Eur Biophys J (1992) 21, 313–320

European Biophysics Journal

The hydration shell of myoglobin

F. Parak [1]*, H. Hartmann [1], M. Schmidt [1], G. Corongiu [2], and E. Clementi [2]

[1] J. Gutenberg-Universität, FB 21 Institut für Molekulare Biophysik, J. Welderweg 26, 6500 Mainz, Federal Republic of Germany
[2] CRS4, Centro Ricerca, Sviluppo e Studi Superiori in Sardegna, P.O. Box 488, 09100 Cagliari, Italia

Received July 3, 1992/Accepted July 24, 1992

Abstract. The space in the unit cell of a metmyoglobin crystal not occupied by myoglobin atoms was filled with water using Monte Carlo calculations. Independent calculations with different amounts of water have been performed. Structure factors were calculated using the water coordinates thus obtained and the known coordinates of the myoglobin atoms. A comparison with experimental structure factors showed that both the low and the high resolution regime could be well reproduced with 814 Monte Carlo water molecules per unit cell with a B-value of 50 Å^2. The Monte Carlo water molecules yield a smaller standard R-value (0.166) than using a homogeneous electron density for the simulation of the crystal water (R = 0.212). A reciprocal space refinement of the water and the protein coordinates has been performed. Monte Carlo calculations can be used to obtain information for crystallographically invisible parts of the unit cell and yield better coordinates for the visible part in the refinement.

Key words: Monte Carlo – Crystal water – Structure refinement

Introduction

The aim of molecular biophysics is the understanding of the functional properties of biologically important molecules. Especially in the case of proteins the X-ray structure determination is a significant step in this direction. However, the architecture of a molecule is not necessarily the only parameter which determines its function. Enzymatic activities or the binding of ligands also depend on the flexibility of proteins. In recent years, a considerable effort has been put into the investigations of protein dynamics. For a review see Frauenfelder et al. (1988). One interesting result obtained was that dry proteins become rigid and lose their functional activities (Rupley and Careri 1989a). There is a close interaction between the protein and the hydration shell. The protein determines the structures of the water up to a certain distance from the protein "surface" and the water influence the protein dynamics (Krupyanskii et al. 1990).

Information on the mobility of water molecules on protein surfaces stems mainly from NMR-investigations (Kuntz and Kauzmann 1974, Kotitschke et al. 1990, Mathur-De-Vre 1979, Otting and Wüthrich 1989, Otting et al. 1991) and dielectric measurements on protein water solutions (Bone and Pethig 1982, Careri et al. 1985, Dachwitz et al. 1989, Pissis 1989, Rupley and Careri 1989b, Singh et al. 1981). Molecular dynamic simulations have given information on the hydration shell (Ahlström et al. 1987, Finney et al. 1985, Goodfellow 1987, van Gunsteren et al. 1983, Steinbach et al. 1991) and made it possible to follow the path of a water molecule interacting with the protein surface (Clementi et al. 1990). Information on the structure of the water bound to a protein has been obtained from X-ray structure analysis (e.g. Blake et al. 1983, Phillips 1980) and neutron structure analysis (Cheng and Schoenborn 1990, Schoenborn 1988). For a detailed discussion see Jeffrey and Saenger (1991). However, without additional assumptions the only water molecules that can be located are those which occupy the same position with respect to the protein in most unit cells. In metmyoglobin crystals one can find about 160 waters per protein molecule (Hartmann et al. 1987). Such waters seen directly in a difference electron density map we call "crystallographically visible waters" in the following. The unit cell contains about three times more. Do these water molecules occupy completely different positions in different unit cells? In this case their contribution to the low order Bragg reflections can be calculated from an appropriate continuous electron density. In an alternative picture one can assume the existence of a mean position for each water molecule. In this case the widths of the distribution around these mean positions vary depending on the distance from the protein surface.

In this contribution, we try to get information on all water molecules in the unit cell of a metmyoglobin crys-

* *Correspondence to:* F. Parak

tal. To obtain discrete positions of crystallographically invisible water molecules we use Monte Carlo simulations. Different models for a water distribution in the protein crystal are assumed to calculate the contribution to the X-ray scattering. In this way, a comparison with experimental results from X-ray structure analysis becomes possible.

Experimental data

Myoglobin crystallizes in the monoclinic space group $P2_1$ with $a = 64.52$ Å, $b = 30.90$ Å, $c = 34.85$ Å and the monoclinic angle $\beta = 105.80°$C. The unit cell contains two molecules and has a volume of 66 854 ± 350 Å^3 (Frauenfelder et al. 1979, Hartmann et al. 1987, Takano 1977).

The structure of crystal water influences the intensity of the Bragg reflections, especially at low scattering angles. In order to test the structure models for the water in myoglobin crystals we compared the calculated structure factors, $|F_H^c|^2$, with the experimentally observed structure factors $|F_H^0|^2$. To obtain $|F_H^c|^2$-values a data set of a met-myoglobin crystal, collected at 300 K to a spatial resolution of 1.5 Å was used (Hartmann et al. 1987).

Coordinates and individual mean square displacements for all non-hydrogen atoms were refined iteratively. After 16 refinements we arrived at a standard R-value of 17.6%. The positions of 160 H_2O-molecules per myoglobin could be determined by assuming 100% occupancy but individual B-factors. Details were given by Hartmann et al. (1987). At this stage of the refinement we have the coordinates of all non-hydrogen atoms of myoglobin and of 160 O-atoms of the water molecules as well as the absolute $|F_H^c|^2$-values calculated from these coordinates. The observed structure factors, $|F_H^0|^2$, were scaled to the $|F_H^c|^2$-values during the refinement procedure and are thus also absolute values. These structure factors and coordinates were used as a reference in the following discussions.

Calculated $|F_H^c|^2$-values are always absolute structure factors, obtained during the refinement of a model structure. These values determine the absolute scaling of the $|F_H^0|^2$-values. In this way $|F_H^0|^2$-values vary slightly when compared with different theoretical models. This makes a discussion difficult. For this reason we generally use here the above described $|F_H^0|^2$-values. For a comparison, we then scale the calculated structure factors to this $|F_H^0|^2$ reference set.

In all fit procedures $|F_H^c|^2 - |F_H^0|^2$ was calculated for all $(h\,k\,l) \equiv$ H-values where experimental data were available.

For a graphic display of the results, however, the $|F_H^c|^2$ and the $|F_H^0|^2$-values, respectively, were averaged in spherical shells of radius $\sin\vartheta/\lambda$ and thickness $\Delta\sin\vartheta/\lambda$ (ϑ: Bragg angle; λ: wavelength of the X-rays). In this way we obtain $\langle|F_H^0|^2\rangle$ and $\langle|F_H^c|^2\rangle$ as a function of $\sin\vartheta/\lambda$.

As a seen in Fig. 1 the $|F_H^c|^2$ and the $|F_H^0|^2$ fit excellently at resolutions higher than 5.5 Å (corresponding to $\sin\vartheta/\lambda = 0.091$ Å^{-1}). This is trivial because the atomic coordinates were determined from this agreement. However, at low resolutions (small scattering angles) there

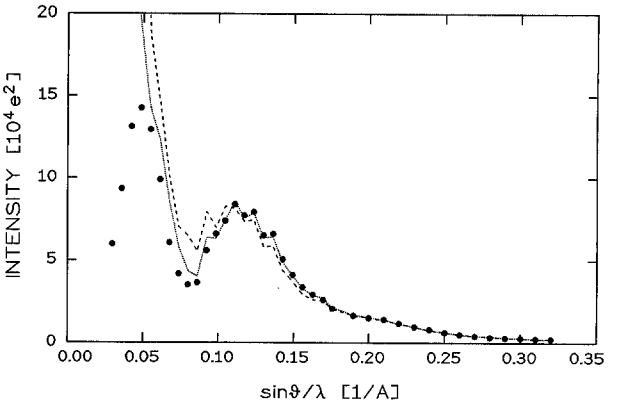

Fig. 1. $|F_H|^2$-values averaged in shells of thickness $\Delta\sin\vartheta/\lambda$ (between 0.0077 and 0.01) as function of $\sin\vartheta/\lambda$. Circles: $|F_H^0|^2$ from Hartmann et al. (1987); dotted line: $|F_H^c|^2$ calculated from coordinates of the non-hydrogen atoms of myoglobin and 160 water O-atoms/myoglobin; dashed line: $|F_H^c|^2$ as before but without the 160 waters

remains a strong discrepancy, indicating that the unit cell is not completely represented by two myoglobin and 320 H_2O molecules. That water molecules are missing is shown by the dashed curve in Fig. 1. Here, $|F_H^0|^2$-values were calculated from the myoglobin molecules only. The discrepancy with the observed structure factors is increased dramatically by omitting the 160 H_2O molecules/myoglobin.

Monte Carlo calculations of water positions in a myoglobin crystal

Possible coordinates of crystallographically invisible water molecules were obtained by Monte Carlo calculations. The Monte Carlo method in the Metropolis version (Metropolis et al. 1953) has been used to simulate an ensemble at $T = 298$ K with constant particle number. The whole unit cell has been chosen as the volume of our simulated system. Periodic boundary conditions were applied according to the crystal symmetry; no symmetry relation was imposed on the water molecules within the simulated volume. During the simulation only water molecules were allowed to move, whereas the protein atoms were kept at their crystallographic position. The coordinates of the atoms were taken from Hartmann et al. (1987). Hydrogen atoms were added according to geometrical considerations (Corongiu 1990). The bond lenghts, bond angles and dihedral angles were chosen from standard literature values depending upon the kind of atoms to which they are bound. We refer to Clementi et al. (1977) for the hydrogen positions in all aminoacide residues.

The interaction between protein and water is represented by atom – atom potential functions obtained by a fit to ab initio computations (Clementi et al. 1977). Each atom of the protein is assigned to a class; atomic number, partial charge and chemical environment are criteria used to define a class (Corongiu and Clementi 1978). Because no net charge was available for the heme group, an ab

initio self consistent field computation was performed and the net charges were determined with the Mulliken population analysis. Keeping in mind that the potentials are transferable for atoms belonging to the same class, we assigned each heme atom to a class according to the criteria previously summarized. Because of charged residues, we found a net charge of + 2 on the myoglobin protein, which under experimental conditions is neutralized by counterions. To simplify the work, instead of including counterions, we neutralized the excess charge by distributing a charge of − 2 over all myoglobin atoms; in this way each partial charge had to be diminished by 0.0008 electrons. Because no water interaction potentials were available for the Fe ion, we substituted the Fe ion with a Zn^{2+} ion. This is not a drastic approximation since we are interested in the structure of the water around the protein and not in reactions involving the heme group directly. The water-water interaction was described by the Monte Carlo potential of Matsuoka, Clementi and Yoshimine abbreviated MCY (Matsuoka et al. 1976).

To determine the number of water molecules needed to fill the simulation cell we proceeded in the following way. As a first step, we considered a system composed of the two myoglobin proteins and two times 160 water molecules, whose oxygen positions were obtained from the crystallographic refinement (Hartmann et al. 1987). The hydrogen atoms were added to each water oxygen by assuming that bond length and bond angle were as for a water molecule in the gas phase. The molecules had random orientations. Then, holding the center of mass of each water molecule fixed, the hydrogen orientation was optimized by performing a Monte Carlo run, using the interaction potentials described. The final configuration was used in the next step. As a second step, we started from an empty monoclinic cell and filled it with equilibrated bulk water (at a density of 0.998 g/cm^3 and $T = 298$ K). Then, the final coordinates from the previous step, for the two myoglobin proteins and the 320 water molecules, were inserted into the unit cell and those water molecules closer either to the protein atoms or to the crystallographic water molecules than a chosen distance were discarded. By assuming 1.6 Å as the minimum distance between the water oxygen atom and any atom (hydrogen included) of the protein and 2.3 Å between water oxygen atoms, 751 water molecules were left inside the cell. It has to be emphasized that these distances are not critical. They only make sure that one gets enough starting coordinates for water molecules. During the Monte Carlo runs the distances relax to final values. We then started a Monte Carlo simulation on this system and we discarded all those water molecules that did not move at all after 4 000 moves/molecule, either because of bad contacts or because of high repulsive interaction energies. After this preliminary run, 732 water molecules were left in the unit cell. A further 14 000 moves/water were performed and only the last 4 000 were considered to collect statistical properties. A graphical display of the last configuration showed the existence of "free space" within the cell, a clear indication that the density of our system was too low. In order to get more waters into the unit cell we repeated the same procedure but used 2% and 4% shorter distances to discard water molecules for the starting configuration. We obtained two additional systems: one containing 856 and another 966 water molecules.

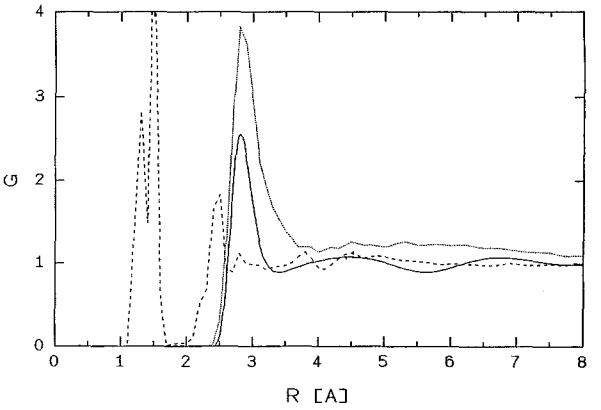

Fig. 2. Radial distribution functions. Solid line: oxygen-oxygen of bulk water; dotted line: oxygen-oxygen of Monte Carlo waters; dashed line: all non-hydrogen atoms of myoglobin and Monte Carlo waters

By monitoring the displacements of the water molecules during the equilibration phase we noticed that big rearrangements had taken place, with a high percentage of the water molecules displaced by more than 6 Å from the starting configuration. However, from the analysis of the "production stage" we found that the hydration patterns at the protein surface were very similar for the three simulations.

For the simulation containing 856 water molecules we continued the Monte Carlo run for an additional 15 M moves and we saved the water coordinates every 1 M moves. We call these configuration "snap shots".

Figure 2 shows the oxygen-oxygen correlation function for the 966 Monte Carlo water simulation and for bulk water (the simulation was performed at the same temperature and at a density of 0.998 g/cm^3). We would like to emphasize that the three simulations yield similar results. When compared with bulk water, an enhancement in the first peak is observed. This is an indication that we are dealing with highly structured water; indeed the majority of the water molecules are in the first hydration shell of the protein. No water molecule is found at a distance greater than 5 Å from the protein atoms because of the missing space. This explains the lack of a well defined second maximum in our pair correlation functions. For comparison the radial distribution function of all non-hydrogen atoms of the unit cell is shown. It is interesting to note that this radial distribution in the well defined myoglobin molecule, together with the Monto Carlo waters, also has a large number of small peaks for $R > 3$ Å. The well pronounced peaks at small R-values around 1.5 Å represent the covalent bond length within the molecule.

In order to get an impression of the structural differences of the crystal water, one can compare snap shots of the structures after a number of moves. The root mean square displacement of the water coordinates is shown in Fig. 3 for the three water sets.

It is interesting to note what happened with the 160 X-ray visible water molecules during the Monte Carlo moves. First we took a snap shot of the coordinates of all water molecules immediately after the Monte Carlo equilibration of the system and looked for those water molecules which were closest to the originally determined water coordinates. The dashed line in Fig. 4 gives the number of Monte Carlo waters at the distance R from the "X-ray" waters. Most of the new water positions are 0.6 to 2.0 Å away from the original positions. At the end of the Monte Carlo analysis the picture changes only slightly. We also looked through all snap shots to find the closest positions of Monte Carlo waters to the previous X-ray visible waters (Fig. 4, solid line). The distance is much smaller than in both other snap shots. A large number of Monte Carlo waters lie less than 0.6 Å away from the original X-ray water coordinates. In the snap shots the root mean square displacements of the Monte Carlo waters from the X-ray water positions is 0.77 Å.

We now look for the influence of the Monte Carlo waters on the structure factors. We started the following analysis with the 966 Monte Carlo water simulation. First we rejected all waters with coordinates of water at forbidden grid points. The definition of such a grid point is given in the next section. We then calculated $|F_H^c|^2$ values from our standard myoglobin structure, including individual $\langle x^2 \rangle$-values and the remaining 814 Monte Carlo waters (dashed line in Fig. 5). For the Monte Carlo waters we used $\langle x^2 \rangle = 0.125$ Å^2 values obtained by averaging the squared displacements of 10 snap shots during the Monte Carlo analysis. The agreement with the $|F_H^0|^2$-values at small scattering angles is now quite good (compare Fig. 1). However, at $\sin\vartheta/\lambda > 0.13$ Å^{-1} the discrepancy between $|F_H^0|^2$ and $|F_H^c|^2$ increases dramatically. Using well defined Monte Carlo water coordinates would imply that the total crystal water structure is identical in all unit cells. In this case the whole crystal water should be crystallographically visible. The $\langle x^2 \rangle$-values obtained from the differences of coordinates between the snap shots are unrealistically small. The number of moves was not large enough to simulate a sufficiently large number of water structures for the averaging procedure. In order to obtain an agreement with the high resolution data we gave a $\langle x^2 \rangle$-value of 0.63 Å^2 to all Monte Carlo waters. From $B = 8\pi^2 \langle x^2 \rangle$ this corresponds to a crystallographic B-factor of 50 Å^2. An optimization of the agreement between experimentally oberved and calculated structure factors was then performed by starting a new refinement, allowing the Monte Carlo waters and the myoglobin atoms to move and to change their $\langle x^2 \rangle$-values. Here we used the program PROLSQ (Konnert and Hendrickson 1980) which was modified to allow for periodic boundary conditions. After 24 cycles of refinement an excellent agreement between observed and calculated structure factors was obtained (see Fig. 5).

Continuous electron density – the simplest model for crystal water

For a comparison with the approach given in the last section we assume in the following that the coordinates of

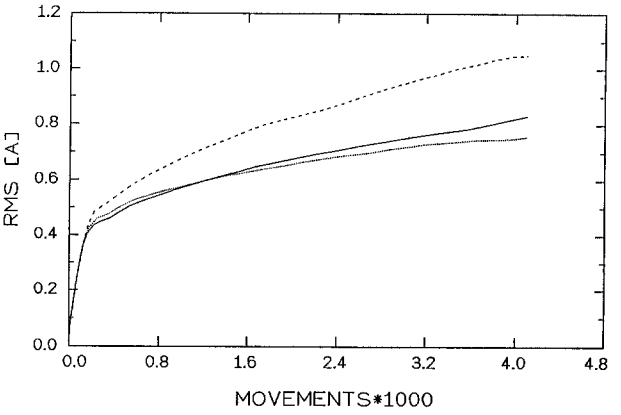

Fig. 3. Root mean square displacements of the water oxygens during the moves to collect statistical properties. Solid line: 966 waters; dotted line: 856 waters; dashed line: 732 waters

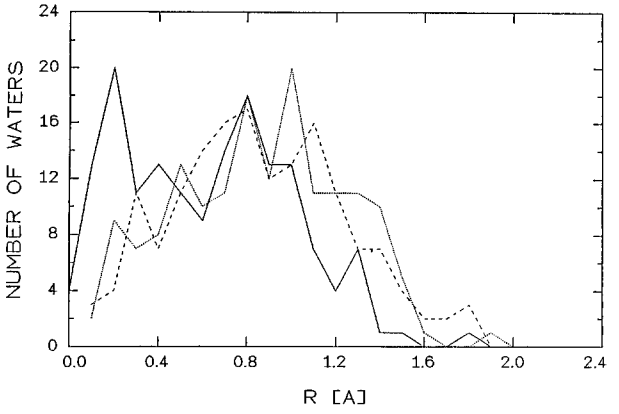

Fig. 4. Displacement of the 160 crystallographically visible water molecules during the Monte Carlo moves. Number of water molecules at a distance R [Å] from the crystallographic coordinates of the originally found water molecules after equilibration (dotted line) and at the end of the Monte Carlo analysis (dashed line). Solid line see the text

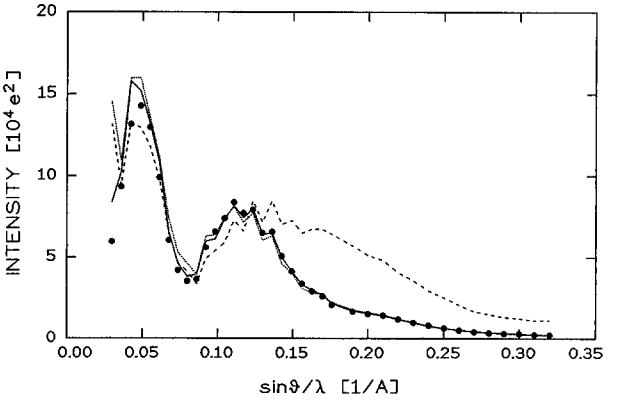

Fig. 5. Comparison of $|F_H^0|^2$ (circles) and $|F_H^c|^2$-values. The $|F_H^c|^2$ are calculated with three different models: dashed line: reference coordinates and $\langle x^2 \rangle$-values of myoglobin together with 814 Monte Carlo waters ($B = 9.86$ Å); dotted line: myoglobin as before, 814 Monte Carlo waters with $B = 50$ Å^2; solid line: After refinement of the Monte Carlo water positions and disorders

most of the crystallographically invisible waters are completely uncorrelated in different unit cells. The X-ray structure analysis averages over all unit cells of the crystal. The result is that these water molecules contribute to the reflections as a continuous electron density where the 10 electrons of the water molecules are smeared over a volume of about 30 Å^3. Such an approach has been used for lysozyme crystals by e.g. Blake et al. (1983) in order to get a better agreement between $|F_H^o|$- and $|F_H^c|^2$-values in the low resolution regime. However, the smearing of the electron density over the "empty" part of the unit cell is not unambiguous. Problems arise on the "surface" of the protein and in the holes in the molecule not occupied by water molecules.

For our calculations the unit cell was enlarged in all directions by $2 \cdot r_w + r_c$ and divided into grid points with an edge length of 0.25 Å. Here r_w and r_c are the van der Waals radii of water and carbon respectively. For r_w we used 1.5 Å. Now all grid points within the van der Waals radii of the individual atoms were determined. These are the protein occupied grid points. We used the same radii as Tilton et al. (1984) for the N ($r_N = 1.85$ Å), O ($r_O = 1.7$ Å), S ($r_S = 1.85$ Å) and Fe ($r_{Fe} = 2.0$ Å) atoms. Tilton et al. (1984) have taken into account the contribution of the H-atoms by increasing the van der Waals radii of the C-atoms (CH: $r_c = 1.95$ Å; CH_2: $r_c = 2.00$ Å; CH_3: $r_c = 2.10$ Å). Instead we used an average value $r_c = 2.05$ Å for all carbon atoms. All remaining grid points were tested to see whether they could be occupied by water. In principle, a hole is water accessible if one grid point exists, which has at least a distance of r_w in all directions to protein occupied grid points. If such a grid point is found all grid points within a sphere of radius r_w are water accessible. All grid points not fulfilling this condition are water forbidden. Note that the water forbidden grid points include all protein grid points and all holes within the protein which are too small to be occupied by water. Water forbidden grid points having contacts to water accessible grid points form the water accessible surface of the protein.

The water accessible part is 32.7% of the unit cell. We now filled these grid points with a homogeneous electron density of $\varrho_w = 0.40$ electrons/Å^3. The structure factors, F_H^c (H_2O), for the homogeneous water density were obtained using a fast Fourier algorithm. The protein occupied grid points gave the protein structure factors F_H^c (prot), with the help of the same algorithm. The total structure amplitude was obtained from

$$F_H^c = F_H^c(\text{prot}) + S\, F_H^c(H_2O)\, e^{-B \sin^2 \vartheta/\lambda^2} \quad (1)$$

A scalling factor S and the B-factor were obtained from a least squares fit of $|F_H^c|$ to the observed structure factors. The scaling corresponds to an adjustment of the homogeneous electron density ϱ_w which became 0.375 electrons/Å^3. The B-factor was introduced in order to smooth the border between water and protein. The fit yielded $B = 31.91$ Å^2 corresponding to $\langle x^2 \rangle = 0.40$ Å^2. Figure 6 compares the $|F_H^c|^2$-values thus obtained with the observed structure factors. The agreement is much better than in the case where only the myoglobin atoms with 160 waters/molecule were used for the calculation of the structure factors.

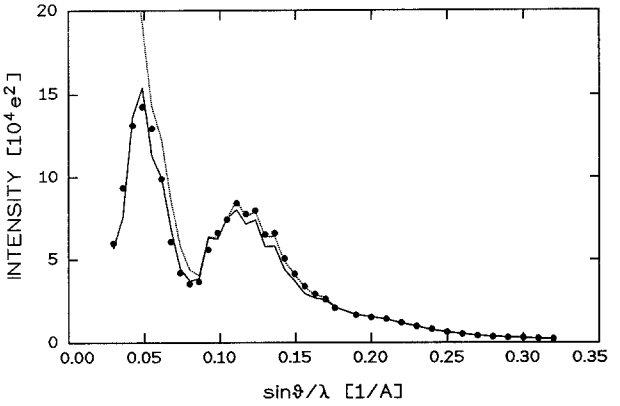

Fig. 6. Averaged $|F_H^0|^2$-values as function of $\sin\vartheta/\lambda$. Circles: $|F_H^0|^2$-values from Hartmann et al. (1987). Solid line: $|F_H^c|^2$ using a homogeneous electron density for the crystal water; dotted line: $|F_H^c|^2$ from myoglobin + 160 waters/myoglobin

Table 1. Standard R-values from different refinements

Resolution	> 5.5 Å	< 5.5 Å	total
No waters	0.598	0.217	0.240
320 X-ray visible only	0.434	0.176	0.193
Homogeneous e^--density	0.207	0.213	0.212
320 X-ray visible and Homogeneous e^--density	0.201	0.176	0.178
814 Monte Carlo waters	0.161	0.167	0.166

A further improvement is obtained by using the coordinates of the crystallographically visible 160 waters/myoglobin and filling the remaining empty space in the unit cell only with a homogeneous electron density (see Table 1).

It should be mentioned that the fast Fourier algorithm was controlled by calculating the structure factors from the protein grid points which were compared with those calculated from the coordinates and the $\langle x^2 \rangle$-values of myoglobin. The standard R-value was only 0.0027.

Discussion

X-ray diffraction data of protein crystals are sensitive to the total content of the crystallographic unit cell. Disorder reveals itself by reducing the structure factors. The information content of the structure factors depends on their Miller indices. The Miller indices, the scattering angle and the corresponding resolution are strictly correlated. Structure factors belonging to low resolution carry the information on crystallographically invisible contents of the unit cell.

Table 1 demonstrates the influence of the crystal water on the structure factors. In the low resolution regime the standard R-value is lowered from 0.598 when the water is completely neglected to 0.161 when refined Monte Carlo water coordinates are taken. Note that the homogeneous electron density yields an unsatisfactory R-factor at high resolutions. It is improved by using the crystallographi-

cally visible waters and filling the remaining space with a continous electron density.

Recently, Cheng and Schoenborn (1990) have used neutron scattering data to get information on the crystal water of carbonmonoxymyoglobin crystals. In the refinement of their data they varied the water density as a function of the distance from the protein surface. Defining as $R = 0$ the center of the surface atoms which are essentially hydrogen or deuterium, density maxima are obtained at about $R = 2.3$ Å and $R = 3.75$ Å. At these distances the B-factors of the hydration water molecules become minimal. From these results one can conceive a clear picture of the hydration shells of a protein. However, there remains the question of whether this interpretation of the low resolution scattering factors is unambiguous. A comparison with our Monte Carlo water molecules allows an independent control of the density fluctuations in the hydration shell. In Fig. 7 the results for the 814 Monte Carlo water molecules are shown. Their coordinates were used to calculate the electron density of the water accessible grid points. We then used our previously defined protein surface and divided the water accessible space into shells with the thickness $\Delta R = 0.3$ Å

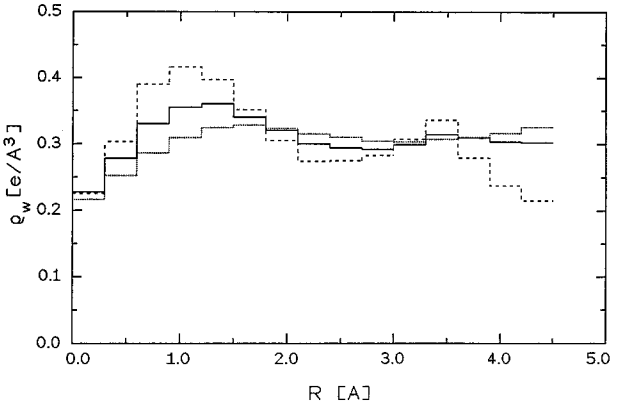

Fig. 7. Electron density in the water accessible space of the unit cell. The surface of the protein is defined by a rolling sphere of 1.5 Å radius. The Monte Carlo waters of the 966 Monte Carlo simulation are smeared out by $B = 20$ Å (dashed line); $B = 50$ Å (solid line); $B = 100$ Å^2 (dotted line)

and a distance R from the protein surface. We obtain similar density fluctuations as Cheng and Schoenborn (1990). Using $B = 50$ Å^2 for the Monte Carlo water molecules we got a density maximum at 1.35 Å. The discrepancy with Schoenborn comes from the different definition of $R = 0$. With $R = 0$ in the center of the surface atoms including the H-atoms added to the X-ray structure as described above, the first density maximum occurs at 2.5 Å in comparison to the 2.3 Å obtained by Cheng and Schoenborn (1990). This agreement is very good. It proves the physical significance of the picture. It also demonstrates that Monte Carlo calculations are the right tool to complement crystallographic information on proteins.

The magnitude of the density fluctuations depends on the B-value which we imposed on the Monte Carlo water molecules. At $B = 100$ Å^2 the maxima are nearly smeared out and approach the continuous water model. Starting with $B = 50$ Å^2 for all Monte Carlo waters the refinement gives individual mean square displacement with $B = 58.6$ Å^2 (corresponding to $\langle x^2 \rangle = 0.742$ Å^2) on the average. The density distribution using the refined water coordinates is close to that shown in Fig. 7. There is a slight tendency to move the waters to larger distances from the protein surface. Note that the crystallographically visible waters had an average B-value of 33.2 Å^2 ($\langle x^2 \rangle = 0.42$ Å^2). A disorder corresponding to a root mean square displacement of 0.86 Å makes an oxygen atom crystallographically invisible.

In the refinements the electron density of the water has to compensate for the missing scattering of the NH_4^+ and the SO_4^{--} ions in the crystal water. For CO ligated myoglobin Schoenborn (1988) found 14 $(NH_4)_2SO_4$ molecules per unit cell from their contributions to neutron scattering. We dissolved some large metmyoglobin crystals and determined the NH_4^+ content using the reaction:

$$\alpha\text{-ketoglutarate} + NH_4^+ + NADH \rightarrow \text{glutamate} + H_2O + NAD^+ \quad (2)$$

catalyzed by glutamate dehydrogenase. As the result we obtained 26 ± 5 ammonium sulfate molecules per unit cell. With a 3.72M ammonium sulfate solution the unit

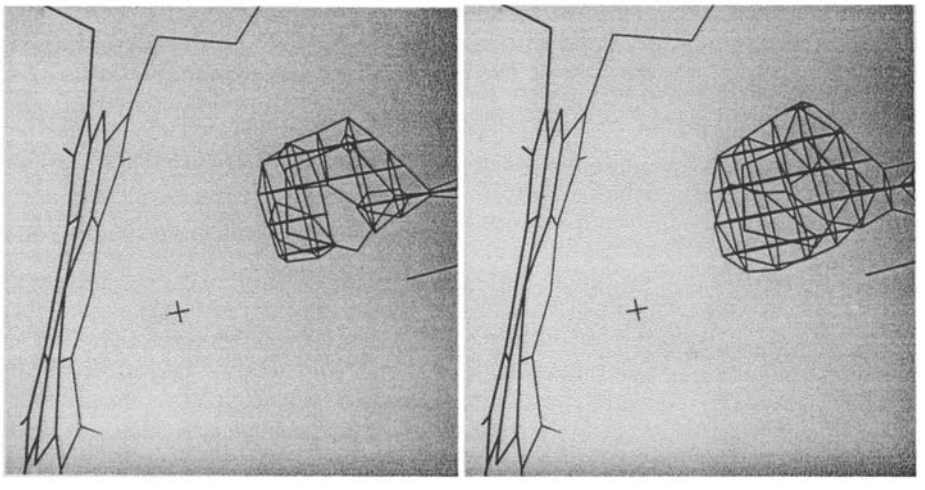

Fig. 8. ($F^0 - F^c$)-maps of the heme region. The sidechain atoms of His E7 have been omitted from the F^c calculation. Both maps are contoured at a density level of 0.7 e/Å^3. *Left:* **a** only X-ray water included; *right:* **b** Monte Carlo water included

cell would contain 49 ammonium sulfates. As expected the crystal water contains less ammonium sulfate than the mother liquor. Calculating the crystal density with 2 myoglobin, 26 ammonium sulfate and 632 water molecules one obtains $\varrho_K^c = 1.25$ g/ml. In good agreement with this value an experimental determination in the ultracentrifuge using a density gradient yielded $\varrho_K^0 = 1.23 \pm 0.06$ g/ml.

The refinement with Monte Carlo waters shows that the water coordinates of the crystallographically invisible water molecules are not completely uncorrelated in different unit cells. In myoglobin crystals the protein structure seems to influence all crystal waters in the unit cell. In our paper we demonstrate a new method to analyze data from X-ray experiments. After the usual structure determination the unit cell is completed by Monte Carlo simulations, yielding coordinates of the invisible atoms. This way, water molecules and hydrogen atoms can be included in a next step of refinement. Appropriate individual *B*-values insure that these atoms contribute only to the low resolution structure factors. Alternating cycles of crystallographic refinements and new energy adjustments by Monte Carlo simulations allow an optimization of the structural model. This procedure provides not only information on the crystal water but may also improve the protein structure. This is shown in the following. We calculated $(F^0 - F^c)$-maps for two models. While for the map with the Monte Carlo waters all reflexions could be used, the resolution range was restricted from 7.0 Å to 1.5 Å in the case with the 320 crystallographically visible waters because of the bad fit of low order reflexions. The noise level is about 15% lower for the map with the Monte Carlo waters ($\sigma = 0.075$ e/Å^3 compared to 0.086 e/Å^3) on the average over the whole unit cell and slightly lower in the central protein regions. The electron density for protein atoms omitted in a calculation of $(F^0 - F^c)$-maps was generally about 15% higher in the case of the Monte Carlo water model. One example is shown in Fig. 8 a for the model with only X-ray waters included and in Fig. 8 b for the model with refined Monte Carlo waters. Here, the atoms of the sidechain of His E7 (H64) are omitted from the $(F^0 - F^c)$-synthesis. Both pictures are contoured at the same density level of 0.7 e/Å^3. One can clearly see that the model with Monte Carlo waters gives better results. The signal to noise ratio in this case is better by about 20%. The improvement of the accuracy of atomic parameters should be of the same magnitude. The estimation of coordinate errors according to the methods of Cruickshank (1949) and Luzzati (1952) shows a reduction of the errors of about 10% for the refined structure with the Monte Carlo water.

Acknowledgements. This work was supported by the Deutsche Forschungsgemeinschaft, the Fond der Chemie and "Regione Autonoma Sardegna".

References

Ahlström P, Teleman O, Jönsson B, Forsén S (1987) Molecular dynamics simulation of parvalbumin in aqueous solution. J Am Chem Soc 109:1541–1551

Blake CCF, Pulford WCA, Artymiuk PJ (1983) X-ray studies of water in crystals of lysozyme. J Mol Biol 167:693–723

Bone S, Pethig R (1982) Dielectric studies of the binding of water to lysozyme. J Mol Biol 157:571–575

Careri G, Geraci M, Giansanti A, Rupley JA (1985) Protonic conductivity of hydrated lysozyme powders at megahertz frequencies. Proc Natl Acad Sci USA 82:5342–5346

Cheng X, Schoenborn BP (1990) Hydration in protein crystals. A neutron diffraction analysis of carbonmonoxy myoglobin. Acta Crystallogr B46:195–208

Clementi E, Cavallone F, Scordamaglia J (1977) Analytical potentials from ab-initio computations for the interaction between biomolecules. 1. Water with amino acids. J Am Chem Soc 99:5531–5545

Clementi E, Corongiu G, Aida M, Niesan U, Kneller G (1990) Monte Carlo and molecular dynamics simulations. In: Clementi E (ed) MOTECC-90. ESCOM, Leiden, pp 805–888

Clementi E, Corongiu G (1989) KGNMOL: A program for molecular interactions. In: Clementi E (ed) MOTECC-89. ESCOM, Leiden, pp 243–306

Corongiu G, Clementi E (1978) Intramolecular and intermolecular interactions for deriving chemical formulae and for simulate complex chemical systems. Gazz Chim Ital 108:273–280

Cruickshank DWJ (1949) The accuracy of electron-density maps in X-ray analysis with special reference to dibenzyl. Acta Crystallogr 2:65–82

Dachwitz E, Parak F, Stockhausen M (1989) On the dielectric relaxation of aqueous myoglobin solutions. Bericht. Bunsen-Ges f phys Chemie 93:1465–1458

Finney JL, Goodfellow JM, Howell PL, Vovelle F (1985) Computer simulation of aqueous biomolecular systems. J Biomol Struct Dyn 3:599–622

Frauenfelder H, Parak F, Young RD (1988) Conformational substates in proteins. Ann Rev Biophys Chem 17:451–479

Frauenfelder H, Petsko GA, Tsernoglou D (1979) Temperature-dependent X-ray diffraction as a probe of protein structural dynamics. Nature 280:558–563

Goodfellow JM (1987) Computer simulation of hydration networks around amino acids. Int J Biol Macromol 9:273–280

Van Gunsteren WF, Berendsen HJC, Hermans J, Hol WGJ, Postma JPM (1983) Computer simulation of the dynamics of hydrated protein crystals and its comparison with X-ray data. Proc Natl Acad Sci USA 80:4315–4319

Hartmann H, Steigemann W, Reuscher H, Parak F (1987) Structural disorder in proteins: A comparison of myoglobin and erythrocruorin. Eur Biophys J 14:337–348

Jeffrey GA, Saenger W (1991) Hydrogen Bonding in Biological Structures. Springer, Berlin Heidelberg New York

Konnert JH, Hendrickson WA (1980) A restrained thermal factor refinement procedure. Acta Crystallogr. A36:344–349

Kotitschke K, Kimmich R, Rommel E, Parak F (1990) NMR study of diffusion in protein hydration shells. Prog Colloid Polym Sci 83:211–213

Krupyanskii YF, Goldanskii VI, Nienhaus GU, Parak F (1990) Dynamics of protein water systems revealed by Rayleigh scattering of Mössbauer radiation (RSMR). Hyperfine Interact 53:59–73

Kuntz ID, Kauzmann W (1974) Hydration of proteins and polypeptides. Adv Protein Chem 28:239–345

Luzzati V (1952) Traitement statistique des erreurs dans la determination des structures cristalline. Acta Crystallogr 5:802–810

Mathur-De-Vré R (1979) The NMR studies of water in biological systems. Progr Biophys Mol Biol 35:103–134

Matsuoka O, Clementi E, Yoshimine M (1976) CI study of the water dimer potential surface. J Chem Phys 64:1351–1361

Metropolis N, Rosenbluth AW, Rosenbluth AH, Teller J, Teller F (1953) Equation of state calculations by fast computing machines. J Chem Phys 21:1087–1092

Otting G, Liepinsh E, Wüthrich K (1991) Protein hydration in aqueous solution. Science 254:974–980

Otting G, Wüthrich K (1989) Studies of protein hydration in aqueous solution by direct NMR observation of individual protein-bound water molecules. J Am Chem Soc 111:1871–1875

Phillips SEV (1980) Structure and refinement of oxymyoglobin at 1.6 Å resolution. J Mol Biol 142:531–554

Pissis P (1989) Dielectric studies of protein hydration. J Mol Liquids 41:271–289

Rupley JA, Careri G (1989a) Enzmye hydration and function. In: Cooper A, Houben JL, Chien LC (eds) The enzyme catalysis process. Energetics, mechanism and dynamics. Plenum Press, New York, pp 223–234

Rupley JA, Careri G (1989b) Percolation process. In: Cooper A, Houben JL, Chien LC (eds) The enzyme catalysis process. Energetics, mechanism and dynamics. Plenum, Press, New York, pp 235–246

Schoenborn BP (1988) Solvent effect in protein crystals. A neutron diffraction analysis of solvent and ion density. J Mol Biol 201:741–749

Singh GP, Parak F, Hunklinger S, Dransfeld K (1981) Role of adsorbed water in the dynamics of metmyoglobin. Phys Rev Lett 47:685–688

Steinbach PJ, Loncharich RJ, Brooks BR (1991) The effects of environment and hydration on protein dynamics: a simulation study of myoglobin. Chem Phys 158:383–394

Takano T (1977) Structure of myoglobin refined at 2.0 Å resolution: Crystallographic refinement of metmyoglobin from sperm whale. J Mol Biol 110:537–568

Tilton RF, Kunz JO, Petsko GA (1984) Cavities in proteins: Structure of metmyoglobin Xenon complex solved to 1.9 Å. Biochem 23:2829–2857