UCLF-8206240-13

THE STRUCTURE OF BOUND WATER AND REFINEMENT OF ACID METMYOGLOBIN *

N.V. Raghavan and B.P. Schoenborn

BNL--34098

Brookhaven National Laboratory
Upton, NY 11973

DE84 006541

INTRODUCTION

The structure of myoglobin as determined from x-ray diffraction data was first reported by Kendrew et al. (1). More recently, Takano (2) described the acidmet and deoxy forms, and Phillips (3) described the oxy form. In these x-ray structures, the positions of the hydrogen atoms cannot be described, but they are important factors in determining the interactions of the heme with its environment. Neutron diffraction work, done in this laboratory, has demonstrated that hydrogen and deuterium positions can be located (4). In addition to the localization of H and D, neutron diffraction provides a unique method for studying the water structure because of the strong scattering ability of D₂O. The scattering factor of deuterium is nearly twice as large as that of hydrogen, and it increases the visibility of water molecules in Fourier maps, so that in a neutron map a water molecule appears about three times as strong as in an equivalent electron-density map. Detailed results of the neutron diffraction analysis of carbonmonoxy myoglobin (5) and oxymyoglobin (6), and the initial results on acid metmyoglobin (7) have been reported. The reciprocal-space refinement of acid metmyoglobin and in particular the details of the water structure and hydrogen exchange are presented here.

CRYSTAL DATA AND REFINEMENT

Crystals were grown from deuterated ammonium sulfate at pH 5.6 and stored in the mother liquor for nearly ten years. The data were collected at the Brookhaven National Laboratory High Flux Beam Reactor. An incident wavelength of 1.52 Å was selected with a pyrolytic graphite monochromator. A graphite filter was used to reduce the $\lambda/2$ contamination. The flux at the crystal was

*Research carried out under the auspices of the U.S. Department of Energy under Contract No. DE-AC02-76CH00016.

MASTER

$\sim 10^6$ neutrons/cm²-sec. Data were collected in ϕ -scans of 1.3° in 21 steps with scan times between 7 and 15 minutes.

The reciprocal-space refinement was carried out by using the restrained parameter least-squares refinement of Konnert and Hendrikson (8). In this technique, information about the stereochemistry of the structure is treated as an additional observation and serves to reduce the number of variables refined independently. There are 2544 atoms, 8400 distances, 207 planes, 182 chiral centers, 762 torsion angles, 21,000 possible contacts, and 2500 non-bonded contacts in the molecule. The total number of variables is 10,200, and the number of structure factor observations $>2\sigma$ is 12,000 to a resolution of 1.8 Å. The real-space refined neutron coordinates of metmyoglobin were used as starting coordinates. Two cycles of refinement reduced the R from 0.39 to 0.30. Comparison of the coordinates with the starting set suggested that most of the large shifts were in hydrogen and deuterium atoms, and that the r.m.s. shift for all the atoms is 0.20 Å. The structure was examined by using computer graphics, and minor corrections were made to a few lysine side chains. The refinement continued as shown in Table 1.

The variation of the isotropic temperature factor, B, as a function of residue number is shown in Figure 1. The average

Table 1. Summary of Refinement

Run No.	0	2	4	8
$R = \frac{ F_o - F_c }{ F_o }$	0.392	0.30	0.265	0.230
r.m.s. bond distance	0.018	0.024	0.050	0.050
σ	0.030	0.030	0.030	0.030
$\langle F_o - F_c \rangle$	94	72	65	56
σ	40	40	25	22
r.m.s. coordinate shift	0.15	0.09	0.05	0.04
r.m.s. B shift (Å ²)	3.4	0.5	1.0	0.7
No. of atoms	2528	2528	2528	2528
No. of reflections used	10533	10533	10533	10533

The values of σ are the input estimated standard deviations that determine the relative weights of the corresponding restraints.

Fig. 1. Variation of temperature factor (B) as a function of residue number. The dotted line indicates the average temperature factor of main-chain atoms, and the solid line, that of side-chain atoms. The helices are indicated at the top of the figure.

main-chain temperature factors for the B, D, and E helices are smaller than those for the C, A, and G helices. Also, the temperature factors for residues in the interhelical regions are generally large, particularly for the bend between the E and F helices. The temperature factors of the side-chain atoms are generally only slightly larger than those of the main-chain atoms except in the case of lysines, as shown in Figure 1. For these lysine surface residues the side-chain temperature factor is 7 to 10 $\text{\AA}^2$ larger than that for the corresponding main-chain atoms.

PROCEDURE FOR THE REFINEMENT OF THE SOLVENT

By using the refined set of coordinates, all the symmetry related atoms are calculated, and a list of all the atoms in the asymmetric unit is obtained. A Fourier map based on $|F_o|$ as coefficient is calculated with 1- $\text{\AA}$ grid spacing and data to 2.8- $\text{\AA}$ resolution. From this Fourier map regions corresponding to the protein and its symmetry relatives are eliminated. The excluded protein volume is calculated from the atomic coordinates and the van der Waals radii. Such a difference Fourier map is shown in Figure 2 with the solvent region shown by dark lines and the protein region cross-hatched. Features in the now excluded regions can be due to structured water, ions, and error peaks that might

Fig. 2. A section of the protein-excluded $|F_0|$ Fourier map. The x direction is along the a-axis and is 64 Å, and the z direction is along the c-axis and is 35 Å. The solvent region is contoured at 0.5 F/Å^3 . The protein region is cross-hatched.

arise from series termination and other errors. In this analysis a Fourier map is used instead of a difference Fourier, to minimize errors in the structure factors and the model. The Fourier map shows that nearly 43% of the cell volume is solvent, and this is equivalent to the calculated value. An asymmetric unit of the cell contains a total of 36,900 points at 1-Å grid spacing with 15,800 points in the solvent region. For the solvent region, the average observed density corresponds to 0.30 F/Å^3 , which is 60% of the theoretical density of 0.49 F/Å^3 for D_2O scattering. Note that no phase contributions from solvent were used in calculating the Fourier map. With use of the theoretical scattering value of 0.49 F/Å^3 for pure D_2O , structure factors were calculated to a resolution of 2.8 Å, from the coordinates of the 15,800 solvent points. A scale factor and an overall temperature factor were determined by χ^2 minimization. The function minimized is $F_0 - F_{\text{total}}$ where

$$F_{\text{total}} = \sqrt{(A_{\text{protein}} + K_1 A_{\text{solvent}})^2 + (B_{\text{protein}} + K_1 B_{\text{solvent}})^2}$$

where the A_{solvent} and B_{solvent} terms include an exponential term

due to the temperature factor of the form $\exp(-k_2 \sin^2 \theta / \lambda^2)$. The values for the scale factor and temperature factor obtained are 1.35 and $156 \text{ \AA}^2$. After applying these factors, a new protein-excluded Fourier map was calculated. The average density in this map was $0.36 \text{ F/\AA}^3$ with a standard error of $0.005 \text{ F/\AA}^3$. This average density is 75% of the theoretical scattering density and agrees well with the $\text{D}_2\text{O}/\text{H}_2\text{O}$ ratio of $80 \pm 10\%$ as determined by the NMR technique. Table 2 shows a sample of such solvent structure factors.

From Table 2 the following observations can be made about the structure factor contributions from the bulk solvent:

1. For the low-angle data the contributions from the protein and the bulk solvent are of the same order of magnitude.

2. The phase of the bulk solvent is always different from that of the protein. For the centrosymmetric reflections the coefficients can be of the same order of magnitude but of opposite sign. For the non-centrosymmetric reflections the phase of the protein and that of the bulk solvent are generally in different quadrants of the phase diagram.

3. The inclusion of solvent in the structure factor calculation reduced R to 0.21.

Table 2. Few Reflections Showing the Solvent Contribution and Phases

h	k	l	$\sin \theta / \lambda$	F_{obs}	Protein		Solvent		Total	
					$ F $	ϕ	$ F $	ϕ	$ F $	ϕ
0	1	4	0.0618	1074.	637.	-97.	342.	-21.	795.	-72.
0	2	1	0.0356	1028.	1745.	10.	1007.	-158.	783.	-4.
1	1	-5	0.0745	966.	1014.	-64.	119.	92.	906.	-61.
1	3	-3	0.0649	1114.	1381.	155.	302.	-28.	1078.	156.
2	1	3	0.0539	734.	852.	-135.	281.	30.	584.	-128.
-2	2	-2	0.0439	667.	1033.	-76.	505.	92.	548.	-64.
2	2	4	0.0733	675.	494.	-63.	251.	-119.	666.	-81.
2	3	0	0.0511	1364.	1373.	165.	426.	14.	1024.	154.
2	0	-2	0.0298	109.	446.	0.	302.	180.	144.	0.
2	0	-1	0.0187	1557.	634.	0.	1738.	180.	1104.	180.
2	0	0	0.0161	1910.	541.	180.	1889.	0.	1348.	0.
2	0	1	0.0247	1022.	271.	0.	1286.	180.	1005.	180.
2	0	2	0.0375	834.	482.	180.	148.	180.	630.	180.
2	0	3	0.0515	125.	258.	180.	164.	0.	94.	180.
3	0	-1	0.0247	133.	840.	0.	1021.	180.	181.	180.
4	0	-1	0.0316	138.	1082.	0.	1440.	180.	358.	180.
4	0	0	0.0322	203.	1060.	180.	600.	0.	460.	180.

A search of this Fourier map for additional bound water molecules using BILDER, an interactive graphics program (9), picked up 30 additional water molecules with a density >0.5 F. After the solvent contribution was included in the observed structure factors, the protein structure was further refined; the results are shown in Table 3. The quantity refined is $\Sigma(F_o - F_c)^2$ where the F_c now includes the contributions from both the protein and the solvent.

It is clear from this table that inclusion of the bulk solvent has a dramatic effect on the conventional R-factor. The value for the shell to 5.0-Å resolution drops from 0.25 to 0.13 in four cycles. Smaller but significant changes can be seen also for the shells from 3.5 to 5.0 and 3.0 to 3.5 Å. No real changes appear in the other shells. This indicates that the bulk solvent contributes to the structure factor to a resolution of 3.0 Å but beyond that its effect is negligible. At the present stage of refinement (Table 4) the overall R-factor is 19%, but for the shells from 2.2 to 1.7 the residual index is still high and further refinement is needed.

DISCUSSION

Protein-excluded Fourier maps are also useful in picking out certain regions of the structure that are disordered. These regions have a larger average density and probably are not weakly bound water molecules since they are close to non-polar atoms. One such region is the GH corner, shown in Figure 3. The residue GH1 hydrogen-bonds to residue B5, which in turn hydrogen-bonds to

Table 3. Conventional R-Factors are Listed, Indicating the Progress of Refinement After Phasing of Solvent Region

Cycle	Solvent Contributes to F			No Solvent Contrib. to F				Total
	$d_{min} = 5 \text{ Å}$	3.5 Å	3.0 Å	2.5 Å	2.2 Å	2.0 Å	1.7 Å	
0	0.25	0.18	0.18	0.18	0.18	0.24	0.28	0.21
1	0.24	0.19	0.18	0.17	0.17	0.21	0.25	0.20
2	0.22	0.18	0.17	0.17	0.17	0.22	0.25	0.20
3	0.20	0.16	0.16	0.17	0.18	0.23	0.26	0.19
4	0.17	0.15	0.15	0.17	0.18	0.23	0.26	0.19
5	0.13	0.13	0.14	0.17	0.18	0.23	0.27	0.19

The temperature factor B of 156 Å^2 reduces the contribution of solvent water so that no contributions are observed beyond 2.8 Å .

Table 4. Present Stage of the Refinement

$d_{\min}$	$\langle\sigma_F\rangle$ Data	$\langle\sigma_F\rangle$ Used	$\langle F_o \rangle$	Shell R	Sphere R	Refln
5.0	6.4	11.0	514.	0.130	0.13	490
3.5	8.6	11.3	409.	0.130	0.13	992
3.0	12.3	13.4	323.	0.140	0.133	899
2.5	15.2	15.7	252.	0.172	0.145	1721
2.2	16.4	16.8	210.	0.182	0.154	1837
2.0	19.5	19.6	177.	0.234	0.166	1743
1.7	22.2	22.2	165.	0.267	0.187	3132

Fig. 3. A section of a Fourier map showing histidine GH1 hydrogen-bonded to histidine B5. Histidine B5 is hydrogen-bonded to aspartic acid B1.

the carbonyl oxygen of residue B1. In the protein-excluded Fourier map there is a large density near His GH1 that is unlikely to be a bound water molecule because of its size and its proximity to non-polar atoms of His GH1. Rotation about the $C_\alpha-C_\beta$ bond of His GH1 shows that this ring fits nicely into the density. His GH1 is disordered, and both positions are partially occupied. In one position, where the histidine hydrogen-bonds to His B5, the coordinates and dihedral angles are similar to those observed in carbonmonoxy myoglobin (5), and in the other position, where the hydrogen bond is broken, the dihedral angles are similar to those observed in alkaline metmyoglobin (10). The relative occupancy between sites 1 and 2 is ~ 0.50 and 0.50 .

Comparison of the heme environments in the three structures shows that in metmyoglobin and oxymyoglobin there is a deuteron on N_ϵ of the distal histidine E7 whereas in carbonmonoxy myoglobin no deuteron is observed: although the $N_\epsilon-O$ distance is only $2.7 \text{ \AA}$, the orientation of $N_\epsilon-O$ is not favorable for hydrogen-bonding. In metmyoglobin and carbonmonoxy myoglobin, no water or sulfate ion is bound to N_δ of His E7 whereas in oxymyoglobin and the x-ray refinement of metmyoglobin (2) a sulfate ion has been observed bound to N_δ of His E7. This difference might be due to differences in pH or in ionic strength between the various studies.

HYDROGEN EXCHANGE

The scattering length of an exchangeable atom is given by the relation $f = 6.7x + (1 - x)(-3.7)$ where 6.7 and -3.7 are the scattering lengths of deuterium and hydrogen respectively, and x is the fractional deuterium present. The refinement is started with a value of $x = 0.355$, at which the scattering length is zero, and no judgment is made as to whether an atom is a hydrogen or deuterium. Also, since the scattering length of an amide hydrogen is strongly dependent on the parameters of the amide nitrogen, and the resolution of the data is only $1.8 \text{ \AA}$, the occupancy parameters of all the main-chain atoms are varied. The positional parameters and scattering lengths are varied in one cycle and the positional and thermal parameters in the next cycle. After five cycles of this alternate refinement the main-chain deuteriums showing little exchange are listed (Table 5). The distribution of occupancy of main-chain amide hydrogen as a function of residue number is shown in Figure 4. As seen in Table 5, there are 10 hydrogens with little exchange, with $x = 0.27$ to 0.42 , but no hydrogen showing no exchange. An estimate of the minimal expected error in the occupancy numbers can be made from the distribution of the observed occupancy of the H_α , which are not exchangeable. This estimate is about one-third the scattering of a hydrogen atom, as shown in

Table 5. Main-Chain Residues Showing Low Exchange in Metmyoglobin and Carbonmonoxy Myoglobin

Residue No.	Residue	Met	CO
6	Glu A4	0.30	0.86
10	Val A8	0.37	0.30
11	Leu A9	0.34	0.94
17	Val A15	0.63	0.26
29	Leu B10	0.42	0.60
33	Phe B14	0.48	0.35
70	Thr E13	0.35	0.41
72	Leu E15	0.14	0.35
104	Leu G5	0.78	0.37
111	Ile G12	0.41	0.28
112	Ile G13	0.40	0.37
115	Leu G16	0.24	0.25
131	Met H7	0.36	0.49
134	Ala F10	0.47	0.24
135	Leu H11	0.42	0.50

Fig. 4. The fractional exchange of main-chain amide hydrogens as a function of residue number. Values below 0.35 are for hydrogens and values above 0.35 are for deuterons. The helices are indicated at the bottom of the figure.

Figure 5. This error can be due to limited data, high-order data with poor counting statistics, and errors in phasing.

A histogram of the scattering lengths of all exchangeable hydrogens including those on side chains is shown in Figure 6. Starting the refinement with all exchangeable hydrogens as deuterons, with a value of $x = 1.0$, causes a strong bias toward full exchange, and only 50 partially exchanged, and no unexchanged hydrogens are observed. Starting the refinement with a value of $x = 0.35$ (no hydrogen or deuteron assumed) results in 67 fully exchanged, 208 partially exchanged, and no unexchanged hydrogens. The median of the distribution is 70% exchange, which agrees well with the H_2O/D_2O ratio observed in the bulk solvent and the ratio obtained by the NMR technique. It is very important that the refinement be initiated with zero scattering ($x = 0.35$) to avoid the strong bias toward full exchange.

Comparison with the refinement of carbonmonoxy myoglobin (11) and oxymyoglobin (6) shows some similarities in the hydrogen exchange behavior despite differences in crystal growth and crystal soaking conditions. It should be reiterated that crystals of met-myoglobin are grown in $(ND_4)_2SO_4$ and D_2O at pH 5.6. In carbonmonoxy myoglobin, the crystals of which are grown at pH 5.6 and soaked in D_2O for several months, residues 10, 17, 33, 72, 104, 111, 112, 115, and 134, which are located in helices, show the least exchange. In oxymyoglobin, the crystals of which are grown at pH 5.6 and then transferred to pH 8.4 and soaked in D_2O for 5 months, Phillips (6) showed that residues 28, 29, 30, 32, 69, 107, 135, 138, 142, and 143 are completely unexchanged, with deuterium occupancies <0.10 , and that other residues 10, 17, 20, 33, 66, 72, 75, 85, 112, 114, 116, and 134 are partially exchanged, with values between 0.10 and 0.40. Clearly there is some similarity between acidmet and carbonmonoxy myoglobin, both at pH 5.6, whereas oxymyoglobin at pH 8.4 seems to have a greater number of residues that are either unexchanged or partially exchanged, but there are residues common to all three. Note that, for oxymyoglobin, not only were the pH and soaking time different, but also a different refinement technique was used (6). A more detailed comparison of the differences as a function of position will be published elsewhere.

ACKNOWLEDGMENTS

The authors would like to thank Dr. S.E.V. Phillips for use of his results prior to publication.

Fig. 5. Distribution of the observed occupancy of the H_A , which are not exchangeable.

Fig. 6. Distribution of scattering length of all exchangeable hydrogens, (a) when the refinement is started with $x = 1.0$ (full exchange), and (b) when the refinement is started with $x = 0.35$ (zero scattering).

REFERENCES

1. Kendrew, J.C., Dickerson, R.E., Strandberg, B.E., Hart, R.G., Davies, D.R., Phillips, D.C., and Shore, V.C., Nature (London) 185:422-7 (1960).
2. Takano, T., J. Mol. Biol. 110:537-68 (1977).
3. Phillips, S.E.V., J. Mol. Biol. 142:531-54 (1980).
4. Schoenborn, B.P., Nature (London) 224:143-6 (1969).
5. Hanson, J.C. and Schoenborn, B.P., J. Mol. Biol. 153:117-46 (1981).
6. Phillips, S.E.V., See paper in this Symposium.
7. Schoenborn, B.P., Cold Spring Harbor Symp. Quant. Biol. 36:569-75 (1971).
8. Hendrikson, W.A. and Konnert, J.A., in: "Biomolecular Structure, Function, Conformation and Evolution," Vol. 1, pp. 43-57, R. Srinivasan, ed., Pergamon, Oxford (1980).
9. Diamond, R., in: "Computing in Crystallography," pp. 27.02-27.15, R. Diamond et al., eds., The Indian Academy of Sciences, Bangalore, India, for the International Union of Crystallography (1980).
10. Schoenborn, B.P., J. Mol. Biol. 45:279-303 (1969).
11. Raghavan, N.V. and Schoenborn, B.P., J. Mol. Biol., in press (1982).

DISCUSSION

ENGELMAN: You have observed that no amide hydrogens are substantially unexchanged in the myoglobin crystal grown in D₂O, whereas other studies have shown some unexchanged positions in crystals grown in H₂O and subsequently changed to D₂O. Do you think this difference arises from some constraint on protein exchange imposed by crystal interactions, or that it represents the uncertainty of determining the occupancy of some positions?

RAGHAVAN: The error in the occupancy determination is still large, but it is safe to say that, in crystals grown in D₂O and left soaking for 10 years, no amide hydrogens are substantially unexchanged. I would hesitate to invoke crystal interactions to explain my results because the errors and the long soaking time are likely to be reasons for the differences from other studies, in which the crystals are grown in H₂O and soaked in D₂O.

WLODAWER: Were the water positions located during the analysis of bulk solvent included in the calculation of structure factors using Hendrickson's program, which led to a low R-factor at very low resolution? Did you include bulk solvent in the calculation?

RAGHAVAN: Yes, the bulk solvent is included in the calculation. In order to do this in Hendrickson's program the observed structure factor amplitudes are corrected for the bulk solvent contribution.

MASON: It is very nice to see a careful treatment of the relative contributions of bulk solvent and protein to the low-angle reflections. Since the degree of deuteration of the crystals is somewhat uncertain, great caution needs to be taken for these reflections. A simple check would be to use the same model as that used for the deuterated data, and to predict the intensities of the low-angle reflections from a non-deuterated crystal. Have you measured these reflections, on such a crystal, as a check on your treatment of solvent?

RAGHAVAN: A data set from a non-deuterated crystal has been measured. It has not been processed yet. It remains to be seen how closely the measured intensities of the low-angle reflections correspond to those predicted from the model.

PHILLIPS: In your solvent correction and refinement procedure, you retain the current Fourier of the solvent area and use it as the model. If there are errors in the protein model which may be reflected in the solvent density, will this procedure tend to "lock in" errors in the model, and give you an unrealistically low R-factor? Perhaps a constant solvent density would be less biased. Did you also fill the cavities inside the protein with density, as will happen automatically in the procedure, or did you assume they are empty and set them to zero?

RAGHAVAN: We are using a constant solvent density equal to the theoretical scattering value of $0.49 \text{ F/\AA}^3$. The Fourier of solvent area is used only to calculate the average density in the area before the solvent correction. Since we do not use this average value, but use the theoretical value of $0.49 \text{ F/\AA}^3$, to calculate the structure factors, the errors in the protein model do not affect the solvent structure factors. We are using van der Waals radii around every atom, and the cavities inside the protein are unfilled in this procedure.

260 BLANK

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.