Acta Cryst. (1982). B38, 784–792

Restrained Least-Squares Refinement of *Themiste dyscritum* Methydroxohemerythrin at 2·0 Å Resolution

BY R. E. STENKAMP, L. C. SIEKER AND L. H. JENSEN

Departments of Biological Structure and Biochemistry, University of Washington, Seattle, Washington 98195, *USA*

(*Received* 26 *May* 1981; *accepted* 2 *September* 1981)

Abstract

A model of methydroxohemerythrin has been refined at 2·0 Å resolution to clarify the structure of the binuclear iron complex in this oxygen-transport protein. An effort was made to minimize the crystallographic bias introduced by the assumed model by omitting the binuclear complex from the initial model and rebuilding it during the refinement. The procedure is extensively described, and some troublesome aspects of the refinement are indicated. The final R is 0·187 for 32254 reflections with d spacings in the range 10–2·0 Å.

Introduction

Over the past five years, two chemically different crystallographic models have been developed for the iron complex in methemerythrin and myohemerythrin, non-heme iron oxygen-transport proteins found in several marine invertebrates (Hendrickson, 1978; Stenkamp & Jensen, 1979). These models, the μ-oxo bridged complex and the confacial bioctahedron, are of interest to structural chemists, and the two models give an indication of the problems associated with extracting detailed molecular structures from protein crystallographic data insufficient to resolve bonded atoms.

The two models have been extensively refined in the belief that computational methods, especially crystallographic refinement, should be able to differentiate between the two and indicate which is consistent with the X-ray diffraction data. A problem with refinement, however, is the bias introduced into the calculation by the choice of the initial model. Calculated phases will tend to generate an image of the assumed model which, when superimposed on the image of the real structure, complicates the analysis of the resulting electron density maps.

In the case of methemerythrin and metmyohemerythrin, the least biased view of the iron complex is provided by the 2·8 Å resolution electron density map for *Themiste dyscritum* hemerythrin (Stenkamp, Sieker, Jensen & McQueen, 1978) where the phases were derived from a single, heavy-atom, isomorphous derivative of the native crystal. However, since 2·8 Å resolution is not sufficient to show the detailed structure of the complex, computational techniques have been used to extend the resolution of the analysis, first to 2·5 Å and then to 2·0 Å resolution.

We describe here a conservative application of a restrained least-squares program (Konnert, 1976; Hendrickson & Konnert, 1980) in refining methemerythrin from *T. dyscritum*, beginning with a model omitting the atoms closely associated with the iron complex. Atoms and groups of atoms in the complex were added when difference electron density maps clearly indicated where they should be placed. The model of the iron complex was carefully developed and should be relatively free of bias from the input model. A detailed description of the molecular structure will appear elsewhere.

Molecular species and crystal form

Previous structural reports of methemerythrin from this laboratory have referred to the molecular species as the metaquo form, but spectroscopic studies (Dunn, Addison, Bruce, Loehr & Loehr, 1977) indicate that the bound ligand is possibly a hydroxide ion. Thus, the form of methemerythrin we have studied crystallographically is presently designated methydroxo- rather than metaquohemerythrin.

The crystals of methydroxohemerythrin used in this study have been described earlier (Loehr, Meyerhoff, Sieker & Jensen, 1975). The space group is $P4$, $a = b = 86{\cdot}58\ (4)$, $c = 80{\cdot}84\ (4)$ Å, with two octameric molecules of D_4 symmetry per unit cell. The asymmetric unit comprises four subunits, two in one octamer (octamer I) with its fourfold axis coincident with the crystallographic fourfold axis at $0,0,z$ and two in the other octamer (octamer II) with its fourfold axis coincident with the crystallographic fourfold axis at $\frac{1}{2},\frac{1}{2},z$ (Stenkamp, Sieker, Jensen & Loehr, 1976). The second octamer is translated 0·5 in z and rotated 20° about z with respect to the first. The subunits are

0567-7408/82/030784-09$01.00

labeled I*A*, I*B*, II*A*, and II*B* with I*A* and I*B* and the fourfold-related subunits forming octamer I. Similarly, II*A* and II*B* along with the fourfold-related subunits make up octamer II. The four molecular twofold axes in each octamer are non-crystallographic.

Data collection and processing

Diffraction intensities have been measured from seven crystals to provide a complete data set to 2·0 Å resolution. Table 1 summarizes the data collection information for the crystals. Friedel pairs were collected for each crystal on a FACS-1 diffractometer with crystal-to-detector distances of 250–300 mm. The diffractometer had been modified by adding a helium-filled path between the crystal and detector to reduce absorption of the diffracted beam.

The ω–2θ step-scan technique (Hanson, Watenpaugh, Sieker & Jensen, 1979) was used to collect data at a rate of approximately 3000 reflections per day. Backgrounds were assumed to be a function only of θ and representative values were obtained at the end of data collection by counting pseudo-reflections between the reciprocal-lattice rows. An empirical absorption correction was derived from measurements of a reflection at various φ values at $\chi = 90°$ (North, Phillips & Mathews, 1968). Standard reflections spanning a range of 2θ values were measured periodically to provide a correction for deterioration as a function of time and 2θ. Table 2 summarizes the absorption and deterioration corrections applied to each crystal.

The Friedel pairs were averaged to eliminate the effects of anomalous scattering, and replicate reflections were averaged to give a set of unique reflections. Table 3 summarizes the results.

The initial refinement cycles were against the data used in the earlier 2·5 Å resolution ΔF refinement (Stenkamp, Sieker & Jensen, 1978). When the resolution was extended to 2·0 Å, data from the two earlier crystals were reprocessed along with those from the four additional crystals. Crystals 1, 2 and 3 covered successive shells of data to 2·8, 2·5 and 2·2 Å respectively. Crystal 4 covered a segment of the shell from 2·5 to 2·0 Å, and crystals 5 and 6 overlapped extensively and covered the rest of the shell from 2·2 to 2·0 Å. Table 4(*a*) lists the R values and numbers of overlapping reflections between pairs among the first six crystals. Comparing the R values suggests that data from crystal 3 are of poorer quality than the data from other crystals. Although crystal 3 was included in the 2·0 Å data set for most of the refinement, it was eventually replaced by crystal 7 which covered the same region of reciprocal space. Table 4(*b*) lists R values and numbers of overlapping reflections between crystal 7 and crystals 2, 4, 5 and 6.

The number of reflections varied somewhat during the refinement procedure owing to the reprocessing of the data. Out of a total of 40422 unique reflections in the data set for the final refinement cycles, intensities for 32254 exceed $2\sigma(I)$.*

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1HMM and R1HMMSF), and are available in machine-readable form from the Protein Data Bank at Brookhaven or one of the affiliated centers at Cambridge, Melbourne or Osaka. The data have also been deposited with the British Library Lending Division as Supplementary Publication No. SUP 37001 (3 microfiche). Free copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

Table 1. *Data collection information*

Space group *P*4

Crystal number	1	2	3	4	5	6	7	Average
Number of reflections collected	21512	25499	22783	19654	8016	18195	22816	
Cell constants (Å) $a = b$	86·58 (4)	86·62 (2)	86·45 (3)	86·50 (3)	86·78 (2)	86·74 (4)	86·44 (2)	86·59
c	80·84 (4)	80·89 (2)	80·99 (2)	80·90 (2)	81·08 (1)	81·15 (4)	81·12 (2)	81·00
ω–2θ steps, number of steps	5	5	5	5	5	5	5	
step size (°)	0·10	0·12	0·10	0·10	0·10	0·11	0·12	

Cu $K\alpha$ wavelength = 1·5418 Å.

Table 2. *Absorption and deterioration corrections*

Crystal number	1	2	3	4	5	6	7
Maximum absorption correction	1·111	1·072	1·174	1·085	1·055	1·087	1·165
Maximum deterioration correction	1·402	2·166	2·210	1·721	1·209	1·884	2·815

Functional form of deterioration correction where the C_i are determined by a least-squares fit of the standard reflections

For crystals 1–6: $\text{scale} = C_1 + C_2 \times |2\theta| \times \text{time} + C_3 \times (2\theta)^2 \times \text{time}^2 + C_4 \times (2\theta)^2 \times \text{time}$

For crystal 7: $\text{scale} = C_1 + C_2 \times |2\theta| \times \text{time} + C_3 \times (2\theta)^2 \times \text{time}^2 + C_4 \times |2\theta| + C_5 \times |2\theta|^2 \times \text{time} + C_6 \times (2\theta)^2$

Table 3. *Statistics from averaging of replicates*

The R from the mean is $R_M = \sum_{j=1}^{nref} \sum_{i=1}^{nobs} |\,|F_{ij}| - |\overline{F_j}|\,| / \sum_{j=1}^{nref} |\overline{F_j}|$,

where *n*ref is the number of reflections and *n*obs is the number of observations for the jth reflection.

The R for the Friedel mates is $R_F = \sum_{j=1}^{nref} 2|\,|F_+| - |F_-|\,| / \sum_{j=1}^{nref} |\,|F_+| + |F_-|\,|$,

where *n*ref is the number of observed pairs and F_+ and F_- are the members of the Friedel pair. R values for F^2 are analogous to these.

Crystal number	1		2		3		4		5		6		7	
	R	*n*ref	R	*n*ref	R	*n*ref	R	*n*ref	R	*n*ref	R	*n*ref	R	*n*ref
R_M F	0·014	4397		0	0·055	67	0·036	6	0·053	6		0	0·072	18
R_M F^2	0·027				0·082		0·070		0·115				0·132	
R_F F	0·021	9041	0·054	8223	0·067	6933	0·059	5686	0·065	1381	0·064	6564	0·057	7487
R_F F^2	0·033		0·088		0·119		0·098		0·120		0·119		0·093	

Table 4. *R values for common reflections*

(*a*) Between crystals 1–6

The R for the overlap reflections between crystals A and B is

$$R_0 = \sum_{j=1}^{nref} 2|\,|F_A^2| - |F_B^2|\,| / \sum_{j=1}^{nref} |\,|F_A^2| + |F_B^2|\,|,$$

where *n*ref is the number of contributing reflections.

Only overlap reflections with $I > 2\sigma(I)$ contribute to the R values.

Crystal numbers A/B	R_0	Number of contributing reflections	Number of overlaps
2/1	0·034	416	501
3/2	0·223	1103	1736
4/2	0·134	619	843
4/3	0·168	2332	3200
5/3	0·163	362	642
6/3	0·177	1006	1305
6/5	0·176	2568	3967

(*b*) For crystal 7

Crystal numbers A/B	R_0	Number of contributing reflections	Number of overlaps
7/2	0·151	1240	1670
4/7	0·077	2449	3215
5/7	0·102	394	675
6/7	0·130	1065	1366

Initial model

We did not use the model from the earlier ΔF refinement at 2·5 Å resolution (Stenkamp, Sieker & Jensen, 1978) because we intended this refinement to be independent of the earlier one. Instead, we went back to the model used to initiate that refinement and derived our starting model from it by deleting the Fe atoms and what had been thought to be a bridging water molecule along with the side-chain atoms beyond the C^β of the amino acids taken to be involved with the complex (His 25, His 54, Glu 58, His 73, His 77, His 101, Asp 106, Tyr 109). The idea was to minimize, in so far as possible, any bias from the initial model while developing the binuclear iron complex during the refinement. The initial model included 3616 protein C, N, O and S atoms.

Computational techniques*

The crystallographic refinement has been carried out using the restrained least-squares program of Hendrickson & Konnert (1980). In each cycle, the subunits were refined one at a time with partial contributions from the other three. The scale factor between F_o and F_c and the overall temperature factor, B, were not allowed to vary in the initial least-squares cycles. In the subsequent refinement, the observed structure factors were rescaled to the calculated values using a linear rescale since least-squares scale factors are more sensitive to the completeness of the model. When individual atomic B values were introduced, the shifts in the thermal parameters were damped by a factor of 0·5 to aid in convergence (E. T. Adman, personal communication). Non-crystallographic symmetry restraints were not applied.

The choice of appropriate weights for the various restraints has been of considerable concern to us since previous experience in this laboratory (E. T. Adman, personal communication; M. Ramanadham, personal communication; R. E. Stenkamp, unpublished data) indicated that this type of refinement is quite sensitive

* All calculations were carried out on a VAX 11/780 computer. Each cycle at 2·0 Å resolution required approximately 36 h of CPU time with virtually no time being lost due to page swapping, in contrast to the experiences of other workers (P. Briley, personal communication; G. Cohen, personal communication). The working set size was enlarged to about 300 pages, and this, rather than any program changes, seemed to control the effects of page swapping. The amount of CPU time necessary for this refinement could have been greatly reduced by the addition of a floating-point accelerator. This and the fact that quite different hardware configurations and operating systems are available should be kept in mind when comparing these running times with those from other laboratories.

Table 5. *Weighting schemes*

	Model with $R = 0.198$		Model with $R = 0.189$	
	σ	$w\Delta^{2}$*	σ	$w\Delta^{2}$*
Distance restraints		0.2369330×10^{4}		0.1778400×10^{4}
Bond lengths	0.050		0.050	
Bond angles	0.100		0.060	
Torsion angles	0.140		0.140	
Planar restraints	0.050	0.2539385×10^{3}	0.050	0.2607325×10^{3}
Chiral-center restraints	0.100	0.4359788×10^{2}	0.100	0.5407101×10^{2}
Van der Waals contacts	0.200	0.1066974×10^{4}	0.200	0.6667873×10^{3}
Thermal-parameter restraints	2.0, 3.0	0.6651211×10^{3}	2.0, 3.0	0.6830048×10^{3}
Structure factors	1/70†	0.6183388×10^{5}	1/70†	0.5705429×10^{5}

* $w\Delta^{2}$ for the restraints are for one subunit only and should be multiplied by four to compare with the structure factor terms.
† $\langle |F_o - F_c| \rangle = 74.02$ and 70.60 for these two models.

to the relative weighting of the structure factors and the restraints. The weights initially used for the restraints are the same as those tabulated in column one in Table 5 and are essentially those suggested by E. T. Adman. As the refinement progressed, the weights on the restraints were generally left unchanged, but the structure factor weights were adjusted to yield models with r.m.s. differences between the calculated and ideal bond lengths of about 0.04 Å. In the last few cycles, the weights for the bond-angle restraints were increased as indicated in Table 5 in order to improve their agreement with the ideal values.

Difference electron density maps, used to indicate regions in the structure requiring special attention, were calculated with a fast Fourier program (Ten Eyck, 1973). An interactive computer graphics system, based on a Vector General 3400, with software developed by Jones (1978), was used to reposition the model in various electron density maps and to generate some of the figures in this paper. Many of the difference maps were scanned visually by inspecting contoured sections showing difference density with the locations of atoms present in the model superimposed.

Refinement at 2.5 Å resolution

Fig. 1 shows schematically the stages in the development of the current model of the protein. Fig 2 is a plot of the R values ($R = \sum||F_o| - |F_c||/\sum|F_o|$) omitting the reflections from ∞ to 10 Å resolution, the r.m.s. shift in the protein atoms, and the r.m.s. differences between the calculated and ideal bond lengths as a function of model number. Solid circles in the figure are for models derived from the preceding one by crystallographic refinement; open circles represent models altered or adjusted by other means.

Stage I (see Fig. 2). This segment of the refinement consisted of eight least-squares cycles involving 3616 protein atoms and the 2.5 Å resolution data set. The overall B was fixed at 8 Å^2 as in the earlier ΔF refinement (Stenkamp, Sieker & Jensen, 1978). R decreased from 0.439 to 0.303 for 16363 reflections. The hydrogen bonds in the largely helical structure were included as restraints for the first five cycles.

A ΔF map calculated at an R of 0.303 was consistent with the confacial bioctahedron model (Stenkamp, Sieker & Jensen, 1976). However, there was no electron density connecting the hydroxyl of Tyr 109 to the Fe atom. It was already known from the ΔF refinement that Tyr 109 was too far from the Fe atom to serve as a direct ligand, but on the basis of the averaged 2.8 Å resolution electron density map which showed density connecting the hydroxyl of Tyr 109 with the Fe atom, we thought that a water molecule might be bound to the Fe atom and be hydrogen

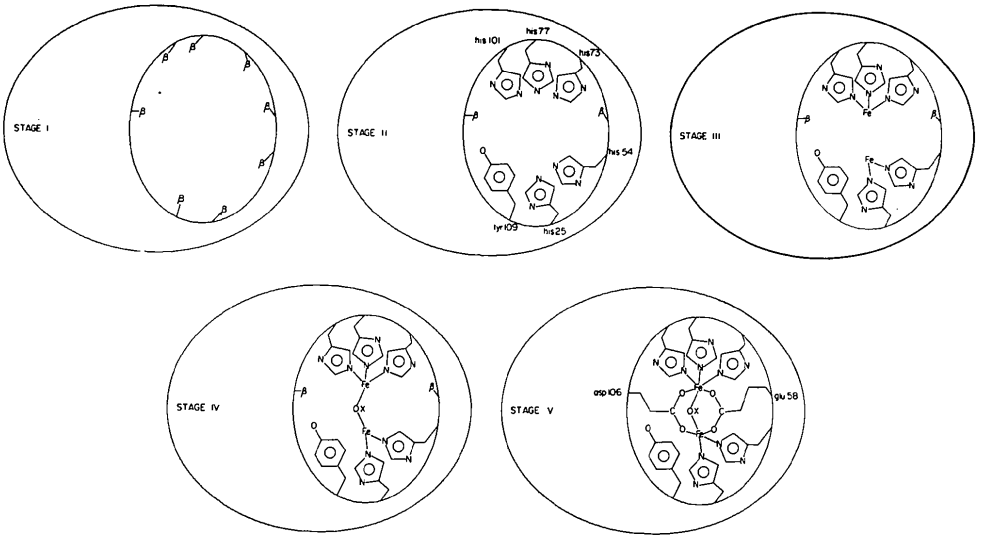

Fig. 1. Schematic diagram indicating the portions of the complex included at each stage of the refinement and development of the model.

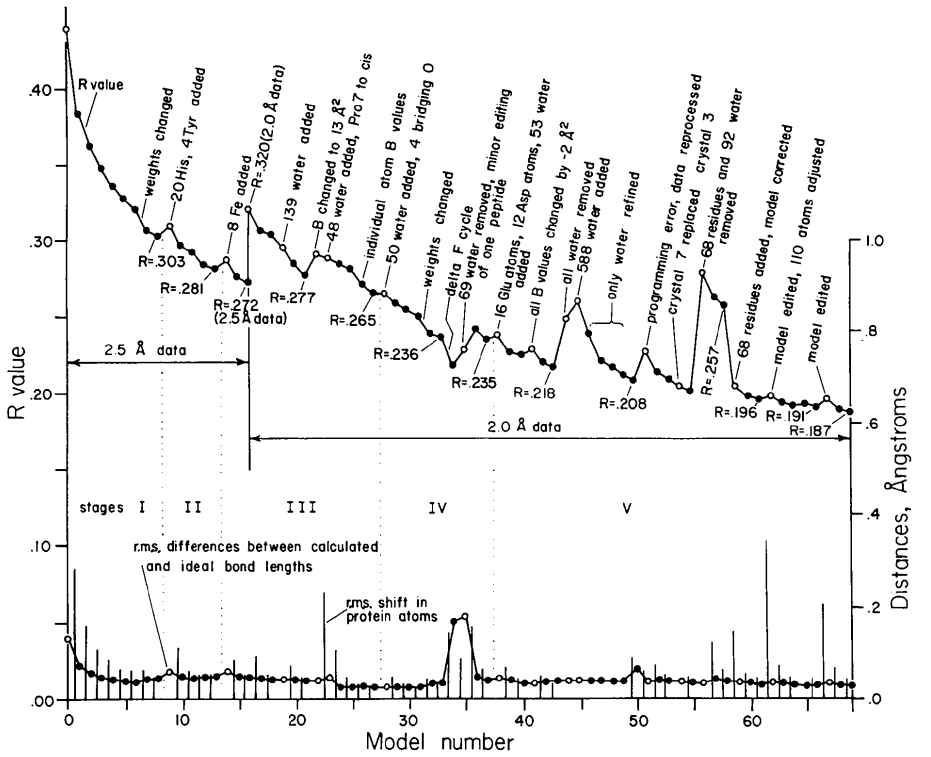

Fig. 2. Plot of R, the r.m.s. shift in the protein atoms, and the r.m.s. difference between the calculated and ideal bond lengths *versus* model number.

bonded to the tyrosyl hydroxyl group. From the present work, it is clear that if it were there, its electron density would be considerably less than that of any of the other atoms in the model. Indeed it would be below the noise level in the free solvent volume of the crystal. This is true for all four subunits. Thus, no positive evidence was found in the ΔF maps for water linking the Fe atom with Tyr 109.

Stage II. The 20 histidine and four tyrosine residues were added to the model in this stage, increasing R to 0·310. Four cycles of least squares reduced it to 0·281.

Stage III. In this stage, the Fe atoms were added, each being restrained to be 2·0 Å from the $N^{\epsilon 2}$ of its ligating histidine residues. No van der Waals restraints were included for the Fe atoms. Inclusion of the Fe atoms increased R to 0·288; the two following cycles of least squares reduced it to 0·272. The average Fe–Fe distance at this point was 2·95 Å.

Refinement at 2·0 Å resolution

Extension of the resolution to 2·0 Å allowed us to estimate better the overall temperature factor since the plot of $\sum|F_c|/\sum|F_o|$ *versus* $(\sin\theta/\lambda)^2$ was based on many more reflections at higher $\sin\theta/\lambda$. The overall B of 15 Å^2 determined from the plot is nearly twice the initial value, and it was held at the new value for several refinement cycles.

R was 0·320 for the 28950 reflections to 2·0 Å resolution. [13134 reflections with $I < 2\sigma(I)$ were omitted from the refinement.] Two refinement cycles reduced R to 0·304. At this point, 139 solvent O atoms were located in a ΔF map and were added to the model, reducing R to 0·295. They satisfied the requirement of hydrogen bonding to the protein or to another solvent molecule.

Two refinement cycles reduced R to 0·277. The overall B was changed to 13 Å^2 on the basis of a plot of $\sum|F_c|/\sum|F_o|$ *versus* $(\sin\theta/\lambda)^2$, and 48 more solvent atoms were added. In addition, the interactive graphics facility was used to convert proline 7 to the *cis* conformation to account better for the electron density in ΔF and $2F_o - F_c$ maps. R increased to 0·289, and two cycles of least squares reduced it to 0·281. At this point, a thermal parameter was included for each atom in the model. Two least-squares cycles reduced R to 0·265.

Stage IV. Fifty water O atoms were added in this stage to give a total of 237 solvent molecules. In addition, what was taken to be a bridging water molecule was clearly evident in the midplane between the Fe atoms, and an O atom was added in that position, being restrained to be 2·0 Å from each Fe

atom. Three cycles of refinement reduced R to 0·250. After reducing the weights on the restraints, two additional least-squares cycles reduced R to 0·236.

Before the addition of the bridging water molecule, the average Fe—Fe distance was 3·09 Å; the five following refinement cycles with the bridging water O atom increased it to 3·18 Å.

In large minimization problems such as this, many local minima undoubtedly exist in the $\sum(\Delta F)^2$ function. Experience suggests that a refinement may lodge in a local minimum and that application of a different technique or algorithm is useful in perturbing the model to follow a different minimization pathway. Prompted by the experiences of M. Guss (personal communication), we subjected the model to an unrestrained, automated ΔF cycle (Freer, Alden, Carter & Kraut, 1975). R decreased steeply to 0·218 (see Fig. 2), but at the same time the r.m.s. difference of bond lengths from ideal values increased drastically, from 0·037 to 0·172 Å. The solvent model was extensively edited, retaining 168 solvent molecules, and R increased to 0·228. In the next cycle of least squares, R increased to 0·242, the increase being due to the idealization of the nonideal structure resulting from the shifts in the ΔF map. In the following cycle, R decreased to 0·235.

Stage V. The remaining protein residues, Glu 58 and Asp 106, were added at this stage using the graphics system to orient the carboxyl groups in the difference electron density. In the earlier averaged 2·8 Å map, both side chains were clearly bound to both Fe atoms, but the detailed mode of binding could not be determined. The improved resolution and the model refinement of the present study show unequivocally that one oxygen from each carboxyl group is bound to one of the Fe atoms and the second oxygen from each is bound to the other Fe (see Fig. 1, stage V). Fig. 3 shows the carboxylate model superposed on the difference density in one of the subunits. No alternative placement of the carboxyl groups is evident. The same mode of binding was found in the other three subunits in the asymmetric unit.

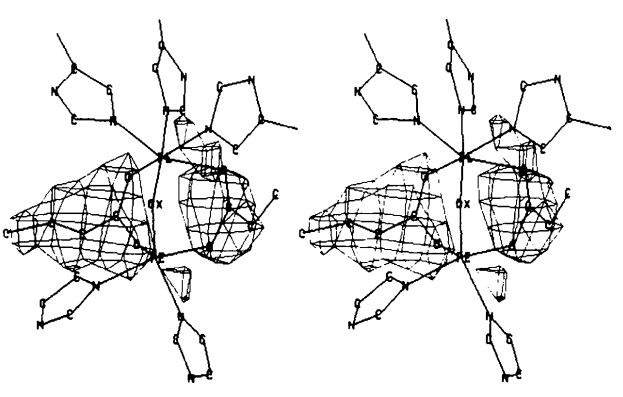

Fig. 3. Stereoscopic view of the difference density in one of the subunits corresponding to the side-chain atoms of glutamic acid 58 and aspartic acid 106 and superimposed model of the binuclear Fe complex.

The O atoms were restrained to be 2·0 Å from the Fe atoms. This model, including 221 solvent molecules, was refined for two least-squares cycles, after which all the B values were reduced by 2 Å^2. The model was then refined for two more cycles, reducing R to 0·218.

Thus far, only the most obvious solvent molecules had been included, and no attempt had been made to impose non-crystallographic symmetry on them, although some of the water molecules were found in sites common to all four subunits [for a preliminary description of the solvent model, see Stenkamp & Jensen (1981)]. To expedite the refinement, we checked the solvent model at this point and extended it by imposing non-crystallographic symmetry. In doing so we calculated structure factors including only the protein atoms (R = 0·248) and extensively checked the peaks in a ΔF map, extending the model to include 588 solvent molecules which obeyed the non-crystallographic symmetry, were located in residual electron density, and were directly or indirectly hydrogen bonded to some part of the protein. R increased to 0·260 upon expansion of the solvent model. Four cycles of refinement, adjusting x,y,z and B of the solvent molecules only, reduced R to 0·213, and one additional cycle on the complete model reduced it to 0·208.

At this point, a programming error in the treatment of the deterioration correction was discovered and led to reprocessing of the data. R increased to 0·227. Two cycles reduced it to 0·209.

Since the data from crystal 3 appeared to be inferior to those from the other crystals (see Table 4), they were replaced by data from crystal 7, reducing R from 0·209 to 0·204. It was encouraging that R decreased when the data set was improved, but an additional cycle of least squares reduced it only to 0·202. The final data set used in the refinement had 32254 reflections with $I > 2\sigma(I)$.

Inspection of a ΔF map indicated model error near residues 11 and 12 and at the carboxyl terminus, but the residual density was not easily interpreted. Accordingly, residues 1–15 and 112–113 and associated solvent molecules in each subunit were removed from the model, which was then subjected to two refinement cycles to remove bias before calculating another ΔF map. This map indicated that the peptide common to residues 11 and 12 in all four subunits needed reorientation, and that the C terminal isoleucine side chains had been placed in the density corresponding to the carboxyl group and *vice versa*. These errors were corrected, O atoms were added to the terminal carboxyl groups (they previously had been treated as carbonyls), and other solvent molecules located near the bridging O atoms in the complexes were added to the model. Two cycles of refinement of this model

(3788 protein atoms, 496 solvent molecules) reduced R from 0·204 to 0·196.

Up to this point, we had relied solely on difference maps to indicate errors in the model. We now calculated the torsion angles in the main chain and side chains and found that some regions of the model which had not been flagged by the difference map required drastic revision. Several ω angles were found with values near 0° rather than the 180° expected for *trans* peptides.* Fig. 4 indicates the types of distortion generated by the refinement, even though the peptide parameters were restrained. Another view of the distortions is shown in Fig. 5 where the ω values are plotted as a function of refinement cycle. The plots with abrupt changes from 0 to 180° are cases where the carbonyl C and O atoms have shifted along the C—O vector, while the plots with gradual transitions in ω result from rotation of the peptide group coupled with distortion. Calculation of the bond angles for the protein also indicated areas requiring special attention.

* The analysis of these regions was facilitated by superposing plots of the affected peptide groups for each model and flipping them quickly to provide an animated view of the movement of the atoms.

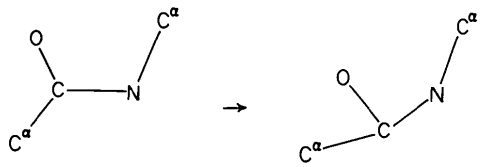

Fig. 4. Schematic drawing indicating the types of distortion of the peptide group generated by the restrained refinement.

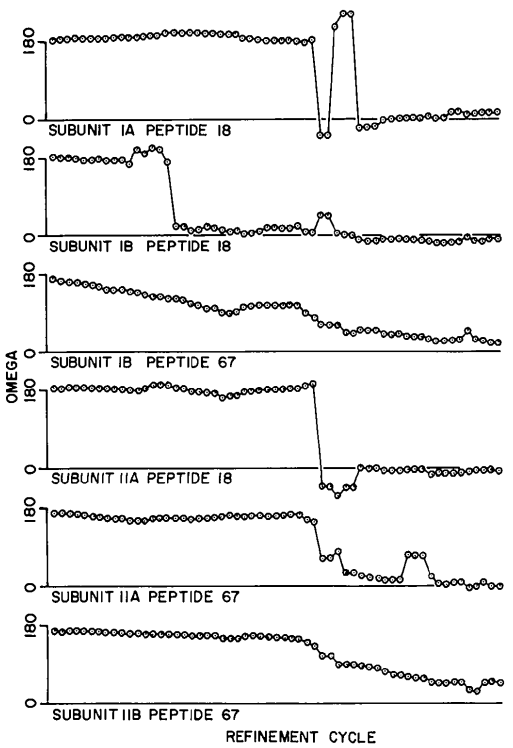

Fig. 5. Values of ω, indicating the peptide distortion, as a function of refinement cycle.

Some of the serious distortions were corrected by use of the graphics system while others were dealt with by transforming into the problem subunits a region in another subunit judged to be correct. Some atoms were also adjusted to improve the consistency of bond and torsion angles in the four independent subunits. In addition, the side chain of Trp 10 was flipped on the basis of two pieces of evidence: no density at the C^{β} atom in a ΔF map, and distortion of the six-membered ring of the aromatic group. No other strong indications of the error or type of correction necessary were evident. In subsequent maps, density appeared at the C^{β} atom, and the side chain became more nearly ideal, indicating that the change improved the model. In all 110 atomic positions were adjusted, leading to an increase of R only to 0·198.

At this point the weights of the angle restraints were increased by a factor of 1·25 relative to the other restraints, and in two least-squares cycles, R decreased to 0·192. The angular restraints were again increased, this time by a factor of 1·33. R increased slightly to 0·193 in the next least-squares cycle and decreased to 0·191 in the following one.

Before the final two least-squares cycles, the first five residues in three subunits along with a number of other atoms (a total of 171) were adjusted using the display system. R increased to 0·196, and the final two cycles reduced it to 0·187. Although some additional changes could be made, the conformational parameters for the four subunits are reasonably consistent with each other and with values expected on the basis of stereochemical considerations.

Discussion

As is commonly observed for protein crystals, the ranges of unit-cell parameters for methydroxohemerythrin listed in Table 1 greatly exceed the standard deviations in the parameters of individual crystals. It is clear that relatively small differences in the crystallizing solution or in handling the crystals cause easily detected differences in unit-cell parameters. Although we assume the crystals to be isomorphous, this is not strictly true and may account in part for the rather high R values listed in Table 4 for the common reflections between pairs of crystals. We note that replicate reflections from individual crystals listed in Table 3 show better agreement.

To compare the present refinement with the earlier one (Stenkamp, Sieker & Jensen, 1978), we have plotted R in Fig. 6 as a function of the cycle number for

the first 14 cycles of the present refinement and the 14 cycles of the earlier one. Although the model in the earlier ΔF refinement included more atoms (all the protein atoms and, beyond the eighth cycle, some solvent), nevertheless the rate of refinement as measured by the decrease in R is similar for the two methods, and the distance restraints applied led to r.m.s. differences from ideal bond lengths similar in magnitude for the two methods (0·04–0·06 Å, except for the initial least-squares cycles). Indeed, we sought to adjust the weights in the restrained least-squares method to ensure a model sufficiently flexible that refinement would not be impaired, but at the same time sufficiently rigid that unacceptable distortions would not occur. Despite our intent, the angular weights did not prevent the serious distortions plotted in Fig. 5. As has been observed by Kim & Holbrook (private communication), our experience suggests that maximum deviations of the restrained parameters from ideal values should be closely monitored during the course of a restrained least-squares refinement.

Two of the abrupt distortions plotted in Fig. 5 (residue 18 in subunits IA and IIA) occurred during the ΔF cycle when no restraints were applied. Checking the coordinates of the preceding least-squares cycles shows, however, that these residues were already substantially distorted, and the restraints in the following restrained least-squares cycles were insufficient to restore acceptable geometry. In retrospect, on the basis of the behavior of R in Fig. 2, the ΔF cycle does not appear to have been advantageous. In any case, reducing or removing restraints will lead to an immediate more rapid decrease in R, but to the extent that the distortions are not real, the refinement will eventually be impaired. When the restraints were reimposed in the least-squares cycle following the ΔF cycle, R increased because of the idealization, and the large r.m.s. difference from ideal bond lengths decreased dramatically (see Fig. 2).

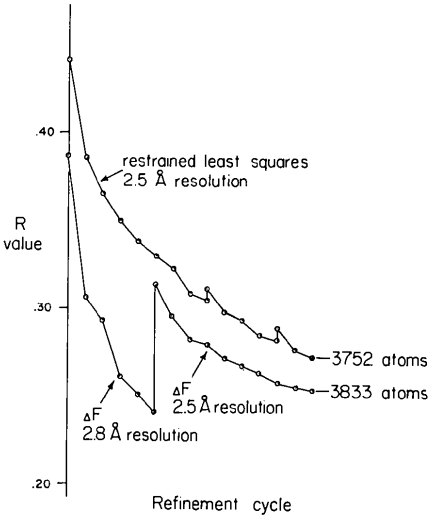

Fig. 6. Comparison of R values for restrained least-squares and automated ΔF refinement.

The effects of changing the weights of the X-ray data relative to the restraints are usually evident in Fig. 2 both in R and in the r.m.s. difference from ideal bond lengths. Thus after the sixth least-squares cycle in stage I, the weights for the X-ray data relative to the restraints were increased by a factor of 1·44; R decreased proportionately more than in the preceding cycle, and the r.m.s. difference from ideal bond lengths increased slightly. Similar behavior was observed after the third least-squares cycle in stage IV when the weights for the X-ray data were increased by a factor of 1·92. A larger effect was observed in the r.m.s. difference from ideal bond lengths before the seventh least-squares cycle in stage III when the weights of the X-ray data were decreased relative to the restraints by a factor of 0·40. Effectively, this increased the restraints, and the r.m.s. difference from ideal bond lengths decreased from 0·05 to 0·029 Å. The accompanying change in R, however, was masked by changes in the two preceding models.

The Fe—Fe distance is an important parameter in the binuclear iron complex, but the presence of unresolved or barely resolved ligand atoms complicates its evaluation. Thus, at the beginning of stage IV, the bridging O atom was added to the model, leading to an increase in the Fe—Fe distance of 0·09 Å in the next least-squares cycles. The increase is no doubt due in part to the overlap of the oxygen electron density with that of the adjacent Fe atoms, but it will also depend on the value used for the restrained Fe—O distance. Any restrained parameter will, of course, be biased toward the value used whether or not it is correct.

A similar effect on the Fe—Fe distance results from applying van der Waals restraints. In a set of test refinements, a van der Waals radius of 1·8 Å was used for iron, and the average Fe—Fe distance increased to about 3·45 Å. The effect was removed by setting the iron radius to 0·00 to eliminate any van der Waals contribution from the Fe atoms.

The sheer size of this structure necessitated the use of automated methods of refinement, requiring as little use as possible of manual adjustment of the approximately 4000 atoms in the molecule. At this point, it is not clear how best to avoid the distortions we encountered while maintaining a model of sufficient flexibility, but the easiest means of locating regions of unacceptable geometry is to monitor the maximum deviation of restrained parameters from their ideal values. If the restraints are too tight, the model will retain acceptable geometry, but atoms may still be in error. In this case, the difficulty should be evident in a ΔF map, which, within the limits set by error and resolution of the X-ray data, is a sensitive indicator of model error if contoured at appropriate levels. Nevertheless, searching through large maps is taxing, and

some regions of a map, particularly at low resolution and high R values, may be difficult or impossible to interpret.

Although this form of methemerythrin has a large asymmetric unit, the four subunits in it enable one to check any feature that is obscure in one subunit against the other three. During the refinement, it was particularly useful to check the behavior and the results in each subunit against the others. It is clear from the plots in Fig. 5 that in the course of the refinement, corresponding regions of different subunits did not necessarily behave in the same way.

The four independent subunits also enable us to make a quantitative estimate of the errors in atomic positions. After superposing the four subunits and averaging the atomic positions, we find the r.m.s distance between all atoms in the subunit and their average positions to be 0·271, 0·287, 0·288 and 0·324 Å for subunits I*A*, I*B*, II*A*, and II*B* respectively. The overall r.m.s. difference for all atoms in all subunits is 0·293 Å, leading to a standard deviation of 0·338 Å. The overall r.m.s. difference for the C^{α} atoms only in all subunits is 0·136 Å, leading to a standard deviation of 0·157 Å.

The final R is considerably lower for this refinement than for the earlier ΔF refinement against the 2·5 Å data set [0·187 *vs* 0·253 (Stenkamp, Sieker & Jensen, 1978)]. This is due in part to the use of restrained least squares which is somewhat more powerful in reducing R, at least at the lower R values (R. E. Stenkamp, unpublished data), but the higher resolution (2·0 Å) also contributes to the lower R since the more extensive data are more useful in correcting model errors.

The model could probably be refined somewhat further, but we do not believe the expected improvement would warrant the effort. To what extent the quality of the data may have affected the convergence rate or whether slow convergence is an inherent characteristic of the restrained least-squares method is not clear. The method has converged rapidly in some cases (Konnert, 1976), but the necessity of many refinement cycles has also been reported (Anderson, Stenkamp & Steitz, 1978; Sielecki, Hendrickson, Broughton, Delbaere, Brayer & James, 1979). The behavior of restrained least squares seems to be quite sensitive to the relative weights of the terms in the function being minimized, and it is possible that different weighting schemes could improve the convergence rate. The size of this problem, however, does not recommend it for operationally checking the effects of different weighting schemes.

This work has been supported by grant GM-10828 from the National Institutes of Health and equipment grant PCM 76-20557 from the National Science Foundation.

References

ANDERSON, C. M., STENKAMP, R. E. & STEITZ, T. A. (1978). *J. Mol. Biol.* **123**, 15–33.

DUNN, J. B. R., ADDISON, A. W., BRUCE, R. E., LOEHR, J. S. & LOEHR, T. M. (1977). *Biochemistry*, **16**, 1743–1749.

FREER, S. T., ALDEN, R. A., CARTER, C. W. JR & KRAUT, J. (1975). *J. Biol. Chem.* **250**, 46–54.

HANSON, J. C., WATENPAUGH, K. D., SIEKER, L. & JENSEN, L. H. (1979). *Acta Cryst.* A**35**, 616–621.

HENDRICKSON, W. A. (1978). *Nav. Res. Rev.* **31**, 1–20.

HENDRICKSON, W. A. & KONNERT, J. H. (1980). *Biomolecular Structure, Function, Conformation and Evolution*, edited by R. SRINIVASAN, Vol. 1, pp. 43–57. Oxford: Pergamon Press.

JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

KONNERT, J. H. (1976). *Acta Cryst.* A**32**, 614–617.

LOEHR, J. S., MEYERHOFF, K. N., SIEKER, L. C. & JENSEN, L. H. (1975). *J. Mol. Biol.* **91**, 521–522.

NORTH, A. C. T., PHILLIPS, D. C. & MATHEWS, F. S. (1968). *Acta Cryst.* A**24**, 351–359.

SIELECKI, A. R., HENDRICKSON, W. A., BROUGHTON, C. G., DELBAERE, L. T. J., BRAYER, G. D. & JAMES, M. N. G. (1979). *J. Mol. Biol.* **134**, 781–804.

STENKAMP, R. E. & JENSEN, L. H. (1979). *Adv. Inorg. Biochem.* **1**, 219–233.

STENKAMP, R. E. & JENSEN, L. H. (1981). *Structural Aspects of Recognition and Assembly in Biological Macromolecules*, edited by M. BALABAN, J. L. SUSSMAN, W. TRAUB & A. YONATH. Glenside, PA: International Science Services.

STENKAMP, R. E., SIEKER, L. C. & JENSEN, L. H. (1976). *Proc. Natl Acad. Sci. USA*, **73**, 349–351.

STENKAMP, R. E., SIEKER, L. C. & JENSEN, L. H. (1978). *Acta Cryst.* A**34**, 1014–1019.

STENKAMP, R. E., SIEKER, L. C., JENSEN, L. H. & LOEHR, J. S. (1976). *J. Mol. Biol.* **100**, 23–34.

STENKAMP, R. E., SIEKER, L. C., JENSEN, L. H. & MCQUEEN, J. E. JR (1978). *Biochemistry*, **17**, 2499–2504.

TEN EYCK, L. F. (1973). *Acta Cryst.* A**29**, 183–191.