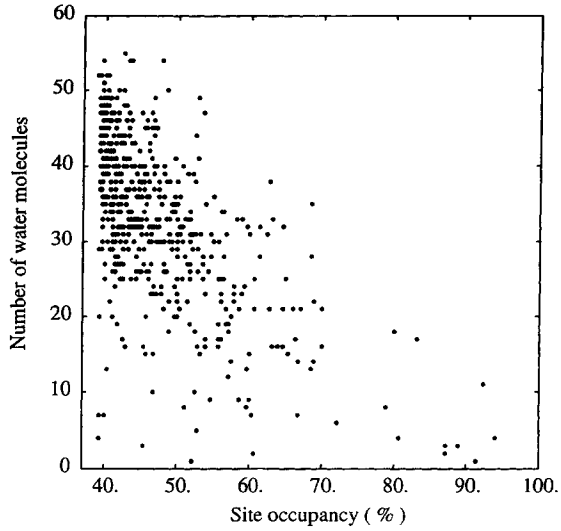

Fig. 12. Distribution of the number of distinct water molecules as a function of the site occupancy weight.

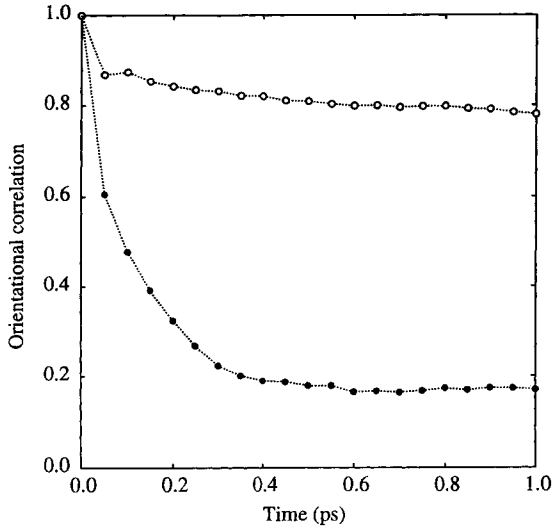

Fig. 13. Comparison of the dipole axis orientational correlation function for water molecules populating the sites W164 (white circles) and W163 (black circles).

sites W164 and W163, respectively, characterized by a high, 86%, and a low, 65%, occupancy weight. The detailed examination of the respective locations of site W164 and W163 with respect to the protein itself is diagrammed in Figure 15. The site W164 is connected to the protein surface via two H-bonds to Leu-115:O and Asp-122:$O_{\delta 2}$ and an additional H-bond with a neighboring hydration site W180 (itself connected to Asp-122:$O_{\delta 2}$ and Lys-15:N_{ζ}). Conversely, site W163 is only indirectly connected to the protein via other hydration sites. Site W164 is clearly involved in three well-defined H-bonds

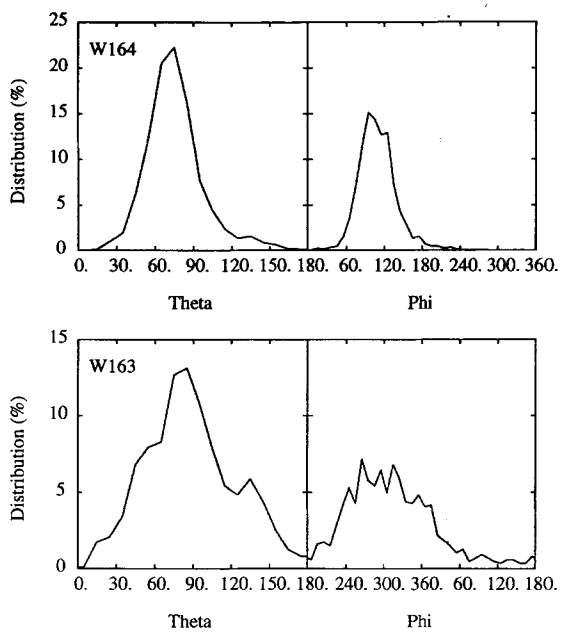

Fig. 14. Comparison of the θ and ϕ angular distributions of the water dipole axis fluctuations for the site W164 and W163. The reference frame is that of the protein backbone and the sampling runs over 100 psec. The angular distribution of the water dipole axis in the site W164 is consistent with librational motion around a single direction of the dipole axis (θ = 70° and ϕ = 110°). Conversely, multiple maxima in the θ and ϕ angular distributions for the site W163 indicates the equilibrium changes in orientation of the water dipole axis during the simulation.

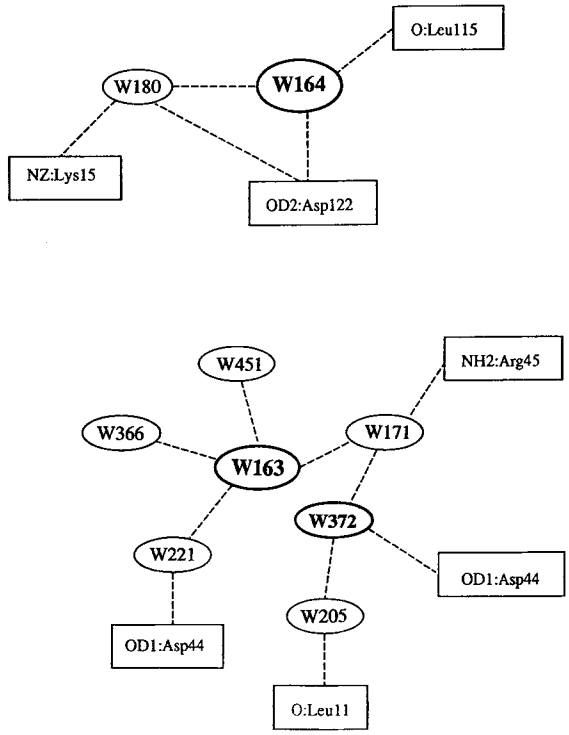

Fig. 15. Two-dimensional schematic description of two different types of hydration sites. Ellipses and dashes represent, respectively, hydration sites and H-bonds. W164 has an occupancy weight of 85% and is directly attached to the protein surface via two strong hydrogen bonds to polar and charged residues. W163 has a lower occupancy weight 60% and is more indirectly connected to the protein via H-bonding with intermediate protein bound water molecules. The sites W366 and W451 have no contact with the protein itself.

whereas the water molecules near the close neighbor pair of sites W163, W372 participate in a network of 8 possible H-bonds (cf. Fig. 15).

Site W164 is populated by very strongly oriented water molecules (cf. Fig. 14) as reflected by the very slow decay of the dipole correlation function (see Fig. 13). In contrast, W163 exhibits a rapid decay of orientational correlation for time less than 0.5 psec. Notice the shoulder observed at short time in the correlation function of W164 which is characteristic of a hydrogen bonded fluid. It is a signature of the librational motion of water molecules in the potential field of their neighbors. Figure 14 displays the normalized probability distribution of the dipole axis orientation of water molecules in sites W164 and W163, respectively. The θ (azimuthal) and ϕ (polar) angles of the dipole axis are calculated with respect to the frame defined by the protein backbone average orientation. The widths of the distributions in θ and ϕ angles computed as the square root of their variance are found to be 23°, and 31° for site W164 and 54° and 68° for site W163. The values around 25° found for site W164 are consistent with librational amplitude of water molecules bound to the protein surface whereas higher values around 60° for site W163 reflect the orientational disorder resulting from the possibility of multiple hydrogen bonds for a single hydration site. Notice the lack of the hydrogen bonding shoulder in the orientational correlation function for this type of site characterized by rotationally free water molecules.

As shown in the previous paper the occupancy weight and temperature factor of the hydration sites are correlated to the number of H-bonds shared with the protein. Here we consider the correlation of occupancy and reorientation (relaxation) times. The reorientation times of the water dipole axes were computed as the relaxation time characterizing the exponential regime best fitting the orientational correlation functions in the [0.25,1.5] psec interval. We found averaged reorientation times of, respectively, 2.3, 3.1, 3.9, 6.5, and 12.1 psec for water in the sites characterized by occupancy weights in the [40,50%], [50,60%], [60,57%], [70,80%], [80,100%] intervals, respectively. The decay of the orientational correlation function appears to be correlated on average to the occupancy weight, as depicted in Figure 16, and thus to the degree of interaction with the protein. The average correlation time of the dipole reorientation for water molecules located at more than 7 Å from the protein atomic surface was

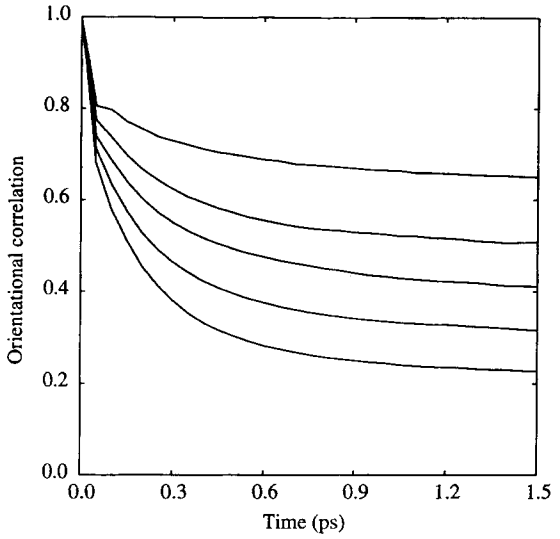

Fig. 16. Average orientational correlation function of the water dipole axis computed for 5 intervals of occupancy weight between 40 and 100%. The curves a, b, c, d, e correspond to hydration sites in the respective [40–50%], [50–60%], [60–70%], [70–80%], and [80–100%] ranges of occupancy. The dashed curve is the mean correlation function for the simulation bulk, i.e., water molecules at more than 7 Å from the protein surface.

computed to be 2.0 psec, which can be viewed as the bulk value for the simulation.

DISCUSSION

The equilibrium aspects of the results presented in the previous section are in excellent agreement with those recently obtained with a Monte Carlo simulation of metmyoglobin in 814 explicit water molecules.[23] In particular, an identical R-factor of 0.17 was obtained for the solvent region characterized by water molecules with a B-value of 50 Å^2. However, the R-factor values we found for the myoglobin hydration network may not be an intrinsic limit. Indeed, the procedure used has introduced two major biases in the reconstruction. The first occurs in the determination of the site loci where the arbitrary choice of a 1.8 Å lower limit in the site separation distance is introduced. The second concerns the B-parameters evaluation, which is done with cartesian coordinates found within 1.5 Å from each hydration site. Despite these two limitations the fit is globally good (see Fig. 10) and confirms that a priori knowledge of the local water equilibrium, which is unknown from X-ray experiments, is essential for the probability density assignment in the solvent boundary region.

However, as reflected by a relatively small reduction of the R-factor when going from isotropic to anisotropic gaussian distributions, the determination of the local maxima of three-dimensional density $\rho^1(\mathbf{r})$ appears to be more important than a more exact description of the thermal motions. Specifically,

TABLE III. Pearson's Coefficient for Various Distributions

Distribution	r*	Distribution	r
R_{iso}/weight	0.20	R_{iso}/A_1	0.02
R_{ani}/weight	0.23	R_{ani}/A_1	0.20
R_{iso}/α_3	0.06	R_{iso}/α_4	0.02
R_{ani}/α_3	0.25	R_{ani}/α_4	0.04
$A_1/\lvert\alpha_3\rvert$	0.17	A_1/α_4	−0.25
$\lvert\alpha_3\rvert/\alpha_4$	−0.19		

*Pearson's coefficent or linear correlation coefficient r.

the individual R-factors of the 551 hydration sites resulting from the methodology we have used are only weakly correlated to the occupancy weight and uncorrelated to either the anisotropy, skewness, or kurtosis parameters, as reported in Table III, when the isotropic gaussian model is used. Conversely, weak correlations to those parameters emerge when using the anisotropic gaussian model and reflect a slight improvement of the overall fit. The refinement of the solvent density region by the use of complete Debye–Waller tensors would not be a trivial problem and might even lead to spurious fitting in some particular cases as depicted in the Results section. However, a complete Debye–Waller tensor can provide an excellent fit when the anisotropy is essentially unidirectional. Crystallographically speaking, the passage from isotropy to anisotropy results in a 6-fold increase in the number of refinement parameters for the purpose of including a better representation of the atomic thermal fluctuations. In this regard the gain of efficiency of anisotropic versus isotropic fit found in this study for solvent would be very poor.

The residence time of water molecules in the different hydration sites around proteins cannot be properly assessed by time-independent X-ray crystallography. A peak in the electron density of a real-space map indicates only the overall probability of finding an atom at a given location. This probability accounts for structural variations between the different unit cells of the crystal as well as for the actual residence time of an occupied site.[24,25] However, NMR techniques, recently developed and applied to biomolecules in aqueous solutions, prove to be informative on both positions and residences times of water inside and around proteins.[26–28]

However, independent of instrumental limitations correlation times are difficult to estimate from experimental results because it is necessary to assume models to interpret the time scales and magnitudes of the motions involved. Usually, the reorientation times measured by NMR, ESR, and EPR methods are in the 10^{-7} to 10^{-10} psec range for the water near the protein surface with a ratio to bulk of less than 1/100.[29–35] In addition, few water molecules appear immobile to NMR time scales. For in-

stance, the high resolution proton nuclear magnetic resonance (3D^{1}H-NMR) studies of polypeptides and proteins in aqueous solution show two distinct classes of hydration water molecules. A well-defined, very small number of water molecules found in the interior of proteins have identical locations in the crystal structure and in solution. Their residence time is measured in the range from 10^{-2} to 10^{-8} which is beyond the time scale of the present simulation. However, protein surface hydration seems to involve water molecules with residence times substantially shorter, in the subnanosecond range, 100 to 500 psec for temperatures less than 10°C.[36,37] Those results suggest that the residence times at room temperature might be even shorter.

Other types of resonance experiments, ^{17}O and ^{2}H, carried out on lysozyme and other proteins in powders and solutions were characterized by correlation times in the picosecond time domain.[38–40] For lysozyme, discontinuities in the NMR response were interpreted in terms of three different classes of water. A restricted number of water molecules, about 20 per lysozyme molecule, were found with a correlation time of 41 psec. About 140 and 1400 water molecules were characterized by, respectively, 27 and 17 psec correlation times. Thus, the results presented in this section are generally consistent with the experiment performed on lysozyme, which is a globular protein of size comparable to myoglobin. The dozen hydration sites with occupancies greater than 80% which present reorientation time significantly longer than other sites may be assigned to the less mobile water molecules or equivalently those with a 41 psec correlation time in the lysozyme NMR experiment.

The correlation times found for water of hydration are comparable with the average residence times described in the section on Residence Times and Water Reorientation confirming that the rate of H-bonds breaking is related to the lifetime of water molecules in a given hydration site. Consequently, reorientation times deduced from resonance experiments NMR, ESR, and EPR can probably be regarded as reliable estimates of the water average residence time around proteins.[17]

CONCLUSION

In this study we investigated the dynamics of water at the protein–solvent interface using a computer simulation of metmyoglobin in water. The conceptual distinction between hydration sites (density maxima) and distinguishable water molecules proved useful in resolving ambiguities in comparison of the density distribution implied dynamics with the experimental data. Correlating between the simulated site occupancies and the thermal motions revealed by B-factors showed that while low B-factors were occasionally found for sites of low occupancy, no sites with a high occupancy and a large B-factor were observed.

Water sites were subjected to a mock refinement. Both a simple isotropic harmonic model and an anisotropic model were considered for the refinement. While the anisotropic model resulted in an R-factor of 0.13 compared with only 0.17 for the harmonic isotropic model, reflecting the expected overall improvement, occasionally individual sites were found for which the quality of the fit actually slightly decreased even with the larger number of parameters. This unwanted apparent correlation in the fitting parameters was traced to strong kurtosis. Similar problems had been seen previously in mock refinements for proteins.

Residence times and dipole reorientation times for the water around the protein surface were compared with NMR and ESR experimental results. For occupancies less than 50% low average residency times for water molecules at a particular hydration site were seen. High occupancies showed much less correlation with the actual residency time at a site. This gives a different interpretation to refined water positions from crystallographic data. The simulation implies that in many cases, the larger the refined occupancy, the less likely a water molecule of long residence time has caused it.

ACKNOWLEDGMENTS

The NIH, Alfred P. Sloan Foundation, and the Robert A. Welch Foundation are thanked for partial support. B. M. P. thanks L. Findsen and S. Subramanian for many conversations in the early stages of this work.

REFERENCES

1. Brooks, C.L., Karplus, M., Pettitt, B.M. Proteins: A theoretical perspective on structure, dynamics and thermodynamics. In: "Advances in Chemical Physics," Vol. 71, Rice, S., Prigogene, I. eds. New York: John Wiley & Sons, 1988; Ichiye, T. The internal dynamics of proteins. Thesis Dissertation, Harvard University, Cambridge, Massachusetts, 1985.
2. Willis, B.T.M., Pryor, A.W. Thermal vibrations in crystallography. Cambridge: Cambridge University Press, 1975.
3. Konnert, J.H., Hendrickson, W.A. A restrained-parameter thermal-factor refinement procedure. Acta Crystallogr. A36:344–350, 1980.
4. Kuriyan, J., Karplus, M., Levy, R.M., Petsko, G.A. Effect of anisotropy and anharmonicity on protein crystallographic refinement. J. Mol. Biol. 190:227–254, 1986.
5. Stout, G.H., Jensen, L.H. "X-ray Structure Determination: A Practical Guide." New York: John Wiley & Sons, 1989.
6. Savage, H., Wlodawer, A. Determination of water structure around biomolecules using X-ray and neutron diffraction methods. Methods Enzymol. 127:162–183, 1986.
7. Yu, H., Karplus, M., Hendrickson, W.A. Restraints in temperature-factor refinement for macromolecules: An evaluation by molecular dynamics. Acta Crystallogr. B41:191–205, 1985.
8. Kendall, M.G., Stuart, A. "The Advanced Theory of Statistics," I. London: Charles Griffin, 1977.
9. Madden, P.A., Impey, R.W. Dynamics of coordinated water: A comparison of experiment and simulation results. Ann. N.Y. Acad. Sci. 482:91–113, 1988.
10. Impey, R.W., Madden, P.A., Tildesley, D.J. On the calcu-

lation of the orientational correlation parameter g2. Mol. Phys. 46:1319–1334, 1982.
11. Koenig, S.H. The dynamics of water-protein interactions: Results from measurements of nuclear magnetic relaxation dispersion. ACS Symp. Ser. 127:157–176, 1980.
12. Eisenstadt, M. NMR relaxation of protein and water protons in diamagnetic hemoglobin solutions. Biochemistry 24:3407–3421, 1985.
13. Findsen, L.A., Subramaniam, S., Lounnas, V., Pettitt, B.M. Molecular dynamics simulation of metmyoglobin in aqueous solution. In "Principle of Molecular Recognition." Buckingham, A.D., ed. London: Chapman and Hall: 1993.
14. Malin, R., Zielenkiewicz, P., Saenger, W. Structurally conserved water molecules in ribonuclease T1. J. Biol. Chem. 266(8):4848–4852, 1991.
15. Otting, G., Liepsinsh, E., Wüthrich, K. Proton exchange with internal water molecules in the protein BPTI in aqueous solution. J. Am. Chem. Soc. 113:4363–4367, 1991.
16. Levitt, M., Sharon, R. Accurate simulation of protein dynamics in solution. Proc. Natl. Acad. Sci. U.S.A. 85:7557–7561, 1988.
17. Madden, P.A., Impey, R.W. Dynamics of coordinated water: A comparison of experiment and simulation results. Ann. N.Y. Acad. Sci. 482:91–113, 1988.
18. Berendsen, H.J.C., van Gunsteren, W.F., Zwinderman, H.R.J., Geurtsen, R.G. Simulation of proteins in water. Ann. N.Y. Acad. Sci. 482:269–286, 1988.
19. Pethig, R. "Dielectric and Electronic Properties of Biological Material." New York: John Wiley & Sons, 1979.
20. Pethig, R., Kell, D.B. The passive electrical properties of biological systems. Phys. Med. Biol. 32:933–970, 1987.
21. Parak, F. Correlation of protein dynamics with water mobility: Mössbauer spectroscopy and microwave absorption methods. Methods Enzymol. 127:196–206, 1986.
22. Singh, G.P., Parak, F., Hunklinger, S., Dransfeld, K. Role of adsorbed water in the dynamics of metmyoglobin. Phys. Rew. Lett. 47:685–688, 1981.
23. Parak, F., Hartmann, H., Schmidt, M., Corongiu, S., Clementi, E. The hydration shell of myoglobin. Eur. Biophys. J. 21:313–320, 1992.
24. Frauenfelder, H., Petsko, G.A., Tsernoglou, D. Temperature-dependent X-ray diffraction as a probe of protein structural dynamics. Nature (London) 280:558–563, 1979.
25. Artymiuk, P.J., Blake, C.C.F., Grace, D.E.P., Oatley, S.J., Phillips, D.C., Sternberg, M.J.E. Crystallographic studies of the dynamic properties of lysozyme. Nature (London) 280:563–568, 1979.
26. Kuntz, I.D., Jr., Kautzmann, W. Hydration of proteins and polypeptides. Adv. Protein Chem. 28:239–345, 1974.
27. Fung, B.M. Nuclear magnetic resonance study of water interactions with proteins, biomolecules, membranes, and tissues. Methods Enzymol. 127:151–161, 1986.
28. Bryant, R.G. Magnetic resonance and macromolecule solvation dynamics. Stud. Phys. Theor. Chem. 38:683–705, 1988.
29. Bryant, R.G., Shirley, W.M. Water-protein interactions: Nuclear magnetic resonance results on hydrated lysozyme. Biophys. J. 32:3–16, 1980.
30. Bryant, R.G., Shirley, W.M. Dynamical deductions from nuclear magnetic resonance relaxation measurements at the water-protein interface. ACS Symp. Ser. 127:147–156, 1980.
31. Shirley, W.M., Bryant, R.G. Proton-nuclear spin relaxation and molecular dynamics in the lysozyme-water system. J. Am. Chem. Soc. 104:2910–2918, 1982.
32. Peemoeller, H., Kydon, D.W., Sharp, A.R., Schreiner, L.J. Cross relaxation at the lysozyme-water interface: An NMR line-shape-relaxation correlation study. Biophys. J. 49: 943–948, 1986.
33. Steinhoff, H.J., Lieutenant, K., Schlitter, J. Residual motion of hemoglobin-bound spin labels as a probe for protein dynamics. Naturforsch. C: Biosci. 44:280–288, 1989.
34. Belonogova, O.V., Frolov, E.N., Illyustrov, N.V., Likhtenshtein, G.I. Effect of temperature and degree of hydration on the mobility of spin labels in surface layers of proteins. Mol. Biol. (Moscow) 13:567–576, 1979.
35. Steinhoff, H.J. A simple method for determination of rotational correlation times and separation of rotational and polarity effects from EPR spectra of spin-labeled biomolecules in a wide correlation range. J. Biochem. Biophys. Methods 17:237–247, 1989.
36. Otting, G., Liepinsh, E., Wüthrich, K., Protein hydration in aqueous solution. Science 254:974–979, 1991.
37. Otting G., Wüthrich, K. Studies of protein hydration by direct NMR observation of individual protein-bound water molecules. J. Am. Chem. Soc. 111:1871, 1989.
38. Lioutas, T.S., Baianu, I.C., Steinberg, M.P. Oxygen-17 and deuterium nuclear magnetic resonance studies of lysozyme hydration. Arch. Biochem. Biophys 247:68–75, 1986.
39. Halle, B., Anderson, T., Forsen, S., Lindman, B. Protein hydration from water-oxygen-17 magnetic relaxation. J. Am. Chem. Soc. 103:500–508, 1981.
40. Bryant, R.G. Magnetic resonance and macromolecule solvation dynamics. Stud. Phys. Theor. Chem. 38:683–705, 1988.