PROTEINS: Structure, Function, and Genetics 18:133–147 (1994)

A Connected-Cluster of Hydration Around Myoglobin: Correlation Between Molecular Dynamics Simulations and Experiment

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

ABSTRACT **An analysis of a molecular dynamics simulation of metmyoglobin in an explicit solvent environment of 3,128 water molecules has been performed. Both statics and dynamics of the protein–solvent interface are addressed in a comparison with experiment. Three-dimensional density distributions, temperature factors, and occupancy weights are computed for the solvent by using the trajectory coordinates. Analysis of the hydration leads to the localization of more than 500 hydration sites distributed into multiple layers of solvation located between 2.6 and 6.8 Å from the atomic protein surface. After locating the local solvent density maxima or hydration sites we conclude that water molecules of hydration positions and hydration sites are distinct concepts. Both global and detailed properties of the hydration cluster around myoglobin are compared with recent neutron and X-ray data on myoglobin. Questions arising from differences between X-ray and neutron data concerning the locations of the protein-bound water are investigated. Analysis of water site differences found from X-ray and neutron experiments compared with our simulation shows that the simulation gives a way to unify the hydration picture given by the two experiments.**

Key words: myoglobin, simulation, hydration

INTRODUCTION

Diffraction experiments carried out on protein crystals provide the time averaged spatial correlation existing between the different elements of proteins at the microscopic level.[1,2] Historically, myoglobin was the first protein for which a complete set of atomic coordinates was resolved thus revealing the intimate detail of the eight α-helices folded together into a packed tertiary structure.[3,4] Because of a simple structure relative to other proteins and its importance in the metabolism of mammals, for dioxygen storage, myoglobin has been extensively considered from both theoretical and experimental point of views.[5] Despite its seeming simplicity, the structure–function relationship of myoglobin is not yet fully understood.[6–10]

The chemical nature of water, the natural medium of biologically active molecules plays a determinant role in protein structure, dynamics and functionality.[11–16] Specifically, the biological activity of myoglobin which consists in binding O_2 molecules reversibly occurs via complex cooperative structural changes where solvent and protein dynamics are coupled over various time scales.[17] Thus, the bath of water molecules around proteins may preserve protein functionality by not only maintaining a proper tertiary structure but also by participating in dynamic structural fluctuations.

Unfortunately, the X-ray diffraction technique that has brought such a wealth of detail regarding the proteins themselves does not always provide precise locations for more than a few dozen water molecules tightly bound at the surface or tucked away in small interior pockets. Detailed solvation structure results are frequently obtained for proteins with a number of amino acid residues typically near 50, such as rubredoxin, 2 Zn-insulin, crambin, and bovine pancreatic trypsin inhibitor for which, respectively, 127, 340, 80, and 63 ordered water molecules have been reported by X-ray studies at resolution better than 1.5 Å.[18]

The smaller the globular protein, often the more packed its crystal. A large percentage of the solvent in the crambin crystal appears to be ordered (about 80%). This is not the case for many crystals of larger proteins where the solvent can occupy the majority of the unit cell volume, nearly 62% in the case of wildtype myoglobin in the P6 crystal form.[19] In such a case, most of the solvent region appears as relatively featureless contours of electron density maps that are essentially obscured by the level of experimental noise. For proteins of the size of myoglobin the resolution is usually in the 1.5–2.5 Å interval. At this level of resolution the solvent features are

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.

evasive except for a few well ordered sites of the first hydration layer and only very few of the second layer. Thus, recent 1.9 Å resolution X-ray data on sperm whale wild-type myoglobin currently reports 117 proximal water molecules.[20] That number is far smaller than the 600 water molecules that are estimated for complete coverage of the protein surface in solution[21] even when crystal contacts (which vary for different space groups) are taken into account.

The apparent reasons of the solvent invisibility are its high mobility, even when trapped in the crystal lattice, as well as the reduced scattering cross section per unit volume as compared with the protein core. Since the different contributions of water and protein to the total diffracted intensity $I = F(\mathbf{S})^2$ cannot simply be separated in the evaluation of the total structure factor, $F(\mathbf{S})$, it is necessary to introduce a solvent model (in addition to the protein) in the course of the refinement procedure. A crude but useful approximation consists in flattening the solvent in the region of the electron density map extending outward from the protein–water interface.[22] The solvent description is then augmented in the course of refinement by introducing models of the highly ordered water molecules at positions located by visual inspection of the difference Fourier map $2F_{obs}(\mathbf{r}) - F_{cal}(\mathbf{r})$.

At lower resolutions, typically between 6 and 4 Å, reasonable R-factors between 0.15 and 0.20 are obtained for the protein itself and can approach 0.06 (which is often near the level of noise in the data) when the solvent density map obtained in the final refinement cycle is added.[23] However, at higher resolution, between 2 to 1.5 Å R-factors again increase toward values greater than 0.15. This is, in part, a consequence of the nonrandom distribution of the water molecules beyond those in immediate contact with the protein surface. Thus, solvent flattening procedures which are effective at revealing protein at a crude level of detail from low resolution X-ray data are actually a source of errors at higher resolution.

Neutron diffraction is another very powerful technique which can be used to analyze the hydration of protein crystals.[24,25] Although similar in many aspects to X-ray experiments, neutron diffraction differs by the nature of the physical interaction involved in the scattering process. Instead of probing an electron probability, neutrons scatter according to the nuclear spin state and isotopic variety of species encountered. The use of H_2O and D_2O as solvents in neutron protein crystallography permits the determination of some bound water molecules. Furthermore, the different neutron scattering length of deuterium atoms (versus hydrogen) enhances the effective solvent density by a factor of two as compared with that observed with X-ray maps.[26] Even though this technique presents some problems in the analysis of larger biopolymers, it has been successfully employed on myoglobin crystals and has demonstrated that a better resolution of the solvent structure leads to an improved overall agreement between observed and calculated structure factors. Two distinct hydration layers were characterized and 87 water molecules were located in that study.[27] Interestingly, the spatial locations of the hydration sites as compared with X-ray results differ.

Strong evidence from sorption thermodynamic measurements, microwave absorption, NMR, ESR, and Raleigh scattering of Mössbauer radiation spectroscopy (RSMR) experiments performed on hydrated protein powders demonstrates that the diffraction picture of protein solvation is incomplete.[28–30] For instance, the specific heat capacity (C_p) isotherms measured as a function of the mass ratio, h, of added water versus protein exhibits discontinuities revealing chemical changes in the nature of hydration. For myoglobin the point of full hydration h, which defines the amount of added water beyond which the measured property does not significantly deviate from bulk, varies from 0.39 in microwave absorption to 0.6 in RSMR. This equivalently represents 400 to 600 nonbulk water molecules which are characterized by a changed rotational and translational mobility as compared with bulk.

Earlier molecular dynamics (MD) simulation studies of proteins in environment similar to their crystal form or in aqueous solution have attempted to correlate the solvent distribution with X-ray data. These simulations were performed on small peptide crystals and proteins such as BPTI.[31–34] They have demonstrated that agreement between X-ray and simulation data could be improved by considering the relative locations of hydration water molecules around peptides. However, absolute positions with respect to the crystal lattice could not be obtained with accuracies significantly better than 1 Å. The following work addresses the general problem of hydration structure for globular proteins by comparing refined diffraction densities with simulated results. This has been carefully considered previously for the protein density.[35]

Diffraction experiments provide the time averaged three-dimensional positions of atoms in a protein system, myoglobin in our case. The dynamic information is essentially encoded in the temperature factor (with the exception of occupancies) which contains some errors and uncertainties due to crystal disorder and conformational variability.[36,37] On the other hand, atomic loci are the explicit result of MD simulations. In this study we compute the spatial density distributions $\rho^1(\mathbf{r})$ analogous to the observed experimental density distributions and simultaneously keep track of the possible dynamic effects characterizing the different solvent regions around the protein. The methods used follow in the next section. Then the results are presented. A de-

tailed comparison with X-ray and neutron diffraction follows ending with our conclusions.

METHOD

The MD trajectory we used resulted from a previous study and was generated in the (N,V,T) ensemble with the Amber force field and suite of programs.[38,39] Briefly, a 170-psec molecular dynamics simulation of metmyoglobin in an explicit aqueous environment of 3,128 SPC water molecules was analysed. The water molecules occupied the majority of the volume in the 56.32 × 56.32 × 44.45 Å^3 simulation box. The MD simulation contained an ensemble of 10,918 trajectories, one for each explicit atom included in the modeled myoglobin–water system. Each trajectory runs over a 170-psec period of time and was analyzed as 3400 phase space points separated by 50 fsec (25 time steps). Volume normalized singlet density distributions, $\rho^1(\mathbf{r})$, have been previously computed for myoglobin and water, respectively, using a three-dimensional grid of points with an even spacing of 1 Å.[40] A small spherical volume of 0.5 Å radius centered at each grid point was used as a volume element $\Delta\tau$ to probe the presence of water for each lattice location in space $\mathbf{r}_{i,j,k}$. For every snapshot, volume intercepts $\Delta v_{i,j,k}$ between the effective spherical core of the SPC water molecules with a diameter of 2.8 Å, and the neighboring volume elements were computed, time averaged, and stored to give partial volume fraction occupancies for water;

$$\langle P^1_{\mathrm{wat}}(\mathbf{r}_{i,j,k})\rangle_{\mathrm{time}} = \frac{1}{\Delta\tau}\sum_{i_t=1}^{N_t}\sum_{i_w=1}^{N_w}\Delta v_{i,j,k}$$

where i_t, i_w, N_t, and N_w, are the time increment, the water molecules indices, the total number of configurations included in the average, and the total number of water molecules. Similarly, the protein volume occupancy $\rho^1_{\mathrm{prot}}(\mathbf{r}_{i,j,k})$ was computed using different distances for the C, N, O, and H atomic sites (respectively, 2.0, 1.8, 1.6, and 1.0 Å). The sum $\rho^1_{\mathrm{tot}}(\mathbf{r}) = \rho^1_{\mathrm{prot}}(\mathbf{r}) + \rho^1_{\mathrm{wat}}(\mathbf{r})$ is the simulation equivalent of the real space density maps obtained after Fourier synthesis of the structure factors measured in X-ray and neutron experiments.[1,2,24,25]

To further develop the parallel with crystallography we have systematically searched for the local maxima of the solvent density distribution. We have used a procedure which starts with finding grid points bearing density values $\rho^1(\mathbf{r}_{i,j,k})$ greater than a fixed minimum value ρ_{min} (see Fig. 1a). Then, the coordinates of these grid points are replaced by the density weighted average coordinates of their six closest grid points in the *X*, *Y*, and *Z* directions. The occupancy weights of the initial grid points are attributed to their new coordinates which collect in groups near the local maxima of the solvent density contour, cf. Figure 1b. Then, each new point is av-

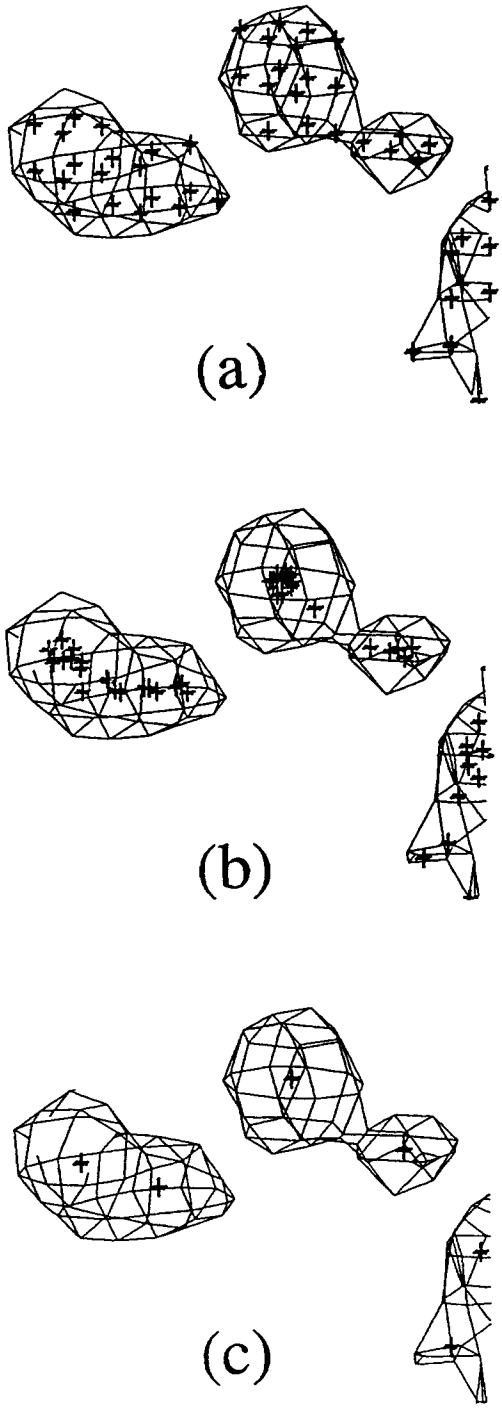


Fig. 1. A schematic summary of the method used to determine the density contour local maxima. The grid points inside the $\rho^1(\mathbf{r}) = 0.4$ contour are collected (**a**). Each grid point is replaced by the weighted average position of its closest neighbors in the *X*, *Y*, and *Z* directions (**b**). The points obtained after step (b) are averaged among themselves within a given grid region (**c**).

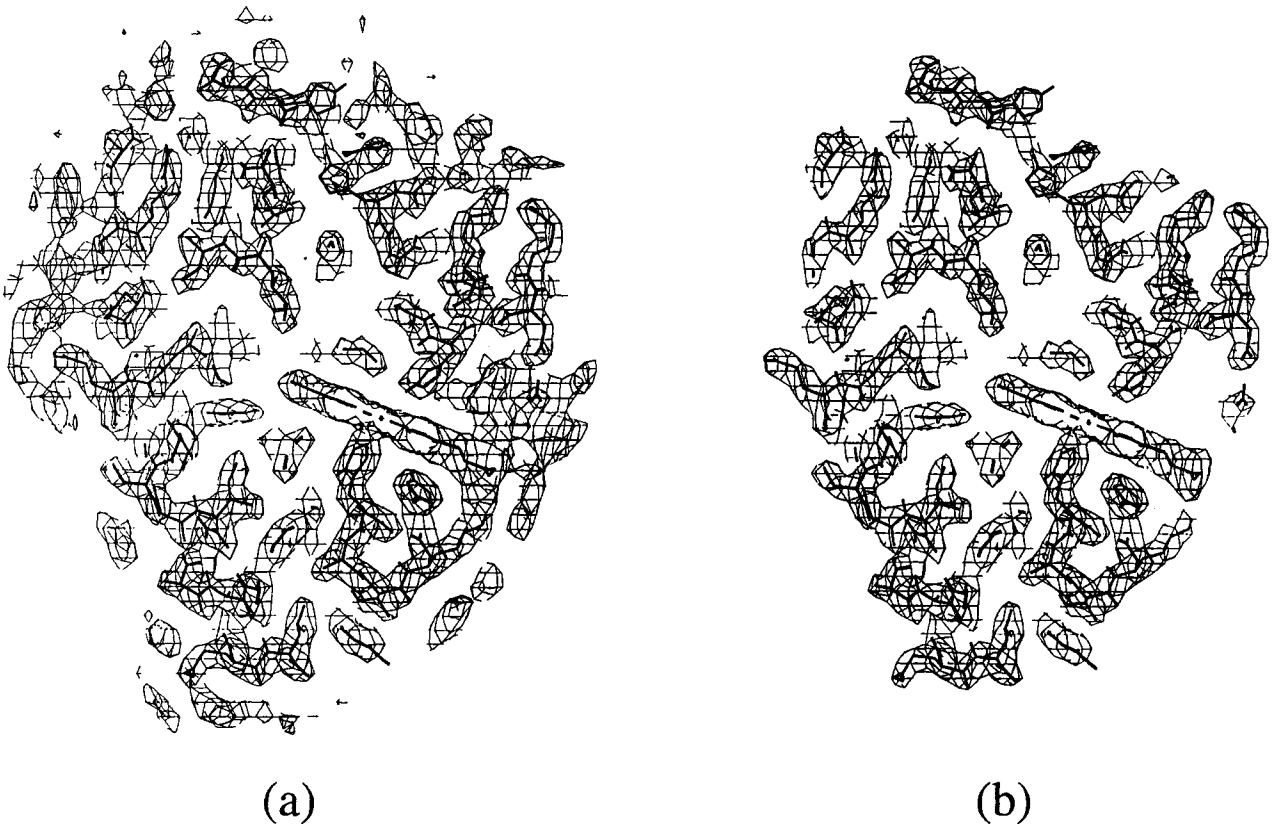


Fig. 2. A two-dimensional projection of the three-dimensional density distributions for the protein/water system $\rho^1(\mathbf{r})_{tot} = \rho^1(\mathbf{r})_{prot} + \rho^1(\mathbf{r})_{wat}$ (**a**), and for the protein $\rho^1(\mathbf{r})_{prot}$ alone (**b**). The grid mesh is 1 Å and the singlet densities $\rho^1(\mathbf{r})$ are computed as partial volume occupancies of the water and protein atoms at the different locations $\mathbf{r}_{i,j,k}$ in space using core diameters of 2.8 Å for the water molecules, and, respectively 1.0, 1.6, 1.8, and 1.9 Å for the H, N, O, C atoms of the protein. The density map results from a 100-psec time average of the protein and water trajectory. The contour for a partial volume occupancy $\rho^1(\mathbf{r}) = 0.4$ is displayed in both (a) and (b). The stick figure represents the mean position of the protein atoms.

eraged with its immediate neighbors contained within a spherical shell of 1 Å radius leading to a new set of less numerous points. This last operation is repeated with 0.1 Å incremental increases of the sphere radius until the maxima locations converge (Figure 1c).

RESULTS

Examples of $\rho^1_{tot}(\mathbf{r})$ and $\rho^1_{prot}(\mathbf{r})$ density maps resulting from the myoglobin MD simulation are presented in Figure 2. Notice how the solvent and the protein densities mesh together to form a continuous contour near the protein–solvent interface. Distinction between the respective contributions of both protein and water would not always be clear without knowing the decomposition of ρ_{tot} from ρ_{prot} from the simulation time averaged cartesians coordinates. The separation of the protein from its solvent surrounding is inherent to simulation methods whereas it is not as simply achieved in diffraction experiment analysis. The rest of the study will deal essentially with the solvent density $\rho^1_{wat}(\mathbf{r})$ which is the focus of our interest.

Solvent Organization at the Protein Interface

We started our analysis by inspection of the solvent density map resulting from the last 100 psec average of the simulation. This corresponds to the [70–170] psec interval of time for which equilibrium of the protein with the solvent had been achieved. The 100 psec average resulting from 2,000 sampled configurations proved to be sufficient to allow characteristic hydration patterns to be recognized. The protein appeared completely coated by an extensive network of high density hydration sites with density values higher than $1.35\rho^0$ where $\rho^0 = 0.295$ is the volume fraction occupancy of the solvent spatially averaged over the simulation box, the volume of the protein being approximately excluded,

$$\rho^0 = \langle \rho^1_{wat}(\mathbf{r}_{i,j,k}) \rangle_{i,j,k}.$$

Volume fraction occupancies are approximately equivalent to occupancy weights which are variable

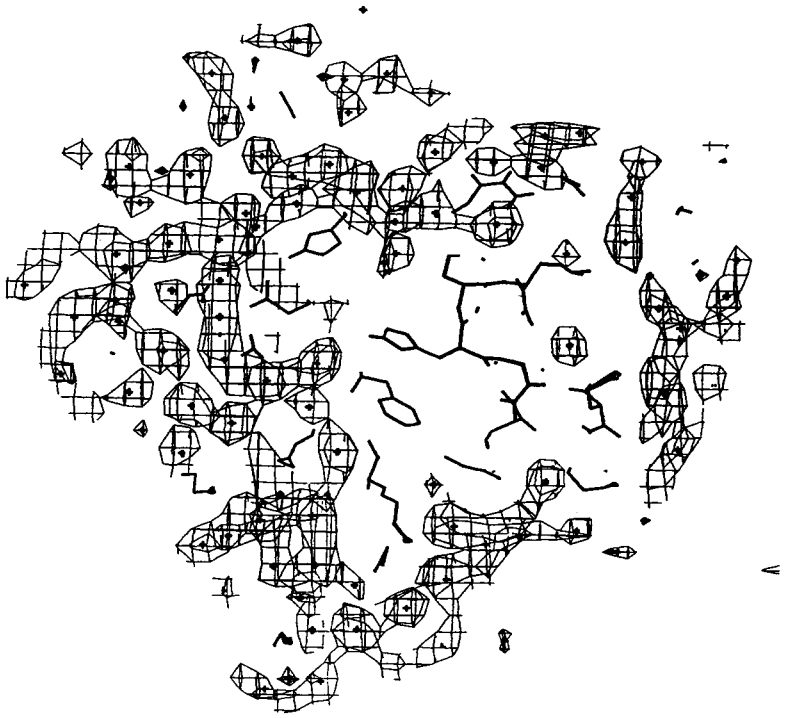


Fig. 3. A partial view of the hydration cluster around myoglobin for the $\rho^1_{wat}(\mathbf{r}) = 0.4$ contour level. Crosses represent mean positions of the hydration locations (density maxima) as determined by the procedure described in the text and shown in Figure 1.

parameters in diffraction experiment analysis. The network of hydration, as exemplified by Figure 3, appears as an extremely complex cluster of interconnected and disjoint sites. Thus, the solvent region delimited by the interior of this contour has an occupancy weight greater than 40%. Some sites are recognizable by their peanut shell like bulges embedded in the tubular networks of density. However, at this level of density, the protein hydration network cannot be uniquely interpreted according to the classical crystallographic picture of sites as well localized entities (presumably individual molecules) with separated ellipsoidal shapes. Only few of them could be well fit by a set of cartesian coordinates augmented by a temperature factor and an occupancy weight. Nevertheless, as the level of occupancy of the contour is increased from $1.35 \times \rho^0$ to $1.65 \times \rho^0$, or equivalently from 40 to 50%, the general shape of the network evolves towards fewer more spatially resolved sites (Figs. 4 and 5) and essentially approaches, as a limit, the crystallographic hydration picture of proteins. However, even at higher density the shape of some hydration sites substantially deviates from ellipsoids which degrades their ability to be fit using the usual Debye–Waller tensor (see Fig. 5).

Network Location

The procedure we used to determine the loci of the hydration sites as described in the methods section does not favor any a priori location around myoglobin such as biasing toward hydrophillic sites etc. The values of ρ_{min} and the radius of search were respectively set at $1.35 \times \rho^0$ and 1.8 Å. The latter value was empirically determined as a compromise between local repartitioning of the sites and the geometry of the density contour, as depicted in Figure 1. The resulting set of points accounted for the coordinates of 551 local maxima of the solvent region with 40% or higher occupancy weight. Since 1.8 Å is smaller than the diameter of the electronic repulsive core of water, these 551 sites do not represent 551 distinct water molecules but 551 locations that have been partially populated by water molecules during the course of the simulation. Interestingly, the number of hydration sites reduces to only 360 when the separation distance is forced up to 2.8 Å which is the effective core diameter of SPC water.[41] This number can be considered as a bound on the average instantaneous number of water molecules participating in the network of hydration. A similar observation was

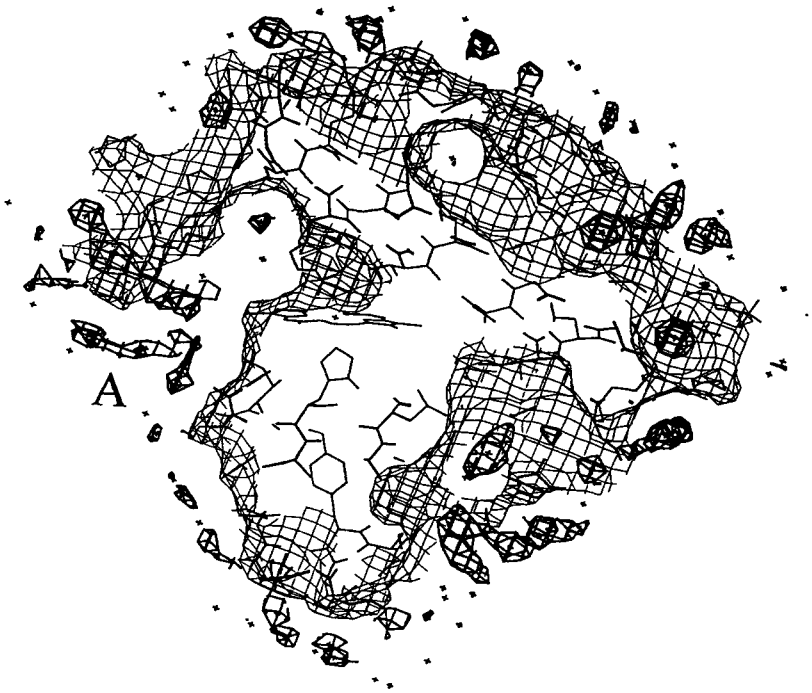


Fig. 4. Partial view of the solvent accessed surface (thin contour) and the hydration network for the $\rho^1(\mathbf{r})$ = 0.6 contour level (thick contour). The solvent accessed surface is defined by the contour for $\rho^1(\mathbf{r})$ = 0.01. It reveals the complexity of the protein water interface at low probability levels and outlines the porphyrin ring above the heme pocket with its entrance channel (region A) showing clusters of hydration. Higher density hydration sites are encapsulated in grooves and near charged groups at the protein surface.

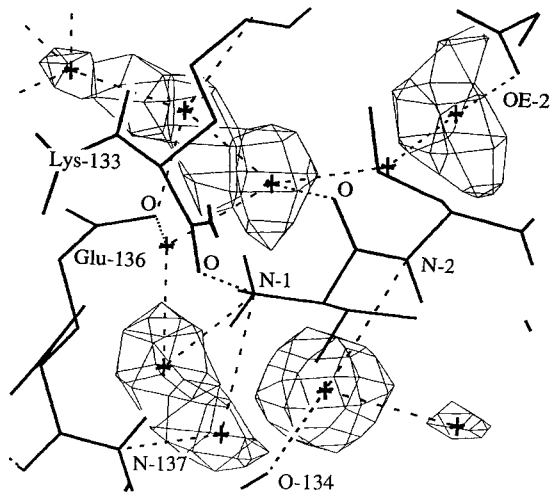


Fig. 5. Local detail of the hydration network near the N terminal residue Val-1 for the contour level $\rho^1(\mathbf{r})$ = 0.6. The protein average structure is represented by sticks and the H-bonds by dashes.

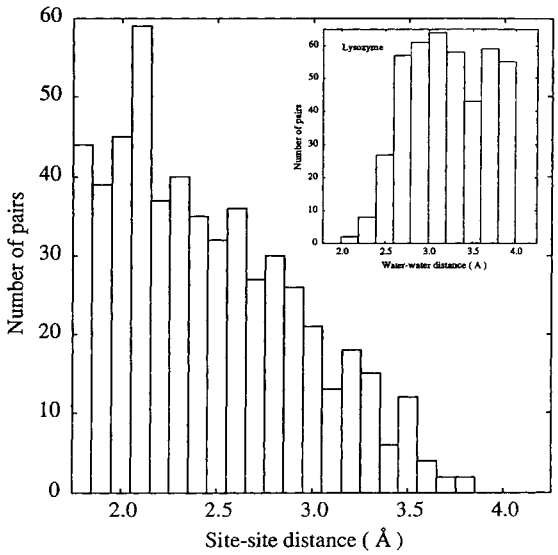


Fig. 6. The distance distribution for pairs of closest neighbor hydration sites in the myoglobin hydration network. The inset figure similarly describes the distance distribution of closest water molecules resulting from an X-ray study of human lysozyme.

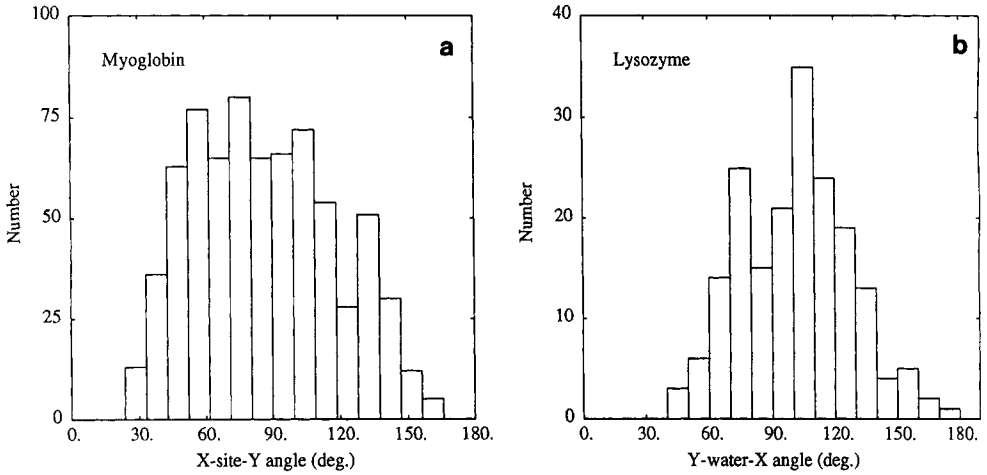


Fig. 7. A comparison of the closest neighbor angular distribution for the 551 hydration sites around myoglobin (**a**), and for the 360 water molecules refined in the human lysozyme study (**b**).

made in an early Monte Carlo simulation of the hen egg white lysozyme (HEWL) crystal where more maxima were observed for the solvent density map than the simulation contained water molecules.[42]

Global Properties of the Hydration Layer

One fundamental characteristic of the hydration network is its remarkable extent around the protein surface as depicted in the section shown in Figure 3. The histogram of Figure 6 shows there is virtually no pair of closest neighbors being separated by more than 3.5 Å which is a reasonable upper limit for hydrogen bonding. Another intriguing aspect of the distribution is the peak at 2.1 Å which indicates that the probability of finding two sites at this short separation distance is higher than for other distances. This is an effect which confirms our previous assertion that hydration sites and water molecules of hydration can be considered as distinct concepts when dealing with averaged solvation structures. Moreover, the rest of the distribution exhibits a sharp decrease in the number of closest neighbors as the separation distance increases with no noticeable enhancement around 3.0 Å separation. This is contrary to the expectation that the H-bond equilibrium distance should be prominent but it is consistent with our previous argument of the difference between the instantaneous versus the averaged (static) picture of protein hydration.

An interesting comparison can be made with a crystallographic study of human lysozyme (HL) at 1.5 Å resolution, which is also a globular protein of a size comparable to myoglobin.[43,44] The mass and volume of the orthorhombic crystal indicated about 350 water molecules per HL molecule which is small for such a protein crystal. Consequently, it was feasible to include 350 water molecules in the refinement. In such a study, the effect we mentioned would be difficult to detect since the crystallographic refinement procedures used built water molecules no closer than 2.8 Å. On the contrary, they observed a peak near 3. Å (see Fig. 6 inset). In our case, 70% of the pairs are found at a separation distance smaller than 2.8 Å. This suggests that the time-averaged hydration network cannot be completely understood as a simple arrangement of hydrogen bounded clusters of water molecules themselves "connected" to the protein polar and charged groups. The details of solvent fluctuations and protein motions are at the origin of this observed effect and will be addressed below and in the following paper.

The near neighbor angles formed by the hydration sites in our myoglobin simulation are compared with those for water oxygens in HL in Figure 7a. In the HL study the distribution of distances between closest ordered water molecules was found consistent with the presence of approximately tetrahedral (100° to 110°) and triangular (70° to 80°) environments with a dominant probability for tetrahedral spatial arrangements between ordered water molecules as shown in Figure 7b. A small but still distinguishable peak at about 150° in the previous diagram can be noticed. For myoglobin the angular distribution of sites separated by distances between 2.6 and 3.4 Å suggests also that angles around 60°, 80°, and 110° may be favored although the distribution of Figure 7a is broad and not strongly structured. However, we see the increased probability around 140° which is also detected, as mentioned above, in the lysozyme hydration water angular distribution. As it is shown in the next section, angles in the 60° to 80°, 80° to 100°, 100° to 110°, and 120° to 150° intervals can be assigned to triangular, square planar, pentagonal, and higher degree polygonal geometries.

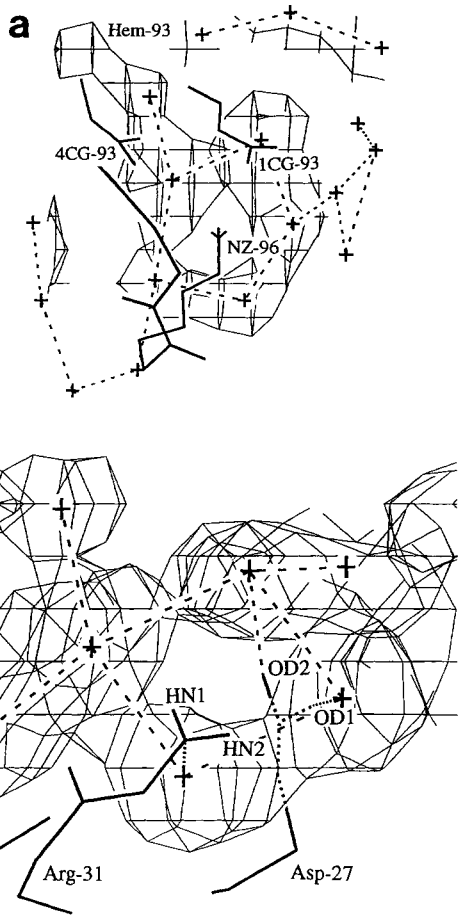


Fig. 8. A pentagonal ring of water density around the terminal atom of the side chain of the hydrophilic residue Lys-96 (**a**). The $\rho'(\mathbf{r}) = 0.4$ contour level is drawn; the dashes connect the hydration sites separated by distances between 2.7 and 3.5 Å. A square arrangement of water density around Arg-31 is shown in (**b**).

Polygonal Analysis of the Hydration Network

The local organization of the hydration cluster around the protein and more particularly near hydrophobic groups has been probed by recent X-ray experiments. The apparent formation of nonplanar pentagonal rings of hydrogen-bound water around hydrophobic residues was observed in the crystal structure of crambin, insulin, and cytochrome *c*.[45,46] Such distorted pentagons and hexagons of hydrogen-bound water were reported earlier for crystal structures at better than 1 Å resolution for α-cyclodextrin and dCpG-proflavine.[47,48] Pentagons appear to predominate over the hypothetical "ice-like" hexagons that were once assumed by earlier theoretical investigations.[49,50] However, pentagonal clathrate-like clusters are not uniformly observed near all such groups.[51–53]

The presence of such pentagons of hydration near hydrophilic groups is presented in Figure 8a. This figure shows the water structure in the vicinity of the hydrophilic Lys-96 side chain, a residue adjacent to the heme pocket entrance. The observed pentagonal ring of density is itself attached to the rest of the hydration network via at least three links to neighboring sites. The edges of the water pentagon have an average length of 2.8 Å and form angles with each other that are between 85° and 130°. Although clearly resolved by our approach, this pentagonal structure has an equilibrium population less than 50% and is consequently near the X-ray lower limit of detection for conservative water building. Such pentagons are also observed for Lys-47 and Lys-98.

Examination of the relative spatial distribution of hydration sites shows that many clusters accommodate the pentagonal geometry. We found that polygons of both lower and higher degrees can be seen in the hydration network. Figure 8b depicts the particular case where water molecules occupy four sites in a square arrangement. They participate simultaneously in the hydration of the charged groups of both Arg-31 and Asp-27. The average equilibrium distance and angle of the square are respectively 3.2 Å and 90°. Figure 9 depicts the hydration geometry at the interface between a group of residues (Glu-105, Phe-106, Ser-108, Ile-112, His-113, Asn-132, Glu-136, Arg-139) and disordered water, while Figure 10 shows the hydration pattern near a close pair of hydrophobic residues (Val-21 and Val-66). Although the angles formed by the pentagon vertices may individually vary between extreme values of 70° and 130°, five-membered rings of density can be capped by an additional site thus creating five triangles sharing vertices with the pentagonal clathrate like structure, cf. Figure 9. Although often incomplete without lowering the minimum density contour, larger structures and rings of density are observed. The hydration site positions for those density rings are found with vertex angles varying around average values of 113°, 123°, and 132°, respectively, for hexagons, heptagons, and octagons. Thus, the angular distribution continuum discussed in the previous section is a reflection of a broad class of geometric structures ranging from triangles to polygons of higher degree. Such water structures were also identified in Monte Carlo simulation studies of small peptides hydration, and in many MD simulations of pure liquid water.[54–58]

Intranetwork and Protein–Network Interactions

Tables I and II present selected statistics concerning the number of potential hydrogen bonds which characterize the hydration network and its pair distribution with the protein. If we consider that potential H-bonds exist between hydration sites separated by no more than 3.3 Å and no less than 2.7 Å, we found that 389 sites (which is 70%) of the total share between one and five H-bonds with their neighbors

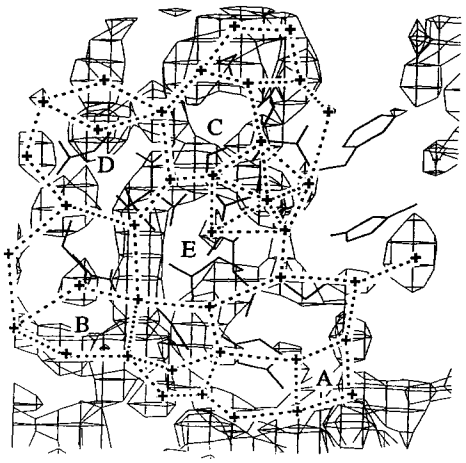


Fig. 9. Detail of the of the hydration geometry in the vicinity of Glu-105, Phe-106, Ser-108, Ile-112, His-113, Asn-132, Glu-136, Arg-139 [contour level $\rho^1(\mathbf{r}) = 0.4$]. Tetragons, pentagons, hexagons, heptagons, and octagons in the connected hydration cluster are identified by dashes.

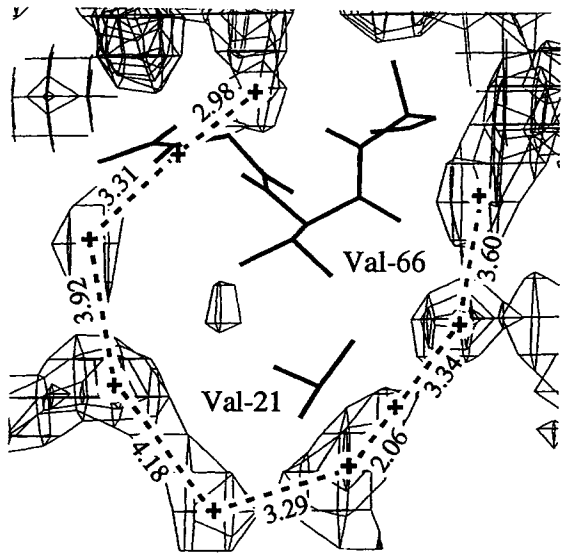


Fig. 10. An example of an extended ring of hydration near a pair of hydrophobic residues Val-21 and Val-66 [contour $\rho^1(\mathbf{r}) = 0.4$].

(Table I). The overall majority of sites have one neighbor in the H-bonding distance range and a total of about 700 different H-bonds accounts for the network itself. Notice that the number of sites decreases by nearly a factor of two for each additional neighbor or potential shared H-bond. However, there is little difference found between sites having zero to five potential hydrogen-bonded neighbors in regards to their average occupancy weight and B-factor which remain remarkably constant around 50% and 30 Å^2, respectively ($B \equiv 8\pi^2/3 \langle \Delta r^2 \rangle$, see following paper for additional information). Thus, there is no simple correlation between the potential H-bond number of a site and either its occupancy or temperature factor. This may be a reflection of the fact that hydration sites are not simply related to individual molecule positions.

A strikingly different effect is observed when examining the interaction between the hydration network and the protein. Indeed the number of H-bonds that a hydration site shares with the protein is correlated, on the average, with its occupancy and temperature factors. As the number of shared H-bonds increases from zero to six the average occupancy increases by a factor of two from 45 to 94% and the B-factor decreases by a factor of two from 31 to 16 Å^2. This expected but nonetheless remarkable effect confirms that the local structure and dynamics of the hydration network is dictated by its degree of interaction with the protein.

TABLE I. Site–Site Distribution Analysis

N*	$2.7 \le d^{\dagger} \le 3.3$ Å	$\langle W(\%) \rangle^{\ddagger}$	$\langle B_{\text{factor}} \rangle^{\S}$
0	162 (29%)	50.0	30.5
1	200 (36%)	51.5	30.5
2	114 (20%)	52.9	30.1
3	55 (10%)	51.7	29.8
4	17 (3%)	52.1	30.4
5	3 (<1%)	55.6	28.3
(Total number of H-bonds: 710)**			

*Number of neighbors within hydrogen bonding distance, i.e., between 2.7 and 3.3 Å from a given hydration site.
†Total number of sites for each case ($N = 1,2,\ldots,5$) and relative percentages.
‡Average occupancy weight.
§Average isotropic factor computed as the average isotropic mean square displacement: $8/3\pi^2 \langle (\Delta \mathbf{r})^2 \rangle$.
**Total number of sites weight averaged with their respective number of neighbors within H-bond distance.

TABLE II. Water–Protein Interaction Analysis

H-bonds*	N_{Site}	$\langle dis \rangle^{\dagger}$	$\langle W(\%) \rangle^{\ddagger}$	$\langle B_{\text{fac}} \rangle^{\S}$
0	209	4.50	44.8	31.5
1	225	3.29	47.2	30.4
2	90	3.33	51.9	29.1
3	18	3.36	60.6	26.8
4	7	3.39	65.8	22.8
5	1	3.32	87.1	20.1
6	1	3.40	94.0	15.7
(Total number of H-bonds: 508)**				

*Number of neighbors within hydrogen bonding distance, i.e., at less than 3.6 Å from a given hydration site.
†Average distance.
‡Average occupancy weight.
§Average temperature factor.
**Total number of sites weight averaged with their respective number of H-bonds to the protein atomic sites.

Outer Hydration Layers Around Myoglobin

Although additional layers of hydration, or solvation shells characterized by water with reduced mobility and higher occupancy, have not been routinely observed by microscopic methods such as X-ray and neutron crystallography their existence is predictable via simulation. They are expected to bridge between strongly ordered water regions in close vicinity to the protein and disordered bulk. Their effective interactions with the protein occurs primarily indirectly via correlations with water molecules belonging to the first layer of solvation. Thermodynamic measurements indicate that water in layers outside the first solvation layer are perturbed an order of magnitude less strongly than water in contact with the protein.[28] However, as demonstrated by dynamic measurements such as ESR, NMR, and RSMR some motional properties are significantly altered above hydration levels for which the principal changes in thermodynamic properties, such as heat capacity, are completed.[28,29,59–64]

A recent neutron diffraction study of carboxymyoglobin crystals was reported with an overall error factor $R = 11.5\%$ in part as a consequence of including in the final refinement procedure a solvent description consisting in several partial layers characterized by different occupancy weight and liquidity.[27] As depicted by the histogram of Figure 11 where the number of sites versus distance to the closest protein atom is plotted, our study confirms that in addition to the primary hydration shell near 3.1 Å average distance from the protein surface there also exist discernible layers of hydration at distances extending up to 6.8 outward from the protein. These layers, although considerably more diffuse than the primary one, are nonetheless clearly recognizable and characterized by water with higher density than bulk (Fig. 12). They represent about 180 of the total 551 hydration sites and therefore constitute a nonnegligible part of the hydration network. Conversely, this is quite a small number of hydration sites when compared with the amount of solvent required to ensure a complete coverage of the protein by second and third solvation layer water molecules. However, the presence of those hydration sites for distances substantially beyond 3.4 Å from the protein surface brings additional evidence of relatively disordered water regions characterized by nonbulk properties as suggested by the experimental techniques mentioned above. The existence of multiple layers of hydration is not suggested in the results from the human lysozyme study which exhibit a remarkable similarity to myoglobin concerning the principal features of the distribution of sites, although it does not indicate the presence of water beyond 4.6 Å (cf. Fig. 11). This similarity is indicative that the solvent organization in the vicinity of globular proteins is essentially universal from one protein to another as well as from solution to crystalline state (ignoring protein crystal contacts).

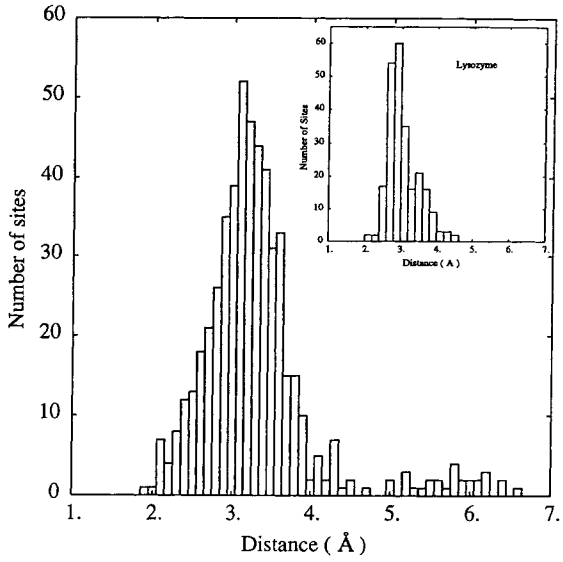


Fig. 11. Distribution of the 551 hydration sites as a function of the distance to the closest nonhydrogenic protein atom; the inset is the equivalent distribution for the 360 water molecules refined in the HL study.

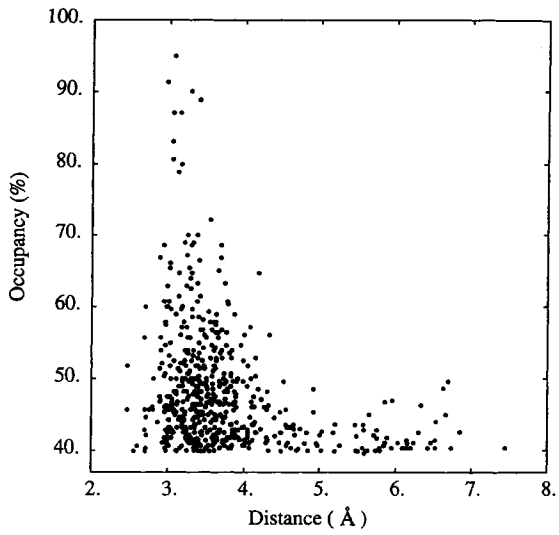
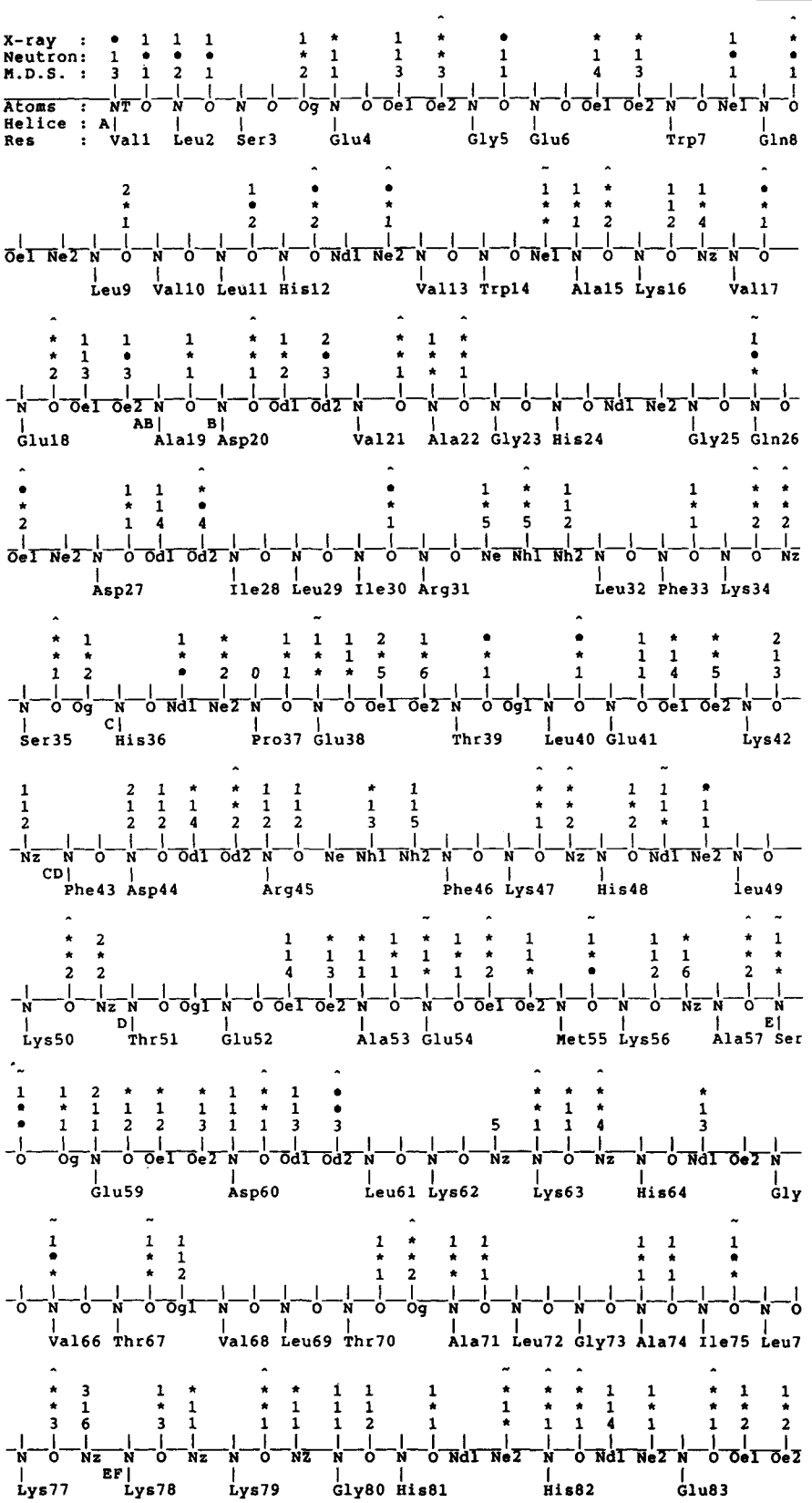


Fig. 12. Occupancy weight distribution for the 551 hydration sites as a function of the closest distance to a nonhydrogenic protein atom.

COMPARISON BETWEEN X-RAY, NEUTRON, AND SIMULATION

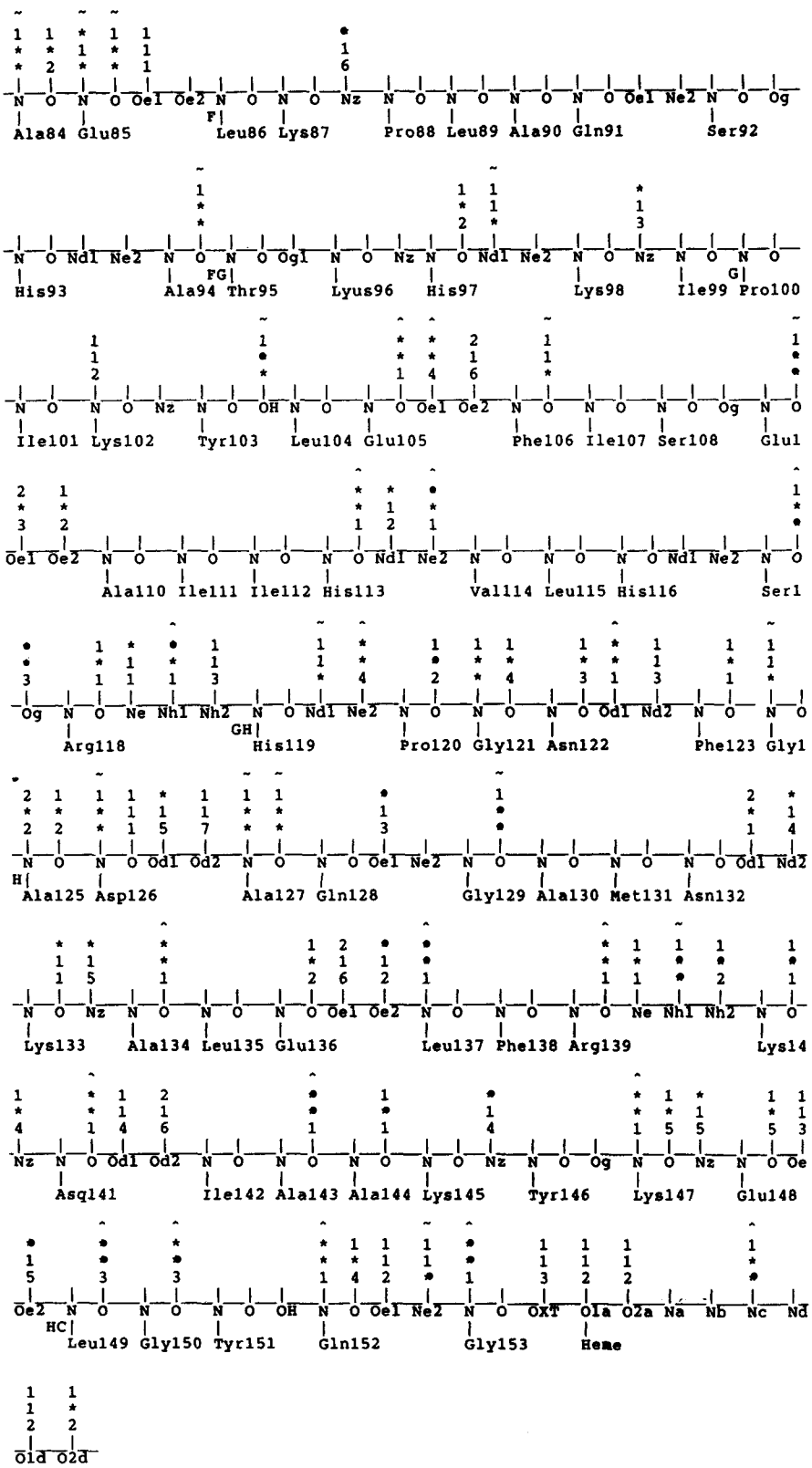
As mentioned above, several X-ray refinements of myoglobin (with water molecules of hydration) have recently appeared. We have chosen to make our

TABLE III. Myoglobin Hydration Sites: Comparison MDS With X-Ray and Neutron Data†



(continued on overleaf)

TABLE III. Myoglobin Hydration Sites: Comparison MDS With X-Ray and Neutron Data[†] (Continued)



[†]Digits indicate the number of hydration waters or hydration sites observed in the vicinity of the protein sites. The asterisks (*) indicate the absence of observed hydration waters whenever hydration waters are observed in either one or both other cases. The carats (^) indicate when MDS hydration sites are found near protein sites whereas no X-ray or neutron waters are observed. The *tilde* (~) indicate when no *MDS* waters are found where either X-ray or neutron waters are observed. The distance criterion for MDS hydration sites is 3.6 Å.

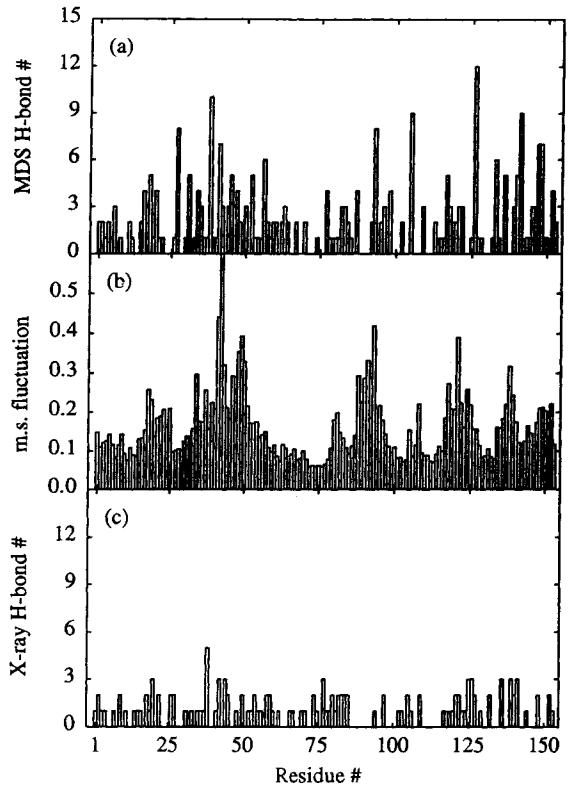


Fig. 13. Comparison of the number of contacts or potential H-bonds between each residue and the hydration network (**a**), and the mean fluctuation of the C_α, N, O backbone atoms for each residue (**b**). A potential H-bond is counted for each hydration site within 3.4 Å from a polar or charged group. The number of contacts per residue for distances < 3.4Å is plotted for the 127 water molecules determined in a recent X-ray crystallographic study of wild-type myoglobin (**c**).

most detailed comparison with one[20] for simplicity. The results should be essentially unaffected by using other structures of similar quality.

A detailed comparison between the hydration sites around myoglobin as determined by the present simulation study with those obtained from recent X-ray[20] and neutron diffraction[27] studies is summarized in Table III. In general the locations of the hydration sites relative to the protein obtained from simulation are nearly a superset of the refined water positions from both X-ray and neutron data. Indeed, with rare exception whenever the two experimental techniques differ the MD results show the presence of one or more hydration sites. The majority of the polar or charged side chains are found to be solvated by continuous hydration density with clusters corresponding to geometries previously described, whereas X-ray and neutron experiments usually refine single water molecules. In addition, many of the surface protein backbone oxygen atoms are found to be in the H-bond range of some hydration sites which were not observed in either X-ray or neutron data. When analyzing the number of contacts within 3.4 Å from the protein surface for both X-ray hydration water and MD simulation hydration sites, we found that the number of simulation contacts per residue is correlated to the backbone mean square fluctuation (see Fig. 13a and b), whereas this effect cannot be seen in X-ray results (Fig. 13c).

X-ray experiments on liquid water measure electron density so that as far as measured intensity is concerned the oxygen positions (density) contribute most to the data. Neutron experiments give more weight to the water hydrogens than the oxygens. Localization of water density around charged groups may appear with different weightings from these experiments depending on the extent of hydrogen vs. oxygen positional and orientational restriction due to interaction (+ vs. −). Thus it is reasonable that the simulation finds a collection of sites that includes the refined solvent positions from both experiments.

Careful examination of both MD and X-ray hydration site occupancy weights shows, as a general trend, that X-ray sites have a larger refined occupancy factor than the simulation results. Indeed, 90% of the X-ray higher density site loci have occupancy factors greater than 80% whereas 90% of the MD sites are found with occupancies smaller than 60% (see Fig. 12). This effect may be partly attributed to the delocalization of disordered water molecules over several hydration sites as the protein evolves from one conformational substate to another during the course of the simulation.[65] Refining molecules of water rather than the probability density in the experimental data contributes to this.

Another noticeable difference concerning the contact distances between the hydration water and the protein polar or charged groups exists between X-ray and MD results. The simulated hydration sites which are potentially interacting with the protein are found at an average distance of 3.3 Å from the time averaged atomic coordinates of the hydrophilic groups (cf. Table II), whereas the refined water molecules of X-ray are localized at about 3.0 Å. This is one more aspect of the averaged dynamic effect resulting from the flexibility of the protein side chains which share multiple but not always simultaneous H-bonds with the hydration network.

CONCLUSION

The dynamic computer simulation approach is informative when addressing the general problem of protein hydration. This study confirms that subnanosecond trajectories are suitable to study many hydration features that can be correlated with experiments.

The extent of the hydration layer around myoglobin as depicted by our approach is corroborated by information provided by a wide range of experimental techniques including thermodynamic, dynamic,

diffraction, and spectroscopic measurements. A particularly good agreement exists between sorption thermodynamics and microwave absorption results obtained on myoglobin powder and the present study concerning the number of water molecules (less than 400) which exhibit a sharply reduced mobility as compared with bulk. Heat capacity and dielectric response are well correlated to the degree of interaction of water with the protein. Such solvent molecules are characterized by reduced rotational and translational motions.

The direct characterization of a multilayer solvation structure demonstrates that distinct hydration layers represented as a probability distribution function could be implemented in refinement procedures of diffraction experiments. Practical considerations dictate that such a distribution should be parameterized with a relatively small number of fitting variables.[23]

Polygonal patterns resulting from the tetrahedral structure of water are clearly recognized as the underlying structural characteristic of partially disordered water around both polar and apolar groups. As a result a completely connected cluster of hydration with a singlet density larger than bulk is seen at values which are essentially invisible to diffraction experiments. In addition, as indicated by the similarity between lysozyme and myoglobin, such patterns might be a universal feature of globular proteins hydration.

Experimental evidence of the existence of an extended hydrogen-bonded network of water molecules around globular proteins is found in proton conduction measurements interpreted with a percolation model.[66–68] The percolation process can be understood as proton transfer along thread of hydrogen-bonded water molecules themselves linked to the protein surface. Each water molecule is a proton conducting element and tubular aggregates of water are thus proton conducting "wires." When the water cluster around the protein is completely connected protons can bypass the topographical details of the protein surface and circulate. The nonconducting to conducting phase transition in proton conduction occurs very sharply for a hydration degree where there is no discontinuity in the thermal properties. The bulk hydration picture proposed in this work is in general agreement with this type of picture from experiment.

Given a consistent picture of the structure of water around globular protein, a reasonable goal would be to define a set of simple rules to reconstruct the network of hydration around globular proteins from the knowledge of their three-dimensional average structure and some elements of their dynamics. Such information could be used to aid in crystallographic refinements. Work along such lines is in progress.[69] The hydration sites here were anticipated in the work of Eisenberg some time ago.[70] Here we attempted to quantify and refine the concept via simulation.

ACKNOWLEDGMENTS

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

REFERENCES

1. Hendrickson, W.A., Konnert, J.H., "Computing in Crystallography." Diamond, R., ed. Bangalore: Indian Academy of Science: 1980:1–25.
2. Blundell, T.L., and Johnson, L.N. "Protein Crystallography." London: Academic Press, 1976.
3. Kendrew, J.C., Dickerson, R.E., Strandberg, B.E., Hart, R.G., Davies, D.R. Structure of myoglobin, a three-dimensional Fourier synthesis at 2 Å. resolution. Nature (London) 185:422–431, 1960.
4. Kendrew, J.C., Watson, H.C., Strandberg, B.E., Dickerson, R.E., Phillips, D.C., Shore, V.C. A partial determination by X-ray methods, and its correlation with chemical data. Nature (London) 190:666–672, 1961.
5. Kagen, L.J. "Myoglobin Biochemical, Physiological, and Clinical Aspects." New York: Columbia University Press, 1973.
6. Pauling, L. "Nature of the Iron-oxygen Bond in Oxyhaemoglobin," Nature, 203:182–183, 1964.
7. Rohlfs, R.J., Mathews, A.J., Carver, T.E., Olson, J.S., Springer, B.A., Egeberg, K.D., Sligar, S.G. The effects of amino acid substitution at position E7 (residue 64) on the kinetics of ligand binding to sperm whale myoglobin. J. Biol. Chem. 265:3168–3176, 1990.
8. Olson, J.S., Mathews, A.J., Rohlfs, R.J., Springer, B.A., Egeberg, K.D., Sligar, S.G., Tame, J., Renaud, J.P., Nagai, K. The role of distal histidine in myoglobin and haemoglobin. Nature (London) 336:265–266, 1988.
9. Johnson, K.A., Olson, J.S., Phillips, G.N. Jr. The structure of myoglobin-ethyl isocyanide: Histidine as a swinging door for ligand entry. J. Mol. Biol. 207:459–463, 1989.
10. Genberg, L., Richard, L., McLendon, G., Miller, R.J.D. Direct observation of global protein motion in hemoglobin and myoglobin on picosecond time scale. Science 251: 1051–1053, 1991.
11. Bone, S., and Pethig, R. Dielectric studies of protein hydration and hydration-induced flexibility. J. Mol. Biol. 181:323–326, 1985.
12. van Gunsteren, W.F., Karplus, M. Protein dynamics in solution and in a crystalline environment: A molecular dynamics study. Biochemistry 21:2259–2274, 1982.
13. Gilson, M.K., Honig, B. Calculation of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies, and conformational analysis. Proteins 4:7–18, 1988.
14. Rogers, N.K. The modeling of electrostatic interactions in the function of globular proteins. Prog. Biophys. Mol. Biol. 48:37–66, 1986.
15. Brooks, C.L. III, Karplus, M. Solvent effects on protein motion and protein effect on solvent motion. J. Mol. Biol. 208:159–181, 1989.
16. Brooks, C.L. III, Karplus, M., Pettitt, B.M., Proteins: A theoretical perspective, structure, and thermodynamics. In: "Advances in Chemical Physics," Vol. LXXI. New York: John Wiley & Sons, Inc., 1988.
17. Richards, L., Genberg, L., Deak, J., Chiu, H.L., Miller, R.J.D. Picosecond phase grating spectroscopy of hemoglobin and myoglobin: Energetics and dynamics of global protein motion. Biochemistry 31:10703–10715, 1992.
18. Savage, H., Wlodawer, A. Determination of water structure around biomolecules using X-ray and neutron diffraction methods. Methods Enzymol. 127:162–183, 1986.
19. Phillips, G.N., Jr., Arduini, M., Springer, B.A., Sligar, S.G. Crystal structure of myoglobin from a synthetic gene. Proteins 7:358–365, 1990.

20. Phillips, G.N. Myoglobin (Met). Protein Data Bank, Chemistry Department, Brookhaven National Laboratory.
21. Richards, F.M. Areas, volumes, packing, and protein structure. Annu. Rev. Biophys. Bioeng. 6:151–176, 1977.
22. Diesenhofer, J., Steigman, W. Crystallographic refinement of the structure of bovine pancreatic trypsin inhibitor. Acta Crystallogr. B29:238–250, 1975.
23. Badger, J., Caspar, D.L.D. Water structure in cubic insulin crystals. Proc. Natl. Acad. Sci. U.S.A. 88:622–626, 1991.
24. Wlodawer, A. Neutron diffraction crystallography. Prog. Biophys. Mol. Biol. 40:115–157, 1982.
25. Kossiakoff, A.A. Neutron protein crystallography: Advances in method and application. Annu. Rev. Biophys. Bioeng. 12:159–182, 1983.
26. Schoenborn, B.P. Solvent effect in protein crystals: A neutron diffraction analysis of solvent ion density. J. Mol. Biol. 201:741–749, 1988.
27. Cheng, X., Schoenborn, B.P. Hydration in protein crystals. A neutron diffraction analysis of carbonmonoxymyoglobin. Acta Crystallogr. B46:195–208, 1990.
28. Rupley, J.A., Careri, G. Protein hydration and function. Adv. Protein Chem. 41:37–172, 1991.
29. Parak, F. Correlation of protein dynamics with water mobility: Mössbauer spectroscopy and microwave absorption methods. Methods Enzymol. 127:196–206, 1986.
30. Goldanskii, V.I., Krupyanskii, Y.F. Protein and protein-bound water dynamics studied by Raleigh scattering of Mössbauer radiation (RSMR). Quart. Rev. Biophys. 22:39–92, 1989.
31. Hagler, A.T., Moult, J. Computer simulation of the solvent structure around biological macromolecules. Nature (London) 272:222–226, 1978.
32. van Gunsteren, W.F., Berendsen, H.J.C., Hermans, J., Hol, W.G.J., Postma, J.P.M. Computer simulation of the dynamics of hydrated protein crystals and its comparison with x-ray data. Proc. Natl. Acad. Sci. U.S.A. 80:4315–4319, 1983.
33. Levitt, M., Sharon, R. Accurate simulation of protein dynamics in solution. Proc. Natl. Acad. Sci. U.S.A. 85:7557–7561, 1988.
34. Thanki, N., Thornton, J.M., Goodfellow, J.M. Distribution of water around amino acid residues in proteins. J. Mol. Biol. 202:637–657, 1988.
35. Kuriyan, J., Petsko, G., Levy, R.M., Karplus, M. Effect of anisotropy and anharmonicity on protein crystallographic refinement. An evaluation by molecular dynamics. J. Mol. Biol. 190:227–252, 1986.
36. Sternberg, M.J.E., Grace, D.E.P., Phillips, D.C. Dynamics information from protein crystallography. An analysis of temperature factors from refinement of the hen egg-white lysozyme structure. J. Mol. Biol. 130:231–251, 1979.
37. Brünger, A.T., Kuriyan, J.K., Karplus, M. Crystallographic R-factor refinement by molecular dynamics. Science 235:458–460, 1987.
38. Findsen, L.A., Subramaniam, S., Lounnas, V., Pettitt, B.M. Molecular dynamics simulation of metmyoglobin in aqueous solution. In: "Principles of Molecular Recognition." Buckingham, A.D., ed. London: Chapman and Hall. 1991.
39. Weiner, S.J., Kollman, P.A., Nguyen, D.T., Case, D.A. An all atom force field of simulations of proteins and nucleic acids. J. Comp. Chem. 7:230–252, 1986.
40. Lounnas, V., Pettitt, B.M., Findsen, L., Subramanian, S. A microscopic view of protein solvation. J. Phys. Chem. 96: 7157–7159, 1992.
41. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., Hermans, J. "In Intermolecular Forces." Pullman, ed. Dordrecht: Reidel, 1981.
42. Hermans, J., Vacatello, M. Modeling water-protein interactions in a protein crystal. ACS Symp. Ser. 127:199–214, 1980.
43. Blake, C.C.F., Pulford, W.C.A., Artymiuk, P.J. X-ray studies of water in crystals of lysozyme. J. Mol. Biol. 167:693–723, 1983.
44. Artymiuk, P.J., Blake, C.C.F. Refinement of human lysozyme at 1.5 Å resolution analysis of non-bonded and hydrogen-bond interactions. J. Mol. Biol. 152:737–762, 1981.
45. Kautzmann, W. Three dimensional structure of proteins. Adv. Prot. Chem. 13:1–63, 1959.
46. Klotz, I.M. Protein hydration and behavior. Science 128: 815–822, 1959.
47. Saenger, W. Circular hydrogens bonds. Nature (London) 279:343–344, 1979.
48. Neidle, S., Bermann, H., Shieh, H.S. Highly structured water network in crystal of a deoxydinucleoside-drug complex. Nature (London) 288:129–133, 1980.
49. Franks, F. "Water-A Complete Treatise." New York: Plenum Press, 1972.
50. Pethig, R. "Dielectric and Electronic Properties of Biological Material." New York: Wiley, 1979.
51. Teeter, M.M. Water-protein interactions: Theory and experiment. Annu. Rev. Biophys. Biophys. Chem. 20:577–600, 1991.
52. Teeter, M.M. Water structure of a hydrophobic protein at atomic resolution: Pentagon rings of water molecules in crystals of crambin. Proc. Natl. Acad. Sci. U.S.A. 81:6014–6018, 1984.
53. Baker, E.N., Blundell, T.L., Cufield, J.F., et al. The structure of 2-zinc pig insulin crystals at 1.5 Å resolution. Proc. Trans. R. Soc. London Ser. B319:369–456, 1988.
54. Goodfellow, J.M., Howell, P.L., Voyelle, F. Monte Carlo studies of water in crystal hydrates. Ann. N.Y. Acad. Sci. 482:179–193, 1988.
55. Howell, P.L., Goodfellow, J.M. Computer simulation of nucleotide crystal hydrates in solutions. Ann. N.Y. Acad. Sci. 482:195–197, 1988.
56. Corongiu, G., Clementi, E. Solvated water molecules and hydrogen-bridged networks in liquid water. J. Chem. Phys. 98:2241–2249, 1992.
57. Rahman, A., Stillinger, F.H. Hydrogen-bond patterns in liquid water. J. Am. Chem. Soc. 95:7943–7948, 1973.
58. Speedy, R.J., Madura, J.D., Jorgensen, W.L. Network topology in simulated water. J. Phys. Chem. 91:909–913, 1987.
59. Bryant, R.G. Magnetic resonance and macromolecule solvation dynamics. Stud. Phys. Theor. Chem. 38:683–705, 1988.
60. Piculell, L., Halle, B. Water spin relaxation in colloidal systems. Part 2. Water oxygen-17 and deuterium relaxation in protein solutions. J. Chem. Soc., Faraday Trans. 1 82:401–414, 1986.
61. Fullerton, G.D., Ord, V.A., Cameron, I.L. Frequency dependence of magnetic resonance spin-lattice relaxation of protons in biological materials. Biochem. Biophys. Acta 869:230–246, 1986.
62. 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.
63. Likhtenshtein, G.I. Study of biological membranes by means of the spin label method. Symp. Pap. Int. Biophys. Congr. 4:145–153, 1972.
64. 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.
65. Frauenfelder, H., Gratton, E. Protein dynamics and hydration. Methods Enzymol. 127:207–216, 1986.
66. Careri, G., Giansanti, A., Rupley, J.A. Proton percolation on hydrated lysozyme powders. Proc. Natl. Acad. Sci. U.S.A. 83:6810–6814, 1986.
67. Careri, G., Giansanti, A., Rupley, J.A. Critical exponents of protonic percolation in hydrated hydrogen powders. Phys. Rev. A 37:2703–2705, 1988.
68. Rupley, J.A., Siemankowski, L., Careri, G., Bruni, F. Two-dimensional protonic percolation on highly hydrated purple membrane. Proc. Natl. Acad. Sci. U.S.A. 85:9022–9025, 1988.
69. Lounnas, V., Pettitt, B.M. A global model of the protein–solvent interface. Biophys. J. (in press).
70. Scanlon, W.J., Eisenberg, D. Solvation of crystalline proteins. J. Mol. Biol. 98:485–502, 1975.