Eur Biophys J (1987) 14:337–348

European Biophysics Journal

Structural disorder in proteins

A comparison of myoglobin and erythrocruorin

H. Hartmann[1], W. Steigemann[2], H. Reuscher[1], and F. Parak[1]*

[1] Institut für Physikalische Chemie der Universität Münster, Schloßplatz 4–7, D-4400 Münster, Federal Republic of Germany
[2] Max-Planck-Institut für Biochemie, Am Klopferspitz, D-8033 Martinsried, Federal Republic of Germany

Received August 5, 1986/Accepted September 11, 1986

Abstract. The refinement of X-ray structural data gives the mean square displacements, $\langle x^2 \rangle$, at each position in the protein molecule. In order to get information on the significance of such values different refinement methods have been compared. The metmyoglobin structure was determined at 300 K and $\langle x^2 \rangle$-values were obtained with the restrained refinement procedure in reciprocal space of Konnert and Hendrickson. A comparison with the results of Frauenfelder et al. was used for an error estimation. The inclusion of surface bound water increases the accuracy of the results but does not change the general picture. For erythrocruorin (CTT3) a refinement was performed in reciprocal space and compared with a refinement in real space performed earlier. The $\langle x^2 \rangle$-values obtained from both procedures are similar although the reciprocal space refinement gives results which are physically more reasonable.

A comparison of the disorder in myoglobin and erythrocruorin showed that the structural similarity results in a similarity in the disorder. Contacts of molecules in the crystal do not dominate the disorder although they locally influence $\langle x^2 \rangle$-values. CTT3 shows large disorder in the heme region in contrast to myoglobin. The differences in the rigidity of the *F*-helix can be correlated with the oxygen affinities supporting models for O_2 binding developed by Frauenfelder et al.

Key words: Structural disorder, metmyoglobin, erythrocruorin, protein

Introduction

Protein molecules have a well defined conformation which can be determined by X-ray structure analysis of a crystal. The accuracy of the average atomic coordinates obtained reaches 0.07 Å in some cases. Individual molecules of the crystal, however, have structures where the coordinates of the atoms differ remarkably from the average.

Artymiuk et al. (1979) and Frauenfelder et al. (1979) have shown that the refinement of the protein structure permits the assignment of individual isotropic atomic mean square displacements, $\langle x^2 \rangle$, to all non-hydrogen atoms. These values measure the deviation from the average position for each atom. Regions of the molecule which have small $\langle x^2 \rangle$-values are structurally well defined while regions with large mean square displacements reflect high structural disorder. It has to be emphasized that X-ray structure analysis allows no differentiation between static and dynamic disorder (compare, for instance, Parak and Knapp 1984).

Artymiuk et al. (1979) compared the individual mean square displacements of hen and human lysozyme which crystallize in different space groups. The high agreement of the mean square displacements along the sequence strongly suggests that the disorder reflects molecular properties and is not determined by intermolecular contacts of the molecular surface in the crystal. Moreover, this correlation strongly underlines the significance of the individual $\langle x^2 \rangle$-values and confirms that the refinement procedures used yield values with physical relevance.

While different lysozymes have been investigated in great detail (see Sternberg et al. 1979; Blake et al. 1983) a similar investigation has not been performed up to now for myoglobin. Since structural disorder and structural fluctuations of myoglobin have been investigated by a large number of techniques such as CO flash photolysis (Austin et al. 1975; Beece et al. 1980; Doster et al. 1982), Mössbauer spectroscopy (Parak and Formanek 1971; Parak et al. 1982; Bauminger et al. 1983; Parak and Knapp 1984), RSMR (*R*ayleigh *S*cattering of *M*öss-

* To whom offprint requests should be sent

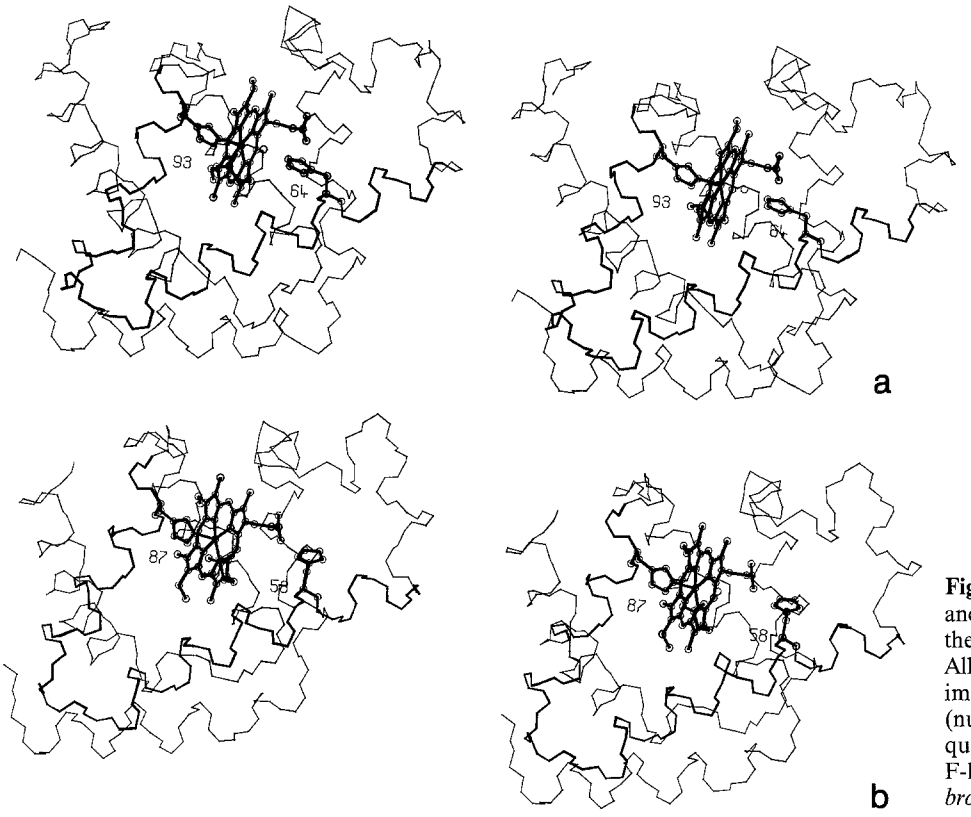

Fig. 1. Stereo picture of Mb (**a**) and CTT3 (**b**). The positions of the backbone atoms are connected. All side groups except the proximal His 93 and the distal His 64 (numbers from the myoglobin sequence) are omitted. Heme group, F-helix and E-helix are drawn with *broader lines*

bauer *R*adiation) (Krupyanskii et al. 1982) or low temperature X-ray structure analysis (Hartmann et al. 1982) a comparison of the mean square displacements of myoglobin with similar proteins seems highly desirable. For this purpose the hemoglobin of the insect larva of chironomus thummi thummi (abbreviated CTT3) is very suitable. This monomeric hemoglobin consists of 136 residues (sperm whale myoglobin (abbreviated Mb): 153 residues). Only 28 residues of these two proteins are identical. Nevertheless the structure of CTT3, which has been determined with high accuracy, is very similar to that of myoglobin (Steigemann and Weber 1979). This is shown in Fig. 1. The two proteins crystallize in different space groups (Mb: $P2_1$, CTT3: $P3_2$). Therefore, as in the case of the two lysozymes, the intermolecular contacts should be different in the crystals.

In this paper we describe the refinement of the Mb and CTT3 structures with the restrained least-squares refinement procedure of Konnert and Hendrickson (Konnert 1976; Konnert and Hendrickson 1980) and compare the mean square displacements of the two proteins. In the case of Mb we collected new reflection data up to a resolution of 1.5 Å. Therefore, it is possible to compare the present results with the results obtained by Frauenfelder et al. (1979). This comparison is helpful for an estimation of the physical significance of the $\langle x^2 \rangle$-values. For the refinement of CTT3 we used the experimental results of Steigemann and Weber (1979) who have already refined this structure in real space.

Crystal preparation and data collection

Crystals of sperm whale metmyoglobin (space group $P2_1$) were grown from ammonium sulfate solution with phosphate buffer at pH 6.1 according to the method of Kendrew et al. (1956). The crystal used for the data collection measured $0.3 \times 0.4 \times 0.7$ mm. Cell dimensions ($a = 64.52 \pm 0.03$ Å, $b = 30.90 \pm 0.01$ Å, $c = 34.85 \pm 0.02$ Å, $\beta = 105.80 \pm 0.02°$) were refined with the program IDXREF (Nyborg and Wonacott 1977) using 1,200 partially recorded reflections from 52 stills.

The intensity measurements were carried out photographically at 300 K on a modified precession camera using the screenless oscillation method with a cylindrical film cassette having a radius of 57.4 mm. This geometry restricts the resolution to 1.2 Å perpendicularly and 2 Å along the rotation axis. CuK_α-radiation from a 1.5 kW X-ray tube, filtered by a graphite monochromator, was used. The focus to crystal distance was 280 mm. The individual oscillation angle of each photograph was between 2° and 3.5° and the total measurement

covered a range of 180°. Between successive exposures there was an overlap of 0.5°. The total exposure time of the crystal was about 680 h. We used 3 films in one pack. These films were evaluated using the programs of Jones et al. (1977). Only fully recorded reflections were taken into account. With the program system PROTEIN (Steigemann 1974) the different exposures were brought to a common scale using all measurements of each reflection. Finally we got 56,864 observations above the 2σ significance level which were reduced to $N = 15{,}903$ independent reflections with an R_{merge} value of 0.052.

R_{merge} is defined as:

$$R_{\text{merge}} = \frac{\sum_{j=1}^{N} \sum_{i=1}^{m_j} |\langle I \rangle_j - I_i|}{\sum_{j=1}^{N} m_j \langle I \rangle_j}, \quad (1)$$

where $\langle I \rangle_j$ is the averaged intensity of an independent reflection, j, with m_j individually measured intensities, I_i. The sum in the denominator includes all intensities in the data set. The reduced data set corresponds to 90% of the possible reflections to a resolution of 2 Å and to 74% up to a resolution of 1.5 Å. The decrease in the measured fraction is partly produced by the limitation of the resolution to 2 Å along the rotation axis.

A linear and exponential scaling of all photographs with $F_{\text{scal}} = F_{\text{unscal}}\, S \exp(-B \sin^2\Theta/\lambda^2)$ and additionally of five exposures which had been repeated every 100 to 150 h showed an artifical increase of the overall B-value of 0.3 Å^2 during the data collection. This increase must be attributed to radiation damage.

In contrast to these measurements the data for the structure of CTT3 (Steigemann and Weber 1979) and for the metmyoglobin structure from Frauenfelder et al. (1979) were collected on a diffractometer. The data set of aquomet CTT3 consists of 24,541 observations and 23,614 independent reflections to a resolution of 1.4 Å with an R_{merge} value of 0.055.

Procedure of the refinement

The refinement of the myoglobin and CTT3 structures repeated here was carried out in reciprocal space using the restrained least-squares method of Konnert and Hendrickson (1980).

For the refinement of the *myoglobin structure* we started with the refined coordinates from Frauenfelder et al. (1979) and an overall B-value of 11.0 Å^2. At the beginning only reflections in the resolution range from 7.0 to 2.0 Å were included. The low order reflections were not taken into account because of the unknown structure of the disordered solvent. The high order reflections were excluded during the first few cycles to get a greater convergence range for the refinement procedure. Since even at 7.0 Å the F_c-values calculated from the myoglobin model are significantly greater than the observed F_0 values this limit was reduced to 5.5 Å in the course of the refinement. In addition, high order reflections up to 1.5 Å resolution were gradually included after a few cycles. From cycle 5 on, individual isotropic mean square displacements were refined for the 1,261 protein atoms. After cycle 24, at a standard R of 0.203, we included 121 water molecules and 2 sulfate ions with full occupancy and a B-value of 20.0 Å^2 (coordinates were taken from the metmyoglobin structures reported from Takano (1977) and Frauenfelder et al. (1979)). For the sulfate ions only the coordinates of the sulfur atoms were available.

The solvent molecules were allowed to move freely. During the next 10 cycles individual mean square displacements, but no occupancies, were refined for the solvent molecules. These refinement cycles lowered the standard R-value to 0.188. For a number of water molecules the B-values increased to 50 Å^2 or more. A few made bad contacts ($d < 2.5$ Å) with protein molecules. For this reason, 23 water molecules were discarded. A difference Fourier map calculated at this stage was already rather clean in the protein region – with a few exceptions. The most prominant peak in the protein region, with a height of 0.48 e/Å^3, occurred at the side chain of Leu E4 (61). In comparison, a normal C-atom with a B-value of 12.0 Å^2 which had been ignored in the calculation of F_c, yielded a peak of height 1.15 e/Å^3. Inspection of a $(2F_0 - F_c)$ map on a graphic display showed that the side chain of Leu E4 (61) can adopt two or three positions separated by rotations of 120° around the $C_\beta - C_\gamma$ bond. The alternate positions of Leu E4 were built into the map with the program FRODO (Jones 1978; Jones and Liljas 1984). In the solvent region there were peaks up to a height of 0.7 e/Å^3. With the sulfur atoms included in the refinement, four positive peaks at a distance between 1.4 and 1.8 Å occurred in a tetrahedral arrangement around the sulfur atoms. Unambigously these are the oxygen atoms of the sulfate ion. The stereochemical restraints of the sulfate group were applied to these atoms and we included them in the further refinement. In addition, 62 positive peaks with peak heights greater 0.35 e/Å^3 were attributed to water molecules and added to the model coordinates. The final model consists of 1,261 protein

atoms, 160 water molecules and 2 sulfate ions. As a consequence of these changes the R-value was reduced by 0.012 in the following 10 cycles of restrained refinement. The resulting R-value is 0.176 for 15,500 reflections in the resolution range from 5.5 to 1.5 Å. Covalent bond distances have a root-mean-square deviation of 0.015 Å from the target values (ideal group data supplied by Konnert and Hendrickson 1980) and bond angles a rms deviation of 4.8°. In comparison, Frauenfelder et al. (1979) reached a standard R-value of 0.18 with a root-mean-square deviation from ideality of 0.05 Å for covalent distances.

The difference in the B-values of two atoms bound to each other has been restrained to an rms-value of 1.2 Å^2 for the backbone and 1.7 Å^2 for the side chain atoms. For small molecular structures, which can be refined without restraints, the corresponding value is in most cases less than 1.0 Å^2. Therefore, the restraints we used should not lead to a large systematic error in the individual mean square displacements. This was tested by using tighter restraints for the thermal parameters, this yielded rms differences of 0.9 Å^2 and 1.3 Å^2 for backbone and side chain atoms, respectively. The alterations of the atomic mean square displacements $\langle x^2 \rangle$ along side chains became smoother but the mean values, $\langle x^2 \rangle_{\text{SIDE}}$ and especially $\langle x^2 \rangle_{\text{NCC}}$, calculated with all side chain atoms of a residue and with the atoms N, C_α, C, respectively were affected only a little. The mean change of $\langle x^2 \rangle_{\text{NCC}}$ for all residues arising from the tighter restraints was about 4%.

The *structure of CTT3* has been determined by Huber et al. (1970, 1971). Steigemann and Weber (1979) and Weber et al. (1978) refined the structure in different ligand states at a resolution of 1.4 Å. For the aquomet form 20 cycles of real space refinement with fixed standard geometry and manual model corrections were performed, resulting in a standard R of 0.183 for 18,100 significant reflections. Individual isotropic temperature factors were determined from the refined atomic radii.

The starting point for our refinement of CTT3 in reciprocal space was the refined model of Steigemann and Weber (1979), with a few exceptions concerning the occupancy factors. The version of the Hendrickson-Konnert program which we used at the beginning of this work accepted only full occupancy for the atoms. Therefore, the occupancy factors of the 94 localized water molecules had to be set to 1.0. Furthermore we had to choose between the two alternative positions which had been found by Steigemann and Weber for two amino acids. The starting R-value was therefore 0.218 instead of 0.183. From the beginning, individual isotropic temperature factors were included in the refinement. After one cycle of refinement, the R-value had already dropped to 0.191. After nine cycles the refinement yielded a standard R of 0.167. At this stage the refinement was continued with a newer version of the Hendrickson-Konnert program which became available to us. Now individual mean square displacements and occupancies were refined for the water molecules and the two alternate positions of Ser G14 (106) and Met H17 (131). After an additional 10 cycles the R-value has dropped to 0.154 for about 22,000 reflections in the resolution range from 5.5 to 1.4 Å.

This value has to be compared with the R-value of 0.183 for 18,100 reflections from the real space refinement. Including the remaining 1,300, mostly weak, reflections the R-value is 0.162. These are reflections with intensities less than 2σ but which had been included with $F = 2\sigma$. The geometry of the refined model is in good agreement with the group data. For example the rms deviation from bond ideality is 0.009 Å. R-value and geometry compare well with the values obtained for the refinement of the oxy- and deoxy structures of CTT3 (Steigemann and Weber 1982; Steigemann 1981) using the method of combined X-ray and energy refinement (Jack and Levitt 1978).

In a restrained refinement one has some freedom to choose the relative weighting of reflection data and restraints in a certain region. Of course this does not only affect the R-value but, in principle, the refined parameters also. Because we were mostly interested in the atomic mean square displacements, we have taken special care not to make them systematically incorrect. Of course one would expect that restraints on the thermal parameters themselves have the greatest influence. This has been checked during the refinement of the Mb structure (see above) but the mean square displacement of an atom can, of course, be incorrect if the atom is misplaced. Minor changes of the model coordinates are affected by the tightness of the stereochemical restraints.

To test the influence of the stereochemical restraints on the thermal parameters we relaxed these restraints by a factor of four. The restraints for the thermal parameters were, however, not relaxed. An additional eight cycles of refinement were performed. The R-value dropped by 0.01 and the deviation of the model from ideal geometry became rather high, e.g. rms = 0.059 Å for covalent bond distances. The individual and the overall B values did not change, however. According to this test the tightness of the geometrical restraints are in a reasonable range and therefore the precise atomic coordinates have a neglegible impact on the B-values.

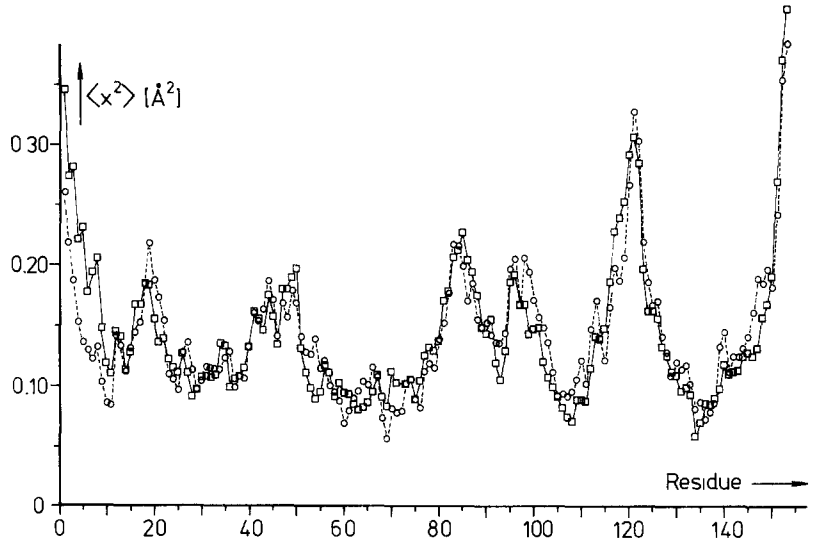

Fig. 2. Mean square displacements averaged over the N–C_α–C atoms of the backbone as a function of the residue number. *Squares:* present refinement of Mb; *circles:* data from Frauenfelder et al. (1979)

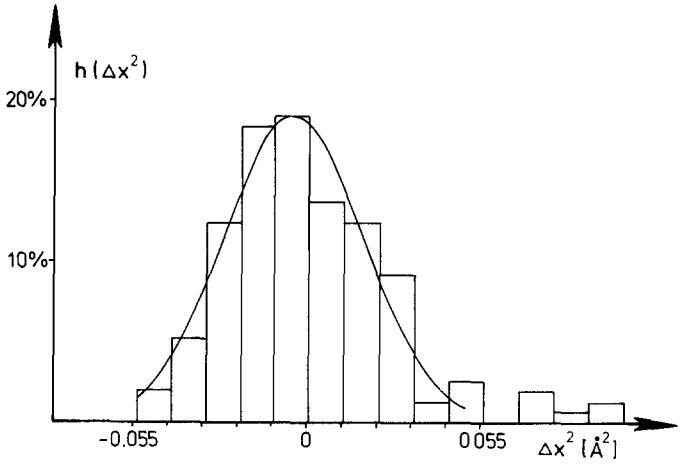

Fig. 3. Probability $h(\Delta x^2)$ of $\Delta x^2 = \langle x_1^2 \rangle - \langle x_2^2 \rangle$ in steps of 0.011 Å^2. The indices 1 and 2 refer to the present Mb measurement and the data from Frauenfelder et al. (1979), respectively. The *solid line* gives a Gaussian distribution fitted to the histogram. 100% corresponds to 153 residues

Results and discussion

Comparison of different investigations on metmyoglobin

Figure 2 shows the mean square displacement of the backbone of Mb. Each point gives the average value $\langle x^2 \rangle_{NCC}$ over the three backbone atoms –N–C_α–C. Our results are compared with the refinement of Frauenfelder et al. (1979). It is obvious that the two refinements give essentially the same disorder along the chain.

To get an assessment of the errors for the atomic mean square displacements we compared our data with the data of Frauenfelder et al. (1979) and calculated rms-values of the mean square displacements. The rms-value is defined as:

$$\text{rms} = \left[\sum_{i=1}^{N} (\langle x_1^2 \rangle_i - \langle x_2^2 \rangle_i)^2 / N \right]^{-1/2}, \qquad (2)$$

where $\langle x_1^2 \rangle_i$ and $\langle x_2^2 \rangle_i$ are the mean square displacements of the amino acid i of our data and the data of Frauenfelder et al. (1979), respectively. For the averages over the backbone atoms, $\langle x^2 \rangle_{NCC}$, we get an rms-value of 0.027 Å^2 using all 153 residues and 0.023 Å^2 if the first 5 residues of the N-thermal end are excluded. The N-terminal end is highly disordered (compare also Takano 1977). We observe large differences of up to 1 Å in the coordinates for a few atoms of these residues. Figure 3 shows the distribution of the differences of $\langle x^2 \rangle_{NCC}$-values between the two data sets. To a good approximation this distribution is Gaussian with a width of about two times the rms-value calculated according to Eq. (2). Consequently the rms-value should be a gross estimate of the standard deviation of the mean square displacements. This error estimation contains the assumption that errors on one residue do not influence errors on another residue. Moreover, only errors of $\langle x^2 \rangle_{NCC}$ independent of the overall B-value, B_{OV}, are considered. The difference of the overall B-values between the two models contributes to the rms-value, but the contribution of the error of B_{OV} to the true standard deviation of $\langle x^2 \rangle_{NCC}$ may be

Table 1. Comparison of the results of Frauenfelder et al. (1979) and the present data for met Mb. The mean square displacements are divided into 5 regions, where N gives the number of residues in a region and $\langle x^2 \rangle_r$ the average value of the atoms in these regions. The rms-value is defined by Eq. (2) and measures the discrepancy between the two structure determinations

$\langle x^2 \rangle$-region [Å^2]	0.05 – 0.10	0.10 – 0.13	0.13 – 0.16	0.16 – 0.20	0.20
N	29	48	30	29	17
$\langle x^2 \rangle_r$ [Å^2]	0.088	0.114	0.145	0.176	0.259
rms [Å^2]	0.0176	0.0190	0.0239	0.0347	0.0426

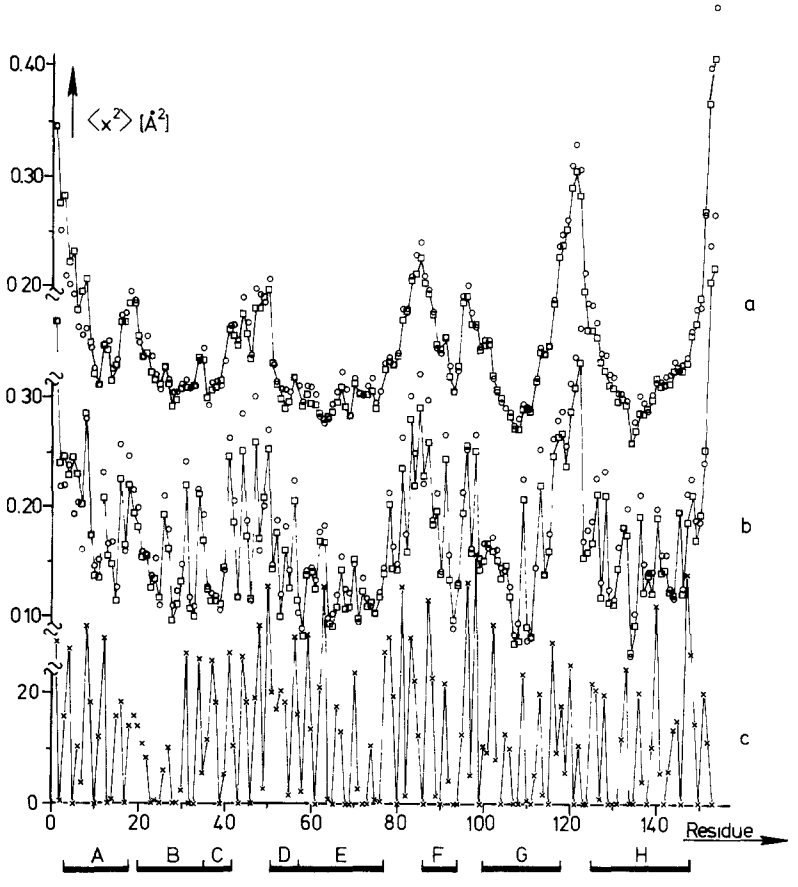

Fig. 4. Mean square displacements, $\langle x^2 \rangle$, from the backbone atoms (**a**) and the sidechain atoms (**b**) obtained from refinements taking into account 160 water molecules (*squares*) and no water (*circles*), respectively. *Crosses* in **c** show the accessible surfaces of the sidechains to the water in Å^2

underestimated, because there are only two independent measurements of B_{OV} available. Calculating rms-values from Eq. (2) in shells of $\langle x^2 \rangle_{NCC}$ shows that the errors are greater for large $\langle x^2 \rangle_{NCC}$-values. This is summarized in Table 1. The overall B-value is 12.5 Å^2 in our case compared to 12.1 Å^2 obtained by Frauenfelder et al. (1979).

Depending on the restraints a periodicity of $\langle x^2 \rangle$ at about 3 residues is generally seen in Fig. 2. Note, that in an α-helix one turn contains 3.6 residues. This reflects the higher mobility at the surface of the molecule (compare also Fig. 4c).

The influence of localized water on mean square displacements

Some refinements of mean square displacements in proteins were performed without considering the surface bound water molecules visible in X-ray structure analysis (see, for example, Hartmann et al. 1982). The question that arises is how strongly such results are influenced by the water. Figure 4 gives mean square displacements where the refinement has been performed without water and with 160 water molecules per myoglobin molecule. It is obvious that the influence on the mean square displacements is rather small. Refinements without water give slightly larger $\langle x^2 \rangle$-values with somewhat more structure, especially for the side chains. The fit tries to compensate for the absence of bound water molecules by giving the protein surface more flexibility.

Figure 4c gives the accessibility of the sidechain atoms to water calculated according to the method of Lee and Richards (1971). For the side chains this accessibility to water correlates strongly with the

mean square displacements. The mobility of a protein is strongly influenced by its surrounding medium. This fact is the basis for explaining protein dynamics inferred from Mössbauer experiments and X-ray structure analysis in a consistent picture (Parak and Knapp 1984).

Different refinements of erythrocruorin

In Fig. 5 we compare the mean square displacements of the backbone of CTT3 derived from the refinement according to the method of Hendrickson and Konnert with the values which have been obtained by Steigemann and Weber (1979). In contrast to our new refinement of the met CTT3-structure, Steigemann and Weber refined individual mean square displacements for each atom without applying restraints to the thermal parameters (which is very difficult to do in real space). Such a procedure is only possible with very high resolution data. It does not introduce systematic errors in the individual $\langle x^2 \rangle$-values, but the statistical errors are high, even at high resolution. Figure 5 shows that regions of major disorder are well reproduced by both methods of refinement. The real space refinement shows more fluctuations along the chain. Restraining the difference of the mean square displacements of bounded atoms to small values has a physical justification and reduces the statistical errors. It is possible, however, that the $\langle x^2 \rangle$-value of an atom become systematically wrong if the disorder changes rapidly from one atom to the other. This may only be relevant in some side chains. We believe that the errors introduced by weak restraints are small compared with the statistical errors which one gets without restraints (compare the preceding section). Therefore, the $\langle x^2 \rangle$-values of the atoms derived in this way seem to be more meaningful.

Comparison of erythrocruorin and myoglobin

Table 2 gives mean square displacements for Mb and CTT using the refinement procedure of Hendrickson and Konnert. The average structural disorder is rather similar in Mb and CTT3. The structure of CTT3 is only slightly better defined, especially for the backbone average. Figure 6 compares the mean square displacements of CTT3 and Mb. In both proteins there is a background of about 0.1 Å^2 common to all atoms in the crystal. As we know from the temperature dependence of the $\langle x^2 \rangle$-values in myoglobin (Hartmann et al. 1982; Parak et al., to be published) this background can be reduced by cooling. Therefore, it can only partly be attributed to static lattice disorder reflecting poor crystallization of proteins. In part it may stem from translational or rotational motions of the whole molecule or of large segments. Segmental motions may be partly unspecific, yielding equal $\langle x^2 \rangle$-values for different regions in the molecule. The background of $\langle x^2 \rangle = 0.1$ Å^2 in Mb and CTT3 can be compared with estimated background values of about 0.13 Å^2 in lysozyme (Artymiuk et al. 1979) and about 0.11 Å^2 in trypsinogen (Walter et al. 1982). These values are astonishingly similar.

In addition to the background there are strong variations of $\langle x^2 \rangle$ along the sequence as already discussed by Frauenfelder et al. (1979) for Mb. The mean square displacements of CTT3 and Mb are rather similar. In most cases the helical regions are relatively well defined in both proteins. Differences in $\langle x^2 \rangle$ between CTT3 and Mb stem partly from missing residues in the sequence of CTT3 in non-helical regions. This is clearly seen in the AB and the GH loops. The non-helical region GH in CTT3 is shorter than in Mb, thereby reducing the possibility of disorder.

Major differences between the mean square displacements of Mb and CTT3 arise in the helical

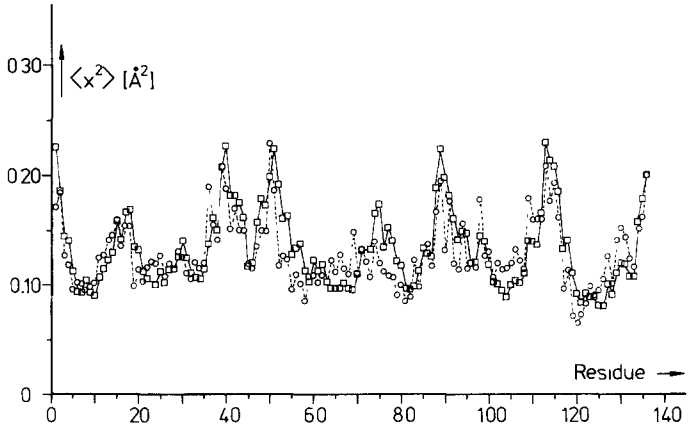

Fig. 5. Mean square displacements, $\langle x^2 \rangle$, of the backbone atoms in CTT3. *Squares:* refinement according to Konnert and Hendrickson (1980) in the reciprocal space; *circles:* real space refinement by Steigemann and Weber (1979)

Table 2. Comparison of mean square displacements in Mb and CTT3. The values in rows II to IV give average values of side-chains in contact with the heme. Characterizing different parts of the heme by the pyrroles A to D (1 to 4 is the numbering used e.g. by Antonini E. and Brunori M.: Hemoglobin and myoglobin in their reaction with ligands, North Holland (1971)) contacts of pyrroles and side-chains are marked in the column Cont. Row V gives $\langle x^2 \rangle$-values averaged over the values in rows II to IV. In row VI 4N means averaging over the N atoms of the four pyrroles. Pyrrole 3 (A) gives $\langle x^2 \rangle$-values averaged over the C-atoms of pyrrole 3 (A). The symbol CO 3 (A) indicates $\langle x^2 \rangle$-values averaged over all atoms connected with pyrrole 3 (A) with the exception of the N and the C forming the connection to the neighbouring pyrrole. Cont A (AS) gives the average of all values marked with A in rows II to IV. Cont A (atoms) gives the average $\langle x^2 \rangle$-value of all individual atoms in contact with part A of the heme. In all cases H-atoms are excluded

	Mb (met)			CTT3 (met)		
	Average over	Cont	$\langle \bar{x}^2 \rangle$ [Å^2]	Average over	Cont	$\langle x^2 \rangle$ [Å^2]
I	molecule		0.158	molecule		0.154
	backbone		0.144	backbone		0.133
	sidechains		0.176	sidechains		0.173
	heme		0.096	heme		0.172
II	Proximal residues			Proximal residues		
	Leu 89	D	0.197	Phe 83 (89)	A	0.131
	Ser 92	C	0.134	Ser 86 (92)	B	0.149
	His 93	ABCD	0.097	His 87 (93)	A CD	0.099
	His 97	BC	0.159	Arg 90 (97)	BC	0.191
	Ile 99	A	0.143	Val 92 (99)	D	0.176
	Tyr 103	A	0.152			
	Leu 104	A	0.135			
	Ile 107	D	0.076			
	Phe 138	D	0.139			
III	Distal residues			Distal residues		
	Thr 39	A	0.111	Ile 34 (39)	D	0.116
	Phe 43	AB	0.117	Phe 38 (43)	C	0.147
	Arg 45	B	0.157			
	His 64	C	0.094	His 58 (64)	C	0.194
	Thr 67	C	0.143	Arg 61 (67)	B	0.195
	Val 68	D	0.107	Ile 62 (68)	AB	0.158
				Phe 66 (72)	A	0.181
				Phe 100 (107)	A D	0.186
IV	"Crossing" residues			"Crossing" residues		
	Lys 42	B	0.186	Lys 37 (42)	CD	0.280
	Ala 71	CD	0.098	Phe 65 (71)	B	0.136
				Gln 96 (103)	D	0.155
V	II – IV		0.132	II – IV		0.166
VI	Fe		0.096	Fe		0.139
	4N		0.087	4N		0.141
	pyrrole 3 (A)		0.061	pyrrole 3 (A)		0.156
	pyrrole 4 (B)		0.094	pyrrole 4 (B)		0.136
	pyrrole 1 (C)		0.084	pyrrole 1 (C)		0.154
	pyrrole 2 (D)		0.070	pyrrole 2 (D)		0.145
	CO 3 (A)		0.071	CO 3 (A)		0.174
	CO 4 (B)		0.122	CO 4 (B)		0.158
	CO 1 (C)		0.119	CO 1 (C)		0.211
	CO 2 (D)		0.073	CO 2 (D)		0.157
	cont A (AS)		0.126	cont A (AS)		0.151
	cont B (AS)		0.143	cont B (AS)		0.166
	cont C (AS)		0.121	cont C (AS)		0.182
	cont D (AS)		0.119	cont D (AS)		0.169
	cont A (atoms)		0.122	cont A (atoms)		0.144
	cont B (atoms)		0.151	cont B (atoms)		0.151
	cont C (atoms)		0.123	cont C (atoms)		0.163
	cont D (atoms)		0.113	cont D (atoms)		0.151

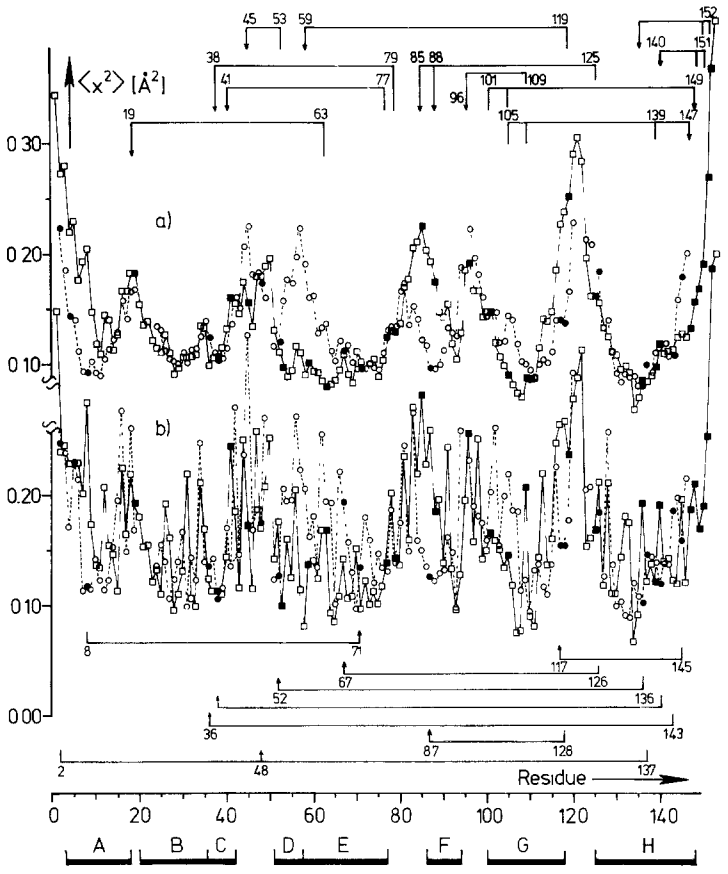

Fig. 6. Comparison of the mean square displacements, $\langle x^2 \rangle$, of Mb (*squares*) and CTT3 (*circles*) for the backbone (**a**) and the sidechains (**b**), respectively. Residue number according to the sequence of Mb. At the position of missing residues in CTT3 the *dotted line* connecting the circles is interrupted. Contacts between molecules in the crystal (*full symbols*) are indicated by lines. *Above* **a**: contacts in Mb, *below* **b**: contacts in CTT3

regions D and E, in the interhelical region EF, and at the beginning of the F helix. In total the helices E and F show greater mean square displacements in CTT3 than in Mb. Before we try to compare this with different functional properties we have to consider the influence of the contacts of the molecules in the crystal. Table 3 gives the contacts with a distance smaller than 3.5 Å. The distances d are sometimes unrealistically short. This reflects the fact that the steric restrictions applied to the coordinates of the atoms in one molecule during the refinement are not used for intermolecular distances. In Fig. 6 the contacts are marked. Inspecting the backbone $\langle x^2 \rangle$-values and starting from the *N*-terminal side CTT3 makes three contacts between residue 1 and 10 while this region is not influenced by crystal packing in Mb. This may explain the smaller $\langle x^2 \rangle$-values in CTT3 in this region. The contact at residue 19 in Mb falls on a maximum of $\langle x^2 \rangle$. No contact occurs in this region in CTT3, nevertheless the $\langle x^2 \rangle$-values also show a maximum. In the CD loop the $\langle x^2 \rangle$-values in Mb are smaller than in CTT3 which may be due to a contact at residue 45. There is a dramatic difference in $\langle x^2 \rangle$ between CTT3 and Mb in the D and E helices. Mb with the smaller $\langle x^2 \rangle$-values has more contacts in this region. That this is not necessarily the explanation for the different disorder is shown by an inspection of the EF loop. The $\langle x^2 \rangle$-values are very much higher in Mb even though there are more contacts as in CTT3. Considering contacts in Mb at residues 19, 85, 109 and in the *C*-terminus proves that the appearance of disorder is determined by the molecular structure and not by the crystal packing. The average $\langle x^2 \rangle$-values of the backbone atoms in the contact regions are 0.177 Å^2 in Mb and 0.147 Å^2 in CTT3 and the corresponding values for the side chains 0.183 Å^2 and 0.137 Å^2. The contact side chain average in CTT3 is smaller than the value averaged over all residues, reflecting an influence of the intermolecular contacts. It has, however, to be considered that the contacts occur on the protein surface where the $\langle x^2 \rangle$-values are, in general, larger. From the considerations above one can conclude that the disorder obtained from measurements on crystals is mainly determined by molecular properties but contacts between molecules do influence the picture. This is of importance for comparisons with other experimental and theoretical investigations. In some cases intra- and intermolecular coupling cannot be separated.

One may now try to correlate differences of the disorder in Mb and CTT3 with functional proper-

Table 3. Contacts between protein molecules in crystals. Metmyoglobin is compared with chironomus CTT3. The distance of atoms calculated from the structure is given by d in Angströms. All contacts with $d \leqq 3.5$ Å are included. For CTT3 the sequence is numbered from 1 to 136 and in brackets from 1 to 153 to allow comparison with Mb. Numbers below the amino acid give the $\langle x^2 \rangle$-value of the backbone or the side-chain, respectively

Mb						CTT3					
		d [Å]						d [Å]			
						Asp 4 (5)					
Ala 19	C	3.4	Lys 63	CE							
19	O	2.4	(0.168)								
(0.138)											
19	O	2.6	63	NZ							
Glu 38	OEL	2.8	Lys 79	NZ		Asp 31 (36)	OD1	3.5	Phe 133 (143)	CD2	
(0.114)			(0.144)			(0.137)			(0.188)		
						Ser 33 (38)	CB	3.5	Gly 130 (140)	O	
						(0.106)			(0.120)		
Glu 41	CG	3.1	Lys 77	CD		Phe 65 (71)	CD2	3.5	Ser 7 (8)	O	
41	OE2	3.1	77	NZ		(0.136)			(0.093)		
(0.246)			(0.139)								
Arg 45	O	3.3	Ala 53	CB		Gly 43 (48)	O	2.7	Leu 1 (2)	N	
(0.157)			(0.100)			(0.175)			(0.226)		
						Leu 46 (52)	N	2.9	Asp 126 (136)	OD2	
						(0.120)	CB	3.3	(0.104)		
Glu 59	OE1	3.4	His 119	CE1		Arg 61 (67)	NE	2.9	Gly 116 (126)	O	
(0.138)			(0.239)			(0.195)	CZ	3.3	(0.185)		
							NH2	2.8			
Glu 85	O	3.3				Asn 81 (87)	ND2	3.0	His 111 (118)	O	
(0.227)						(0.126)			(0.137)		
85	CG	2.5	Ala 125	CB							
85	CD	3.3	(0.169)								
85	OE2	3.3									
(0.291)											
Pro 88	CG	3.1	125	O							
(0.185)			(0.162)								
Lys 96	NZ	2.9	Glu 109	OE1							
96	CE	3.2	(0.308)								
(0.257)											
Glu 136	O	3.4	Gly 150	CA							
(0.086)			(0.198)								
136	OE2	3.0	Gln 152	O							
(0.193)			(0.370)								
Lys 140	CA	3.3	Leu 149	O							
(0.118)			(0.168)								
140	CG	3.5	Tyr 151	OH							
(0.192)			(0.254)								
Lys 147	O	3.4	139	NE							
(0.132)		2.9	139	NH2							
Lys 147	CE	3.4	109	OE2							
	NZ	3.1									
(0.188)											
147	NZ	2.7	Glu 105	OE1		Lys 135 (145)	NZ	3.0	Ala 110 (117)	O	
147	CD	3.4	(0.147)			(0.160)			(0.140)		
Glu 148	CA	3.4	Glu 105	OE2							
(0.157)											
148	O	3.3	Arg 139	NE							
		3.4	139	CD							
			(0.122)								
148	O	3.2	Ile 101	CG2							
			(0.167)								

ties. An O_2 molecule taking a short path from the outside to the heme iron probably passes the E helix of the molecule. According to Case and Karplus (1979) there are two major paths for oxygen binding in Mb; the potential barriers hindering the passage are formed by the residues His E7 (64), Thr E10 (67) and Val E11 (68) for the classical path and Leu E4 (61) and Phe B14 (33) for another path. According to Fig. 6 these regions are less ordered in CTT3 than in Mb. Therefore, the binding path for the oxygen should be less hindered in CTT3. However, a comparison of the oxygen affinities of the two proteins does not show such a correlation. At 25 °C and pH 7.0 the oxygen affinities measured by p_{50}-values are 1.3 mm and 0.9 mm Hg in CTT3 and Mb, respectively. The tendency becomes even more pronounced on going to lower pH-values where crystallization is performed. Nevertheless, the results of Ringe et al. (1984) show that the residues His E7 (64), Arg CD3 (45) and Val E11 (68) can be forced a side by phenylhydrazine, marking a path for O_2. According to Austin et al. (1975) and Doster et al. (1982) there exists a preequilibrium of O_2 in the heme pocket. The oxygen affinity is mainly determined by the final step where the O_2 molecule in the heme pocket binds to the iron and not by the access of O_2 to the heme pocket. The binding of oxygen changes the structure, via the proximal histidine, F8 (93), the F-helix is moved. The flexibility of the F-helix can be measured by the $\langle x^2 \rangle$-values of the backbone averaged from the middle of the EF loop (residue 82) to the middle of the FG loop (residue 97). One obtains $\langle x^2 \rangle_F = 0.171\,\text{Å}^2$ and $\langle x^2 \rangle_F = 0.138\,\text{Å}^2$ for Mb and CTT3, respectively. This supports the picture of Frauenfelder. The protein with the higher flexibility of the F-chain can more easily go from the unliganded to the liganded conformation and thus bind O_2 more easily. This correlation is, however, no proof of this picture. It is obtained from data of met Mb and met CTT3, not using the deoxy and oxy structures. Moreover, the result could be influenced by intermolecular contacts in the crystals although one should remember that there are more contacts in Mb than in CTT3.

Finally we compare the disorder in the heme region. Table 2 gives the residues in contact with the heme together with the $\langle x^2 \rangle$-values of their side chains. As already mentioned the heme shows more disorder in CTT3 than in Mb. In Mb the disorder varies strongly over the heme. This can be seen from the $\langle x^2 \rangle$-values averaged over the C-atoms in the 4 pyrroles. The pyrroles II(D) and III(A) buried deepest in the molecule show rather small disorder. The pyrroles I(C) and IV(B) carrying the propionic acids are clearly more disordered. The $\langle x^2 \rangle$-value averaged over the few N-atoms bound to the Fe is between the values of the inner and the outer pyrroles while the iron itself has the largest value, close to the value of the proximal His F8 (93). In Table 2 the residues which come close to the different pyrroles are marked. Their averaged $\langle x^2 \rangle$-values correlate with the values of the pyrroles, although the $\langle x^2 \rangle$-values of the side chains are clearly larger. This discrepancy is diminished if we characterize the four parts A–D of the heme by the $\langle x^2 \rangle$-values obtained from the C-atoms of the pyrroles together with the other atoms linked with them, excluding only the N-atoms and those C-atoms which are between two pyrroles. Nevertheless, the disorder or flexibility of the heme is always smaller than that of the contacting groups.

The picture obtained in CTT3 is somewhat different. The average disorder of the heme is nearly identical with the average over the side chains in contact. The disorder varies only slightly for different heme regions. The relatively best defined part is the pyrrole B region carrying a propionic acid with contact to the molecular surface. In comparison to the protoporphyrin IX the disorder of the Fe is reduced by the ligation to the proximal His 87 (93). This also decreases the disorder of the four N-atoms of the heme bound to the Fe. In Mb just the opposite tendency was found. In comparison to Mb the heme is rotated by 180° about the $\alpha-\gamma$-meso axis in CTT3 (Huber et al. 1971); NMR experiments gave evidence for two types of CTT3 where the second species has the same heme orientation as Mb (La Mar et al. 1978). According to the NMR data the ratio of the different species is 60:40 and this would explain the high disorder. X-ray structure investigations (Steigemann and Weber 1979) find that the second component should be less than 25%, but even such a small contribution may explain the disorder. However, one could also explain the data by the assumption that, in crystals, only one CTT3 species exists. In this case the high disorder would suggest that the heme can be easily exchanged by a rotated heme since it does not fit so well into the protein moiety as in Mb.

Conclusions

The generally used restrained refinement procedure of Konnert and Hendrickson yields mean square displacements $\langle x^2 \rangle$, which allow a physical interpretation even if the water bound to the protein surface is not included. The errors of $\langle x^2 \rangle$ increase with increasing $\langle x^2 \rangle$. The similarity of the tertiary structure of Mb and CTT3 results in a similarity of the $\langle x^2 \rangle$-values within the molecule. Significant differences of $\langle x^2 \rangle$-values in Mb and CTT3 at certain parts

of the molecule can be correlated with the oxygen affinity although the present data set is not large enough for a final conclusion. Contacts of the molecules in the crystal influence but do not dominate the disorder in the molecule.

Acknowledgements. This work was supported by the Deutsche Forschungsgemeinschaft. We gratefully acknowledge that R. Huber allowed us to use facilities of his institute.

References

Artymiuk PI, Blake CCF, Grace DEP, Oatley SI, Phillips DC, Sternberg MJE (1979) Crystallographic studies of the dynamic properties of lysozyme. Nature 280: 563–568

Austin RH, Beeson KW, Eisenstein L, Frauenfelder H, Gunsalus IC (1975) Dynamics of ligand binding to myoglobin. Biochemistry 14: 5355–5372

Bauminger ER, Cohen SG, Nowik I, Ofer S, Yariv I (1983) Dynamics of heme iron in crystals of metmyoglobin and deoxymyoglobin. Proc Natl Acad Sci USA 80: 736–740

Beece D, Eisenstein L, Frauenfelder H, Good D, Marden MC, Reinisch L, Reynolds AH, Sorensen LB, Yue KT (1980) Solvent viscosity and protein dynamics. Biochemistry 15: 5147–5157

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

Case DA, Karplus M (1979) Dynamics of ligand binding to heme proteins. J Mol Biol 132: 343–368

Doster W, Beece D, Bowne SF, Dilorio EE, Eisenstein L, Frauenfelder H, Reinisch L, Shyamsunder E, Winterhalter KH, Yue KT (1982) Control and pH dependence of ligand binding in heme proteins. Biochemistry 21: 4831–4839

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

Hartmann H, Parak F, Steigemann W, Petsko GA, Ringe Ponzi D, Frauenfelder H (1982) Conformational substates in a protein: Structure and dynamics of metmyoglobin at 80 K. Proc Natl Acad Sci USA 79: 4967–4971

Huber R, Epp O, Formanek H (1970) Structures of deoxy- and carbonmonoxy erythrocruorin. J Mol Biol 52: 349–354

Huber R, Epp O, Steigemann W, Formanek H (1971) The atomic structure of erythrocruorin in the light of the chemical sequence and its comparison with myoglobin. Eur J Biochem 19: 42–50

Jack A, Levitt M (1978) Refinement of large structures of simultaneous minimization of energy and R-factor. Acta Crystallogr A 34: 931–935

Jones A, Schwager P, Bartels K (1977) In: Arnd UW, Wonacott AJ (eds) The rotation method in crystallography. North-Holland, Amsterdam, pp 105–117, 139–151

Jones TA (1978) A graphics model building and refinement system for macromolecules. J Appl Crystallogr 11: 268–272

Jones TA, Liljas L (1984) Crystallographic refinement of macromolecules having non-crystallographic symmetry. Acta Crystallogr A 40: 50–57

Kendrew JC, Parrish RG (1956) The crystal structure of myoglobin; III. Sperm-whale myoglobin. Proc R Soc A 238: 305–324

Konnert JH (1976) A restrained-parameter structure-factor least-squares refinement procedure for large asymmetric units. Acta Crystallogr A 32: 614–617

Konnert JH, Hendrickson WA (1980) A restrained-parameter thermal-factor refinement procedure. Acta Crystallogr A 36: 344–349

Krupyanskii YF, Parak F, Goldanskii VI, Mössbauer RL, Gaubmann EE, Engelmann H, Suzdalev IP (1982) Investigation of large intramolecular movements within metmyoglobin by Rayleigh scattering of Mössbauer radiation (RSMR). Z Naturforsch 37c: 57–62

La Mar GN, Overkamp M, Sick H, Gersonde K (1978) Proton nuclear magnetic resonance hyperfine shifts as indicators of tertiary structural changes accompanying the Bohr effect in monomeric insect hemoglobins. Biochemistry 17: 352–361

Lee B, Richards FM (1971) The interpretation of protein structures: Estimation of static accessibility. J Mol Biol 55: 379–400

Nyborg J, Wonacott AJ (1977) In: Arnd VW, Wonacott AJ (eds) The rotation method in crystallography. North-Holland, Amsterdam, pp 139–145

Parak F, Formanek H (1971) Untersuchung des Schwingungsanteils und des Kristallgitterfehleranteils des Temperaturfaktors in Myoglobin durch Vergleich von Mößbauerabsorptionsmessungen mit Röntgenstrukturdaten. Acta Crystallogr A 27: 573–578

Parak F, Knapp EW (1984) A consistent picture of protein dynamics. Proc Natl Acad Sci USA 81: 7088–7092

Parak F, Knapp EW, Kucheida D (1982) Protein dynamics. Mössbauer spectroscopy on deoxymyoglobin crystals. J Mol Biol 161: 177–194

Ringe D, Petsko GA, Kerr DE, Ortiz de Montellano PR (1984) Reaction of myoglobin with phenylhydrazine: A molecular doorstop. Biochemistry 23: 2–4

Steigemann W (1974) Die Entwicklung und Anwendung von Rechenverfahren und Rechenprogramme zur Strukturanalyse von Proteinen am Beispiel des Trypsin-Trypsininhibitor-Komplexes des freien Inhibitors und der L-Asparaginase. Dissertation TU München

Steigemann W (1981) Proceedings of the Daresbury Study Weekend: Refinement experiences using chain constraints in real space and energy restraints in reciprocal space, pp 40–46

Steigemann W, Weber E (1979) Structure of erythrocruorin in different ligand states refined at 1.4 Å resolution. J Mol Biol 127: 309–338

Steigemann W, Weber E (1982) Structure of deoxy- and oxy-erythrocruorin and related ligand states at 1.4 Å resolution. In: Chien Ho (ed) Hemoglobin and oxygen binding. Elsevier North-Holland, Amsterdam, pp 19–24

Sternberg MIE, Grace DEP, Phillips DC (1979) Dynamic information from protein crystallography. J Mol Biol 130: 231–253

Takano T (1977) Structure of myoglobin refined at 2.0 Å resolution. J Mol Biol 110: 537–568

Walter J, Steigemann W, Sing TP (1982) On the disordered activation domain in trypsinogen: Chemical labelling and low temperature crystallography. Acta Crystallogr B 38: 1462–1472

Weber E, Steigemann W, Jones T, Huber R (1978) The structure of oxy-erythrocruorin at 1.4 Å resolution. J Mol Biol 120: 327–336