J. Mol. Biol. (1984) **175**, 159–174

The Crystal Structure of Human Deoxyhaemoglobin at 1·74 Å Resolution

G. Fermi, M. F. Perutz, B. Shaanan†

Medical Research Council, Laboratory of Molecular Biology Hills Road, Cambridge CB2 2QH, England

AND

R. Fourme

LURE (CNRS-Universite Paris-Sud) and Laboratoire de Biologie Physicochimique (GR13, CNRS), 91405 Orsay, France

(Received 20 December 1983)

The structure of human deoxyhaemoglobin was refined at 1·74 Å resolution using data collected on film at room temperature from a synchrotron X-ray source. The crystallographic *R*-factor is 16·0%. The estimated error in atomic positions is 0·1 Å overall, 0·14 Å for main-chain atoms of internal segments, and 0·05 Å for the iron atoms. The effects of intermolecular contacts on the structure were investigated; such contacts cause only highly localized distortions, as judged from the degree of molecular asymmetry that they induce.

The geometry of the iron–nitrogen complex closely resembles that of the deoxymyoglobin structure of Takano (1977) and of the 5-co-ordinated model compounds of Hoard (1975) and Jameson *et al.* (1980). The distance of the iron from the mean plane of N(porphyrin) is 0·40(5) Å and 0·36(5) Å, respectively, at the α and β haems, in contrast to the corresponding distance of +0·12(8) Å and −0·11(8) Å in oxyhaemoglobin (Shaanan, 1983); the Fe–N_ε(F8) bond length is 2·12(4) Å and the Fe–N(porphyrin) bond length is 2·06(2) Å; the last is also in good agreement with extended X-ray fluorescence spectroscopy measurements on deoxyhaemoglobin (Eisenberger *et al.*, 1978; Perutz *et al.*, 1982). The haems are domed toward the proximal side; the separation between the mean planes of N(porphyrin) and C(porphyrin) being 0·16(6) Å and 0·10(6) Å, respectively at the α and β haems. At the α haems, the normals to the mean pyrrole planes are tilted uniformly toward the haem centre, by about three degrees relative to the haem normal, and there is a folding of about four degrees of the haem about an axis running between the methene carbons that are between the pyrrole rings bearing like-type side-chains. At the β haems, there is no such folding, and only pyrroles II and IV (those eclipsed by His F8) are appreciably tilted, by about eight degrees. The independence of these parameters from restraints imposed on the model was verified by unrestrained refinement of the entire molecule starting from a structure with modified haem geometry.

† Present address: Department of Structural Chemistry, The Weizmann Institute of Science, 76 100 Rehovot, Israel.

0022–2836/84/140159–16 $03.00/0 © 1984 Academic Press Inc. (London) Ltd.

1. Introduction

The least satisfactory portion of current qualitative descriptions of the structural basis of the allosteric mechanism in haemoglobin (Monod *et al.*, 1965) is the basis of the low oxygen affinity of the α subunits in the T(deoxy) quaternary structure. A plausible structural basis for the greater interunit binding forces in the T structure relative to the R(oxy) quaternary structure has been provided in terms of the relative numbers of salt bridges (Perutz, 1970) and amount of buried surface area (Chothia *et al.*, 1976) in the two quaternary forms. The relationship between quaternary structure and at least the gross tertiary structure in the vicinity of the haem has been traced (Baldwin & Chothia, 1979): the T to R transition enforces shifts of both the α and β F helices (relative to the haem) of the order of 1 Å, and a shift of about 2 Å of the β E helix, with but little effect on the α E helix, relative to the haem. The relative positions of the E helix and the haem provide a ready explanation for the low oxygen affinity of the β subunits in the T state, for in the T state the ligand site is blocked by Val E11 and in the R state it is not. Other factors may well influence or even dominate the T state affinity of the β haems, but a plausible and sufficient explanation is in hand. For the α haems, however, more subtle stereochemical factors must be involved, for there the ligand site is blocked only by a water molecule in the T state; moreover, a water molecule is also present near the ligand site of deoxymyoglobin (Takano, 1977), which has high oxygen affinity. Perutz (1970) first proposed that the out-of-plane position of iron in the deoxy form (compared with an in or near planar position in the oxy form) must play a role, and subsequently much effort has been spent on attempting to provide a satisfactory explanation on this basis (for a review, see Perutz, 1979); e.g. steric hindrance between the iron-linked His F8 and the haem has been investigated as a possible source of the low T state oxygen affinity. These efforts have been hampered by the lack of detailed knowledge of the deoxy haem structure; in particular, the degree of doming of the haem, and hence the positions of the porphyrin nitrogens relative to the haem and the iron, could not be established with confidence of the basis of 2·5 Å data (Fermi, 1975; Fermi *et al.*, 1982). We undertook refinement of the deoxyhaemoglobin structure at 1·74 Å resolution primarily to improve this situation.

2. Materials and Methods

(a) *Data collection and processing*

Hb† crystals were grown in buffered ammonium sulphate solution as described by Perutz (1968). Intensities were recorded at room temperature on film by the rotation method (Arndt & Wonacott, 1977), utilizing monochromatic X-radiation with a nominal wave length of 1·4 Å from the storage ring DCI at LURE. Films were mounted 3 per pack in flat cassettes at a crystal-to-film distance of 53 mm, allowing reflections down to 1·7 Å spacing to be recorded. Crystals were mounted with either the *a* or *b* axes parallel to the rotation

† Abbreviations used: Hb, human deoxyhaemoglobin A; HbO_2, human oxyhaemoglobin A; Mb, sperm whale deoxymyoglobin; MbO_2, sperm whale oxymyoglobin; TPP, *meso*-tetraphenyl-porphyrin; TpivPP, *meso*-tetra(α,α,α,α-*o*-pivalamidophenyl)porphyrin; MeIm, methylimidazole; r.m.s., root mean square.

axis. A total of 211 usable exposures (rotation range 1·1 to 1·5 deg.) were taken from 8 crystals, using exposure rates of 200 to 750 s/deg. The total angular range recorded was somewhat greater than the minimum necessary for a complete data set.

Films were densitometered on a flat-bed "hybrid" densitometer (Mallett *et al.*, 1977) redesigned in our laboratory by J. F. W. Mallett. The new instrument has improved optics and includes an automatic film changer that allows measurement of up to 20 films in an unattended overnight run.

Integrated intensities from films within each pack were scaled together and corrected for Lorentz and polarization (Kahn *et al.*, 1982) factors. No absorption correction was applied. Data beyond 1·74 Å were discarded because of poor internal agreement. The resulting intensities were merged together with the 20·0-2·5 Å diffraction data of Ten Eyck & Arnone (1976), with scaling by the method of Fox & Holmes (1966). The merging R factor (on intensity) was 6·7% for all multiply measured or symmetry-related reflections among the 292,000 input reflections. About 20% of the input reflections were partially recorded. The merged intensities were converted to amplitudes with a correction applied to weak and negative intensities based on the *a priori* intensity distribution (French & Wilson, 1978). The final data set comprised 56,287 independent reflections (99·6% of the total in the 20·0-1·74 Å resolution range) together with 54,034 anomalous differences (99·0% of total non-centric reflections). The R factor for Friedel pairs was 3·6% on intensity. 2·8% on amplitude.

Cell dimensions were determined only to an accuracy of about 0·1 Å and did not differ significantly from those more accurately determined by Ten Eyck & Arnone (1976); the latter were used for refinement purposes.

(b) *Refinement of the atomic model*

The starting point for the refinement was the atomic model of Hb refined against data to 2·5 Å previously described (Fermi *et al.*, 1982), which showed an R factor of 26·8% against the full present data set. Refinement was carried out by the procedure of Jack & Levitt (1978), which is equivalent to diagonalized least-squares refinement in reciprocal space with simultaneously applied energy-like restraints between atoms to maintain regularity of model geometry. All observations were included at equal weight throughout the refinement. The energy parameters used were those of Phillips (1980), except that the iron atoms were unrestrained in order to avoid any possible bias in their positions. The scale factor between energy and X-ray terms was varied between 0·00025 and 0·0005 to maintain the r.m.s. variation of C–C single bonds between 0·02 and 0·03 Å. Structure factors were calculated with a 2-gaussian approximation to atomic scattering factors (5-gaussian for the irons) and with 4460 hydrogen atoms of determinable position added to the heavy atom model (Phillips, 1980). Water molecules were tentatively identified by a partially automated procedure that selects those positive peaks in the $F_{obs}-F_{calc}$ difference map above a given threshold that are at good hydrogen-bonding distances from O and N atoms in the protein. The final selection of water molecules was based on visual examination of the $F_{obs}-F_{calc}$ difference map. The final model contained 221 water molecules and 2 ions (probably phosphates).

A constant solvent density was introduced outside of the molecular volume by a method similar to that of Phillips (1980), but with the parameters of the solvent determined by refinement rather than by trial and error. The solvent correction was re-calculated from time to time during refinement. In the final model, the solvent scattering level was $\rho_s = 0{\cdot}38\ e/A^3$, the distance of the solvent edge from the nearest model atom centre was $d_e = 2{\cdot}2$ Å for C and $d_e = 1{\cdot}8$ Å for N and O, and the solvent smoothing temperature factor was $B_s = 41\ Å^2$; the ratios of d_e for C, O and N were kept fixed at values corresponding to the van der Waals' radii of the atom types. These values are similar to those used by Phillips in the oxymyoglobin model: $\rho_s = 0{\cdot}385\ e/A^3$, $d_e = 2{\cdot}2$, 1·2 and 1·8 Å for C, N and O, $B_s = 40\ e/A^3$. Addition of the solvent correction during the 8th cycle of refinement dramatically lowered the R factor at both low and high resolution (from 28% to 18% in

the 20·0-5·39 Å shell and from 36% to 28% in the 1·83-1·74 Å shell); the effect at high resolution was primarily due to improved scaling between F_{obs} and F_{calc} (with a single scale factor and single overall temperature factor).

Refinement was carried out for a total of 14 cycles; positional parameters were refined during 12 cycles and atomic temperature factors during 10. Adjustments of some 30 side-chains of the model were made at cycle 9 after visual inspection of the $2F_{obs}-F_{calc}$ and $F_{obs}-F_{calc}$ maps in a graphics display (Diamond, 1978). This examination revealed that several side-chains (and also the main-chain at the amino termini of the β chains) appear to have two reasonably well-defined positions. However, no attempt was made to include multiple side-chain positions in the model.

The final R factor between F_{obs} and F_{calc} was 16·0% (24·2% in the 1·83-1·74 Å resolution shell), the highest peak in the difference map was 0·7 e/Å^3, and the r.m.s. variation of C-C single bonds was 0·023 Å. The r.m.s. difference in atomic positions between the final and starting models is 0·37 Å. The final model and associated electron density maps show some evidence of significant asymmetry between the non-crystallographically related $\alpha_1\beta_1$ and $\alpha_2\beta_2$ dimers (see Results and Discussion section (a)). However, a symmetry averaged model was calculated, which should be adequate for many purposes and may provide greater co-ordinate accuracy in internal regions of the molecule. For this purpose, the final atomic co-ordinates were symmetry averaged about the molecular dyad and the resulting structure re-regularized by energy refinement without reference to the X-ray data. Fifty-five pairs of symmetry related water molecules were included in the averaged model. The r.m.s. difference in atomic positions between the averaged and unaveraged models is 0·28 Å (protein only). The averaged model with a solvent scattering correction applied, showed an R factor of 20% against the X-ray data.

(c) *Error analysis*

The random error in atomic positions, σp, in the 2·5 Å haemoglobin model was estimated from asymmetry with respect to the (non-crystallographic) molecular dyad and confirmed by a Luzzati (1952) plot of the R factor against reciprocal resolution (Fermi, 1975). At the resolution of the present model, the former method may seriously overestimate σp because of possible true asymmetry of the crystal structure (some evidence of which could be discerned in the electron density map), and the Luzzati method is of dubious relevance to a fully refined structure. Values of σp can be calculated from the diagonal least-squares matrix, but these are uncertain because of the diagonalization procedure, because of approximations used to calculate the matrix elements, and because of the uncertain effects of restraints applied to the model. We have assumed that values of σp calculated from the

TABLE 1

Estimated r.m.s. positional error, σp *(Å) of atoms of the unaveraged model*[a]

	All atoms	Main-chain atoms	Side-chain atoms
Whole molecule	0·21 (0·37)	0·18 (0·22)	0·24 (0·49)
Internal segment[b]	0·17 (0·30)	0·14 (0·14)	0·20 (0·41)
Haems[c]	0·19 (0·28)	0·16 (0·13)	0·23 (0·41)
Irons[d]	0·05	—	—

[a] From matrix-derived σp, scaled by factor 1·7 to make r.m.s. σp from matrix equal to that estimated from molecular asymmetry (values given in parentheses) for main-chain atoms of internal segment.

[b] Segments B6 to CD1, FG5 to G17, H5 to H18 of the α and β chains.

[c] Main-chain = porphyrin core.

[d] See the text.

matrix (neglecting restraints) correctly estimate the relative values of σp for the various atoms, but have applied a scale factor to make the r.m.s. value of the matrix-derived σp match that derived from asymmetry for main-chain atoms of internal segments, a class of atoms for which asymmetry is expected to be minimal. The resulting estimates of r.m.s. σp are listed in Table 1 for various classes of atoms. The listed estimate of overall r.m.s. σp is in good agreement with $\sigma p \approx 0{\cdot}2$ Å estimated from a Luzzati plot. The iron positions were independently determined by least squares refinement against the anomalous dispersion data, thus providing a third estimate of σp. The values of σp for the irons obtained from the scaled real least squares matrix, from asymmetry, from discrepancy between their real-data and anomalous-data positions, and from their r.m.s. distance from the haem axis (i.e. assuming the irons in fact lie on the haem axis) are 0·03 Å, 0·07 Å, 0·05 Å and 0·05 Å, respectively. The value $\sigma p = 0{\cdot}05$ Å in Table 1 is simply the mean of these.

A test was made of the effect of starting position and of the applied energy restraints on the model structure by means of unrestrained refinement from a starting position with a modified haem configuration. The atomic positions of the final unaveraged model were modified as follows: the 32 atoms of each haem plane (porphyrin plus first side-chain carbons) were moved to a strictly planar configuration, and the iron atoms were moved 0·5 Å from their final positions, in a direction away from the haem plane for the $\alpha_1\beta_1$ irons and towards the haem plane for the $\alpha_2\beta_2$ irons. Seven cycles of unrestrained refinement (with both position and atomic temperature factors refined on all but the first) of the entire protein model were then carried out. The R factor after unrestrained refinement was 13·5%. The estimated error in atomic positions for the unrestrained model (calculated as for the final model) was 0·25 Å overall, or about 25% greater than for the final (restrained) model. The overall r.m.s. shift of atomic positions between the final and unrestrained models was 0·21 Å, which equals the overall estimated $\sigma p = 0{\cdot}21$ Å for the final (restrained) model alone; the maximum r.m.s. shift in any one residue was 0·7 Å, and the maximum shift of any atom was 1·8 Å. This is better agreement than would be expected on the basis of independent errors in the 2 models, but of course the errors are not independent because both models were refined against the same data. Haem geometry returned to a close approximation of the final (restrained) model, as is indicated in the Tables of haem geometry in Results and Discussion section (b). The r.m.s. difference in positions of the iron atoms between the final and unrestrained models was 0·03 Å (0·05 Å maximum), compared with $\sigma p = 0{\cdot}05$ Å for the restrained model alone. However, each of the 4 irons showed a bias towards its starting position before unrestrained refinement, ranging from 0·01 to 0·05 Å (0·02 Å mean), despite the fact that the refinement appeared to be complete, as judged from the oscillation of the iron positions on the last few cycles of unrestrained refinement and the very small shift ($<0{\cdot}004$ Å) of the irons on the last cycle. The bias may result from the relatively low observation to parameter ratio (approx. 3 : 1), which, especially in combination with the diagonal-matrix approximation, may allow other parameters initially to partially "take-up" the large shift applied to the irons, with only extremely slow convergence toward their true values. The phenomenon was not investigated further, since the iron bias was small compared to the estimated error in the final (restrained) model.

3. Results and Discussion

(a) *Intermolecular contacts and asymmetry*

The unique intermolecular contacts of the $P2_1$ crystal structure of haemoglobin are listed in Table 2. There are three unique extensive regions of contact (each including 29 to 48 atom pairs) and three less extensive regions (each including only 4 to 9 atom pairs); each molecule is thus in contact with 12 neighbours, as in a structure of close-packed spheres. Intermolecular contacts are asymmetric with respect to the molecular dyad; even in the two cases of quasi-equivalence (namely

TABLE 2

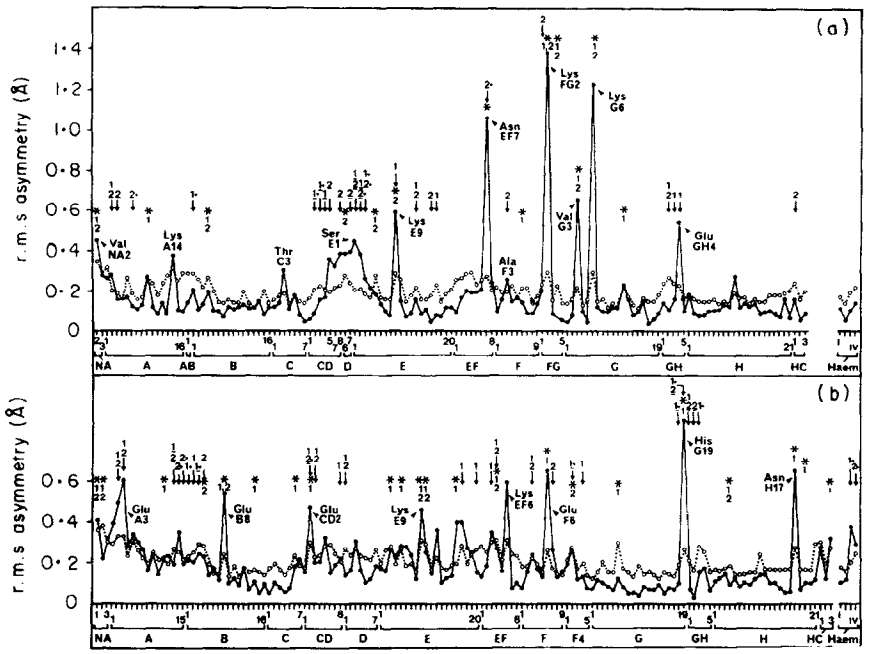
Unique intermolecular contacts of haemoglobin in $P2_1$ *crystals*[a]

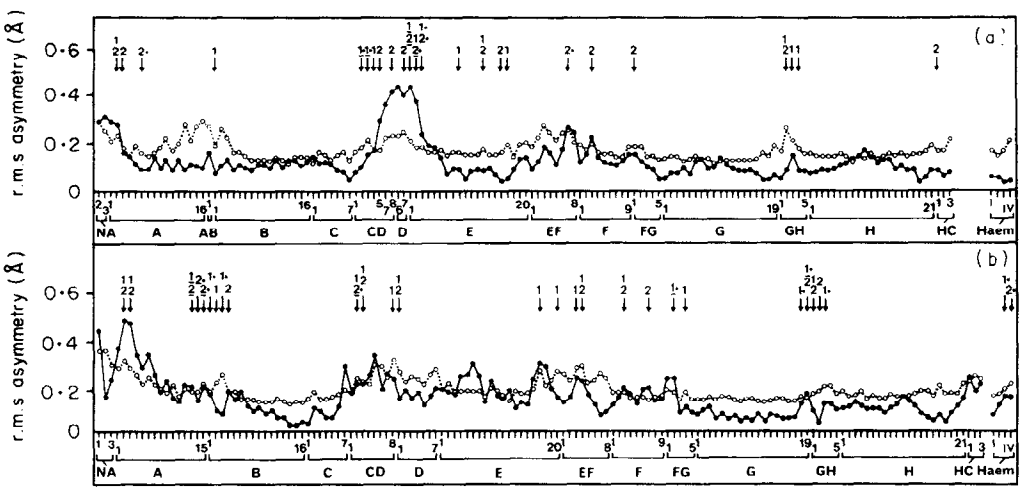
Transformation[b]	Residues of central molecules	No. atom pairs	Residues of contacting molecules
$x+1, y, z$	B1α_1	2	B4β_2
	GH2, GH3 α_1	12	G19, GH1, GH2 β_2
	GH4 α_1	1	GH2 α_2
	G18, G19, GH1, GH3 β_1	14	A13-A15 β_2
		29	
$x, y, z-1$	CD2-CD4 α_1	11	FG2, FG4 β_1
	E1-E3, E9 α_1	15	CD2, CD3, FG2, FG4, haem β_1
	CD2, CD3 β_2	22	CD5, CD8, D7, E1-E3 α_2
		48	
$1-x, y+\frac{1}{2}, 2-z$	A2α_1	5	F3 β_1
	CD8, D1 β_1	5	EF3. EF4 β_1
	EF7, F3 α_2	23	A13, B1-B3 β_1
	FG1, HC1 α_2	5	E17, E20 β_1
		38	
$-x, y+\frac{1}{2}, 1-z$	A2, A3, A6 α_2	6	F3, F7, haem β_2
	D1 β_2	3	EF4 β_2
		9	
$-x, y+\frac{1}{2}, -z$	E13, E16 α_2	8	A2, A3 β_2
$1-x, y+\frac{1}{2}, 1-z$	E13, E17 α_1	4	A2, A3 β_1
		136	

[a] All interatomic distances $\leqq 4{\cdot}0$ Å counted as contacts; water molecules disregarded.

[b] Crystallographic co-ordinates of contacting molecule equivalent to x, y, z of central molecule. For each contact, the roles of central and contacting molecule are reversed when the inverse transformation is applied.

the intermolecular contacts between E13α_1–A2β_1 and E13α_2–A2β_2 and between D1β_1–EF4β_1 and DIβ_2–EF4β_2) the contacts of the $\alpha_1\beta_1$ and $\alpha_2\beta_2$ dimers differ significantly in detail. It should be possible, therefore, to assess the extent to which intermolecular contacts distort molecular structure by examining to what degree they are correlated with molecular asymmetry. Figure 1 is a plot of r.m.s. asymmetry, by residue, with intermolecular contacts indicated. The Figure also plots the r.m.s. estimated positional error by residue and indicates those residues lying in regions of poorly resolved or ambiguous electron density in the $2F_{obs}-F_{calc}$ map. Figure 2 is a similar plot for main-chain atoms only. As indicated in the Figure legend, asymmetry at a residue may be regarded as statistically significant where it exceeds the r.m.s. estimated error. Examination of the Figures indicates that distortion of the molecule, as judged by molecular asymmetry, may indeed be caused by intermolecular contacts but that the effect is highly localized. From Figure 1(a) it may be seen that in the α subunit there is only one region, of about eight residues in the region CD5 to E3, of extended significant asymmetry; this appears to be clearly related to the high density of





FIGS 1 and 2. Molecular asymmetry of haemoglobin. Fig. 1 (above): all atoms included; Fig. 2 (below); main-chain atoms only. For each Figure, (a) shows the alpha chain and (b) the beta chain.

Open circles: r.m.s. estimated error, $\sigma p = 1{\cdot}7$ times value from diagonal least squares matrix (see Materials and Methods section (c)). Filled circles: r.m.s. asymmetry, by residue, defined as the r.m.s. distance between atomic positions in one dimer and the corresponding symmetry averaged positions; molecular dyad determined by least squares fitting with all atoms included at equal weight. Arrows, with numerals above denoting $\alpha_1\beta_1$ or $\alpha_2\beta_2$, mark residues in intermolecular contacts (interatomic distances $\leq 4{\cdot}0$ Å counted as contacts); with a dot (.) to indicate intermolecular hydrogen bond, and underscoring to denote main-chain atom of central molecule in the contact. Asterisks (*), with numerals below to indicate $\alpha_1\beta_1$ and $\alpha_2\beta_2$ dimers, denote residues where weak or ambiguous density was noted during examination of the $2F_{obs} - F_{calc}$ electron density maps.

If asymmetry results entirely from random errors in atomic positions that are uncorrelated between the $\alpha_1\beta_1$ and $\alpha_2\beta_2$ dimers (but not necessarily uncorrelated within each residue), then the expected value of r.m.s. asymmetry is $1/\sqrt{2}$ times the r.m.s. σp. R.m.s. asymmetry may reasonably be regarded as statistically significant for residues where r.m.s. asymmetry exceeds r.m.s. σp; the probability that r.m.s. asymmetry will exceed r.m.s. σp by chance alone depends upon the number of atoms in a residue and the degree of correlation between errors in the co-ordinates of different atoms within a residue and ranges from about 0·1 (marginal statistical significance) if co-ordinate errors are totally correlated within a residue down to less than 0·01 for residues with more than 6 atoms if co-ordinate errors are independent within a residue (chi-square distribution assumed).

intermolecular contacts, but the region is bounded by regions where asymmetry is small in comparison to error. There are also about ten individual residues showing significant asymmetry, all of which are bounded by regions of low asymmetry. Of these, at least two (at Ala F3 and Glu GH4) seem to be clearly related to intermolecular contacts; the remainder are either associated with regions of ambiguity in the electron density map (NA2, E9, EF7, FG2, G3 and G6) or have no evident cause (C3 and H10) and may possibly be due to random accumulation of errors. Several intermolecular contacts cause no significant asymmetry, most notably those at CD2 and CD3, which include main-chain hydrogen bonds. Figure 2(a) shows that the asymmetry of the CD5 to E3 region affects also the main-chain in that region and also indicates marginally significant asymmetry near the amino terminus of the α chain.

The situation for the β chain is similar, though somewhat less clear. It could be argued, for example, that the rather high (though statistically only marginally significant) asymmetry in the region A4 to A13 might be influenced by intermolecular contacts at either end of the region, but the very close correlation of asymmetry with positional error suggests that underestimation of positional error in the A4 to A13 region is a more likely cause of the apparent statistical significance of asymmetry in the region. Thus, the conclusion that distortions caused by intermolecular contacts are highly localized appears to hold for both the α and β chains of haemoglobin.

(b) *The haem environments*

The stereochemical relationship between the iron, N_ε and the haem of Hb is presented in comparison with that of several other structures in Table 3. (See Fig. 3 for an illustration of the haem structure of Hb and for explanation of the numbering of pyrrole rings.) The distances of the iron from the haem plane ($Fe–P_{haem}$) differ slightly in the α and β subunits of haemoglobin (whereas in the 2·5 Å model only a common value 0·6(1) could be determined; Fermi, 1975); as in HbO_2, the α iron is further from the haem plane than is the β iron. However, the distances of the α and β irons to the mean N(porphyrin) planes ($Fe–P_N$) are the same within experimental error. The difference between the α and β $Fe–P_{haem}$ distance thus results mainly from differing degrees of doming of the α and β haems, the doming parameter $P_N–P_C$ being 0·06 Å greater in α than in β. The $Fe–P_N$ distance and the $Fe–N_\varepsilon$ and Fe–N(porphyrin) bond lengths in Hb are the same within experimental error as the corresponding values for Mb and 5-co-ordinated model compounds. Our measurements of mean Fe–N(porphyrin) bond lengths also agree with those obtained in haemoglobin by extended X-ray fluorescence spectroscopy (EXAFS), which yield a mean value of 2·06(1) (Eisenberger *et al.*, 1978; Perutz *et al.*, 1982).

Several authors (for a review, see Perutz, 1979) have suggested that the low oxygen affinity of T-state haemoglobin (in comparison with that of R-state haemoglobin, myoglobin and free chains) derives not from any abnormality of the iron–nitrogen geometry but from haem–globin interactions that hinder the movement of the iron into the P_N plane upon oxygenation. Our finding that the

TABLE 3

Haem stereochemistry of haemoglobin, myoglobin and 5-co-ordinated model compounds[a]

	Distances to/between mean planes[b]					Bond lengths	
	$Fe–P_{haem}$	$Fe–P_N$	$N_\varepsilon–P_{haem}$	$P_N–P_C$	$P_{porph}–P_{haem}$	$Fe–N_\varepsilon$	$Fe–N_{porph}$ (mean)
Hb[c]							
α mean	0·58(3)	0·40(5)	2·72(6)	0·16(6)	0·05(3)	2·16(6)	2·08(3)
	0·59	0·41	2·76	0·16	0·05	2·20	2·14
β mean	0·50(3)	0·36(5)	2·58(6)	0·10(6)	0·06(3)	2·09(6)	2·05(3)
	0·48	0·36	2·58	0·08	0·06	2·10	2·04
αβ mean		0·38(4)		0·13(4)	0·05(2)	2·12(4)	2·06(2)
		0·38		0·12	0·05	2·15	2·09
r.m.s. asymmetry[d]	0·02	0·04	0·04	0·03	0·01	0·04	0·02
HbO_2[e]							
α	0·16(8)	0·12(8)	2·1(1)	0·04	0·01	1·94(9)	1·99(5)
β	0·00(8)	−0·11(8)	2·1(1)	0·06	0·06	2·07(9)	1·96(6)
Mb[f]	0·47	0·42	2·67	(—)	(—)	2·22	2·03(10)
MbO_2[f]	0·22(3)	0·18(3)	2·28(6)	0·00	0·03	2·07(6)	1·95(6)
Fe(TPP)(2-MeIm)[g]	—	0·42	—	0·13	—	2·161(5)	2·086(4)
Fe(TpivPP)(2-MeIm)[h]	—	0·399	—	0·03	—	2·095(5)	2·072(5)

[a] Distances in Å, figures in parentheses are standard deviation of last digit, where available.

[b] Abbreviations of mean planes: P_{haem}, porphyrin plus first side-chain atoms (32 atoms total); P_{porph}, porphyrin plane (24 atoms); P_N, porphyrin nitrogens; P_C, porphyrin carbons. Distances positive toward proximal side.

[c] Mean of α_1 and α_2 (resp. β_1 and β_2) haem values from unaveraged final model; standard deviations apply to the listed mean values and are calculated from co-ordinate errors (see Table 1). Overall mean listed where α–β difference is statistically insignificant. Values from model refined without restraints are shown below the above; the errors for the unrestrained model are about 25% greater than those of the final model.

[d] R.m.s. difference between value for each haem and the listed α or β mean values.

[e] Shaanan (1983).

[f] Phillips (1980).

[g] Hoard (1975).

[h] Jameson *et al.* (1980).

iron–nitrogen geometries of Hb and Mb are identical within experimental error confirms this idea. One factor, based on the 2·5 Å haemoglobin model, that has been implicated in opposing movement of the iron towards P_N is steric repulsion between C_ε of His F8 and N of pyrrole IV, resulting from asymmetric placement of His F8 (i.e. of the Cδ–Cε line) relative to the haem plane; the latter is constrained by the geometric relationship between the haem and helix F8, which in turn is dependent upon quaternary structure (Gelin & Karplus, 1979; Baldwin & Chothia, 1979). As may be seen from Table 4, the present structure confirms the asymmetry of $C_\delta–C_\varepsilon$ relative to the haem plane (about 11 deg.) and the relatively close contact distance (3·1 to 3·3 Å) between C_ε and N(IV). It should be noted, however, that the C_ε–N(IV) distance is about 0·1 to 0·3 Å shorter in HbO_2 than in Hb, which is only a little less than the difference of 0·3 to 0·5 Å between the $Fe–P_N$ distances in Hb and HbO_2. Thus only a weak argument can be made for an

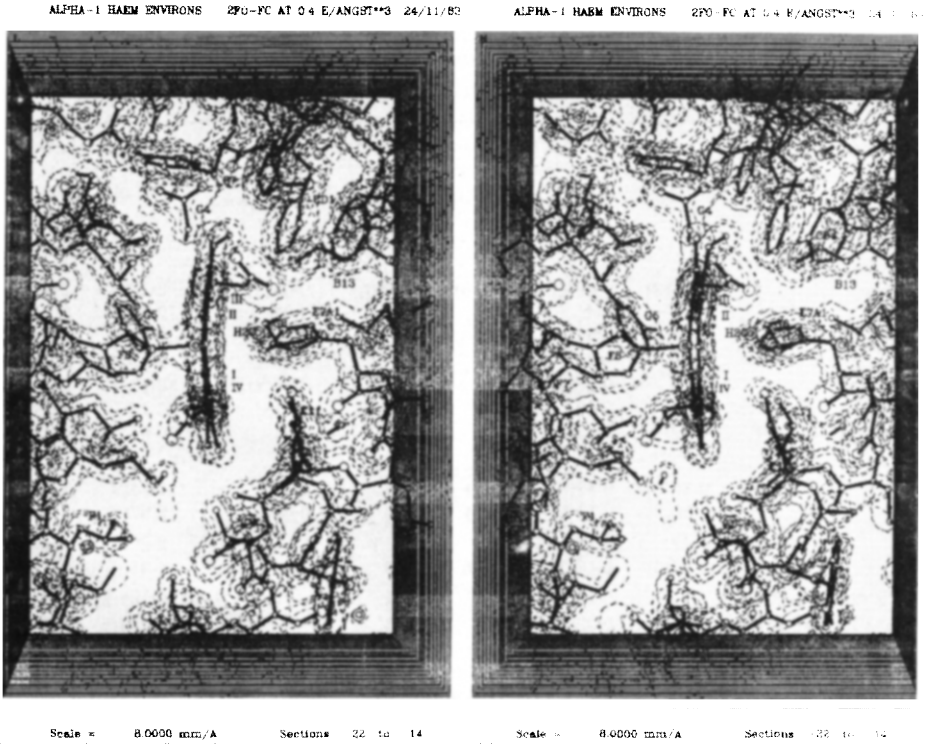


FIG. 3. Stereo view of the α_1 haem environs of Hb in the final (unaveraged) $2F_{obs} - F_{calc}$ electron density map. Only the single contour at 0·4 e/Å^3 is shown. The roman numeral notation for the pyrrole rings (I to IV) is indicated on the Figure. The axis of folding of the haem (see the text) is along the line of view. The Figure is intended only to illustrate the quality of the map; hence, the other 3 haems are not illustrated. For purposes of visualization of the structure, previously published illustrations based on the 2·5 Å model are adequate (Fermi, 1975; Fermi & Perutz, 1981; Shaanan, 1983).

important role of steric hindrance by C_ε–N(IV) in reducing the oxygen affinity of the T-state. Rather than being due to a single short interatomic contact, low oxygen affinity of the T-state is likely to arise from constraints by the globin on the geometry of the entire haem complex; this view appears to be in agreement with recent theoretical studies of Hb (Gelin *et al.*, 1983). The blockage of the ligand site of the β subunit of Hb by Val E11 found by Bolton & Perutz (1970) is confirmed in the present structure (Table 4).

Baldwin & Chothia (1979) attempted to explain the allosteric mechanism of haemoglobin by analysing differences between the then-available R structures, namely HbCO and methaemoglobin, and the 2·5 Å model of deoxyhaemoglobin. We have re-calculated many of the quantitative differences between the R and T structures analysed by Baldwin & Chothia, using the present Hb model (the re-regularized, symmetry averaged model was used for this purpose) and the high resolution model of HbO_2 (Shaanan, 1983) with results presented in Table 5. It can be seen from the Table that the shifts between Hb and HbO_2 differ but little from those between Hb(2·5 Å) and HbCO. In particular, the shifts of the E and F

TABLE 4

Distances and angles in the haems and their environs in Hb and HbO_2[a]

	Hb			HbO_2[b]	
	α mean	β mean	r.m.s. asymmetry	α	β
A. *Angles (deg.) between*					
Fe → N_ε and haem normal	8(2) 9	5(2) 6	1·2	3	5
F8 imidazole plane and haem normal	3(1) 3	2(1) 1	0·8	0	4
F8 imidazole plane and haem N(II)–N(IV)	18(1) 18	24(1) 23	0·5	11	27
C_δ → C_ε(F8) and haem normal	77(3) 78	79(3) 79	2·6	84	91
B. *Distance (Å) between*					
N_{porph} Ct[c] (mean)	2·04(3) 2·10	2·02(3) 2·01	0·02	1·98	1·95
N_ε–N_{porph} (mean)	3·26(4) 3·34	3·17(4) 3·18	0·03	2·9	2·8
C_ε(F8)–haem N(IV)	3·26(8) 3·29	3·12(8) 3·17	0·02	2·9(2)	3·0(2)
C_δ(F8)–haem N(II)	3·78(8) 3·76	3·74(8) 3·72	0·06	3·2(2)	3·1(2)
Ligand site[d] and $C\gamma_2$ Val E11	3·5(1) 3·6	2·2(1) 2·2	0·07	3·2(1)	3·4(2)
H_2O (haem-linked)	1·7(1)		0·07		

[a] See footnotes [a], [c] and [d] to Table 3, which also apply to this Table.
[b] Shaanan (1983).
[c] Ct, N(porphyrin) centre.
[d] Taken to lie on the haem axis, at distances from the haem plane of 1·5 Å and 1·8 Å, respectively, at the α and β haems (See Shaanan, 1983).

helices relative to the haem plane, which are most important to the mechanism of co-operativity, are nearly identical.

Nagai & Kitagawa (1980) measured the stretching frequencies of the Fe–N_ε bond in the α and β subunits of R-state and T-state haemoglobin by resonance Raman spectroscopy and concluded that in Hb the bond in the α subunit is under greater stress than it is in the β subunit. They estimated that the difference in stress would be equivalent to a bond length difference (α minus β) of about 0·02 Å. Our measurement of the Fe–N_ε bond length itself is not sufficiently accurate to detect such a small difference: although we find a positive α-minus-β bond-length difference, the difference is not statistically significant (Table 3). Other structural parameters that might affect tension in the Fe–N_ε bond yield ambiguous results (Table 4): the deviation of Fe–N_ε bond from the haem normal is greater at α than at β (the difference being of marginal statistical significance) as would be expected; but the mean N_ε–N(porphyrin) distances, which may contribute to strain in the iron–nitrogen complex (Warshel, 1977), are longer at α than at β (the

TABLE 5

Helix shifts in going from unliganded to liganded haemoglobin

	α subunit		β subunit	
Helix	Hb–HbO_2	Hb(2·5 Å)–HbCO	Hb–HbO_2	Hb(2·5 Å)–HbCO
After fitting haems				
E helix: magnitude	0·4	0·1	2·0	2·3
x	−0·4	0·0	−0·6	−0·5
y	0·0	0·0	1·3	1·4
z	0·0	0·1	1·4	1·7
F helix: magnitude	1·3	1·1	0·7	0·8
x	0·2	0·4	0·2	0·2
y	−0·2	−0·1	−0·4	−0·4
z	−1·3	−1·0	−0·6	−0·7
After fitting $\alpha_1\beta_1$ interface (magnitude)				
NA	1·3	1·4	1·7	1·4
A	0·4	0·7	0·6	0·6
B	0·3	0·3	0·6	0·4
C	0·3	0·5	0·6	0·6
CD	0·9	1·1	0·6	0·7
D			0·5	0·5
E	0·7	0·7	1·0	1·1
EF	1·0	1·3	0·9	1·3
F	1·6	1·6	1·6	1·3
FG	1·7	1·3	1·9	1·9
G	0·6	0·5	0·7	0·4
GH	0·6	1·2	0·2	0·3
H	0·6	0·5	0·5	0·4
HC	3·3	2·5	3·2	2·5
haem	0·9	0·6	1·7	1·6

Hb to HbO_2 shifts in Å from the re-regularized symmetry averaged Hb model (Results and Discussion) and the HbO_2 co-ordinates of Shaanan (1983). Hb (2·5 Å model) to HbCO shifts from Baldwin & Chothia (1979), which should be consulted for the fitting methods employed (e.g. only main-chain included). For shifts relative to haem: *x*, along haem normal, positive toward distal side; *y*, haem centre toward methene III-IV; *z*, haem centre toward methene I-IV.

difference again only marginally significant statistically), contrary to expectations if the α complex were more strained than the β complex due to short contacts between the imidazole carbons of His F8 and the porphyrin nitrogens. Rather, the stretching of the Fe–N_ε bond of the α haems observed by resonance Raman spectroscopy must be due to the clamping of the porphyrin by its contacts with globin side-chains (see Table 7).

The detailed structure of the haems is described in Table 6, in terms of the angles between the normals to the mean pyrrole and mean haem planes. In principle, eight parameters per haem are required to describe the tilt and twist of the four pyrroles, but an adequate description can be made by fitting only two parameters each to the α and β haems, as follows. At the α haem, the four pyrroles are assumed to be uniformly tilted (normals displaced toward the haem centre) through an angle *t*, and additionally the haem is assumed to be folded about an

TABLE 6

Pyrrole–haem angles (deg.)[a]

Pyrrole[b]		Observed				Predicted[c]			
		α mean		β mean		α mean		β mean	
		Restricted	Unrestricted	Restricted	Unrestricted	Restricted	Unrestricted	Restricted	Unrestricted
I(2)	Tilt[d]	5·3	4·2	−0·6	2·4	6·2	6·7	0·4	0·2
	Twist[e]	2·1	1·2	−1·0	−0·2	2·8	2·3	0	0
II(3)	Tilt	8·2	10·5	6·9	6·6	6·2	6·7	8·2	8·2
	Twist	−2·6	−3·5	1·6	3·2	−2·8	−2·3	0	0
III(4)	Tilt	5·8	5·7	1·4	−2·0	6·2	6·7	0·4	0·2
	Twist	3·6	2·2	−0·8	−2·5	−2·8	−2·3	0	0
IV(1)	Tilt	5·4	6·2	9·6	9·9	6·2	6·7	8·2	8·2
	Twist	−2·9	−4·6	−0·5	−1·4	−2·8	−2·3	0	0
r.m.s. residual[f] (observed minus predicted)									
	Tilt	1·7	2·7	2·0	2·3				
	Twist	0·9	1·9	1·4	3·1				
	Overall	1·4	2·3	1·7	2·7				
	Overall, αβ r.m.s.			1·6 restrained 2·5 unrestrained					

[a] Values listed for α and β haems are means of the values from the unaveraged final model. The estimated error in the listed observed values is about 2 deg.; the r.m.s. difference between the values at individual haems and the listed mean values is 1·2 deg.

[b] Nomenclature of Fermi & Perutz (1981) with that of Phillips (1980) in parentheses.

[c] See the text.

[d] Angle between haem normal and projection of pyrrole on plane through haem normal and pyrrole N; positive direction pushing N toward proximal side.

[e] Right-handed rotation of pyrrole plane about vector from pyrrole N to haem centre.

[f] Calculated over all 4 haems; i.e. includes a component due to asymmetry between $\alpha_1\beta_1$ and $\alpha_2\beta_2$.

TABLE 7

Haem contacts[a]

Residue[b]	α subunit		β subunit	
B13 (D)	Met	>4·0	Leu	3·9 (1) II: V
C7 (E)	Tyr	3·4 (1) III: M	Phe	3·7 (4) II: V, III: M
CD1 (D)	Phe	3·5 (8) III: R, M, II/III	Phe	3·6 (5) III: M, IV: M, II/III
CD3 (E)	His	2·8 (4) III: P*	Ser	>4·0
CD4 (D)	Phe	3·4 (4), 3·8 (4) III: P	Phe	>4·0
E7 (D)	His	3·2 (9) I: R	His	3·2 (14), 3·2 (11) I: R, IV: R, III/IV
E10 (D)	Lys	3·4 (5), 3·6 (1) IV: M	Lys	3·8 (4) IV: P 3·0 (5) IV: P*
E11 (D)	Val	4·0 (1) I/IV	Val	3·6 (8) I: R, V, M
E14 (D)	Ala	>4·0	Ala	3·8 (2) I: M, IV: M
F4 (P)	Leu	3·8 (2) IV: M 3·7 (5) IV: M, P, I/IV	Phe	3·9 (3) IV: M >4·0
F7 (P)	Leu	3·4 (2) IV: P	Leu	3·8 (1), 3·5 (2) IV: P
F8 (P)	His	3·1 (15)I-IV: R, I/IV	His	3·0 (18)I-IV: R, I/IV
FG3 (P)	Leu	3·3 (7) III: R, P, III/IV	Leu	3·5 (5) III: R, P, III/IV
FG5 (P)	Val	3·5 (6) II: R, V, II/III	Val	3·9 (2) II: V, II/III
G4 (P)	Asn	3·4 (4) II: V, M	Asn	3·5 (3) II: M
G5 (P)	Phe	3·5 (5) II: M, I: R, V, I/II	Phe	3·6 (5) II: R, M, I/II
G8 (D)	Leu	3·5 (2) I: R, I/II 3·8 (1) I/II	Leu	3·4 (3) I: R, V
H19 (P)	Leu	3·5 (4) I: R, V, M	Leu	3·4 (3) I: R, V

[a] Interatomic distances ≤4·0 Å counted as contacts. The Table gives the minimum contact distances in Å (followed by the number of contacting pairs in parentheses), followed by the number of the pyrrole(s) with which contact is made, with abbreviations: R, ring; M, V, P, methyl, vinyl, propionic side-chain; I/II, methene between rings I and II, etc.; asterisk (*) denotes hydrogen bond. Separate listings are shown for α_1 and α_2 (β_1 and β_2, respectively) when contact distance differs by more than 0·2 Å or number of contacts differs by more than 2, otherwise distance is minimum and number contacts is maximum of α_1 and α_2 (β_1 and β_2, respectively).

[b] Letters in parentheses denote: D, distal; P, proximal; E, edge-on contact.

axis passing between the I/II and III/IV methene carbons (apex of folding on proximal side) through an angle f (for notation see Fig. 3); least squares fitting yields the values $t = 3{\cdot}4$ degrees and $f = 3{\cdot}9$ degrees, resulting in predicted tilt and twist angles as listed in the Table. For the β haem, it is assumed that there is no folding of the haem and no twisting of the pyrroles, and that pyrroles opposite to the haem centre are tilted by the same amount, $t_{I,III}$ for one pair and $t_{II,IV}$ for the other, the fitted values being $t_{I,III} = 0{\cdot}4$ degrees and $t_{II,IV} = 8{\cdot}2$ degrees. As indicated in Table 6, the r.m.s. residual between the observed and predicted pyrrole angular parameters of 1·6 degrees (2·5 degrees for the unrestrained model) is comparable to the estimated error of about 2 degrees in each parameter, which indicates that any additional irregularity of the pyrrole planes is small in comparison to co-ordinate accuracy. The pyrrole geometry described above must result from globin–haem contacts rather than through direct influence of the F8 imidazoles since the unequal tilting of the β pyrroles is opposite to that which

might be expected from steric repulsion between the imidazole and the haem nitrogens. However, no simple explanation of pyrrole geometry is evident from an examination of haem contacts (see Table 7).

Haem contacts are listed in Table 7. Apart from the addition of more detailed information, the Table is similar to those previously published (Fermi, 1975; Fermi & Perutz, 1981). An exception is that we now find propionate IV β_2 (but not β_1) to form a hydrogen bond with N_ζ of Lys E10, a bond that was previously obscured by symmetry averaging; propionate IV β_2 also forms an intermolecular hydrogen bond (with Thr A6α_2), while propionate II β_2 is free. Propionate III β_1 forms an intermolecular hydrogen bond (with Ser E1α_1) and is also linked by a water molecule to N_δ of His E7. The propionate side-chains of the α_1 and α_2 haems are not involved in intermolecular contacts and are quite symmetric. The existence of a hydrogen bond between propionate III α and N_δ of His CD3, included in the initial 2·5 Å model but cast in doubt by further refinement against the same data (Fermi *et al.*, 1982) is now reaffirmed. Propionate III α is also linked to a water molecule that is hydrogen-bonded to N_ζ of Lys E10 and to the carbonyl of E7, while propionate IV α is free.

We thank P. Marin and his collaborators from the Laboratoire de l'Accélérateur Linéaire at Orsay who operated the storage ring DCI, and Anne Graeme-Barber for assistance with densitometry.

REFERENCES

Arndt, U. W. & Wonacott, A. J. (1977). Editors of *The Rotation Method in Crystallography*, North Holland Publ. Co., Amsterdam.

Baldwin, J. & Chothia, C. (1979). *J. Mol. Biol.* **129**, 175–220.

Bolton, W. & Perutz, M. F. (1970). *Nature (London)*, **228**, 551–552.

Chothia, C., Wodak, S. & Janin, J. (1976). *Proc. Nat. Acad. Sci., U.S.A.* **73**, 3793–3797.

Diamond, R. D. (1978). In *Symposium on Biomolecular Structure, Conformation and Evolution* (Srinivasan, R., ed.), Pergamon, Oxford.

Eisenberger, P., Shulman, R. G., Kincaid, B. M., Brown, G. S. & Ogawa, S. (1978). *Nature (London)*, **274**, 30–34.

Fermi, G. (1975). *J. Mol. Biol.* **97**, 237–256.

Fermi, G. & Perutz, M. F. (1981). *Haemoglobin and Myoglobin: Atlas of Molecular Structures in Biology* (Phillips, D. C. & Richards, F. M., eds), vol. 2, Oxford University Press.

Fermi, G., Perutz, M. F., Dickinson, L. C. & Chien, J. C. W. (1982). *J. Mol. Biol.* **155**, 495–505.

Fox, G. C. & Holmes, K. C. (1966). *Acta Crystallogr.* **20**, 886–891.

French, S. & Wilson, K. (1978). *Acta Crystallogr. sect. A*, **34**, 517–525.

Gelin, B. R. & Karplus, M. (1979). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 801–805.

Gelin, B. R., Lee, A. W.-M. & Karplus, M. (1983). *J. Mol. Biol.* **171**, 489–559.

Hoard, J. L. (1975). In *Porphyrins and Metalloporphyrins* (Smith, K. M., ed.), pp. 317–380, Elsevier, Amsterdam.

Jack, A. & Levitt, M. (1978). *Acta Crystallogr. sect. A*, **34**, 931–935.

Jameson, G. B., Molinaro, F. S., Ibers, J. A., Collman, J. P., Brauman, J. I., Rose, E. & Suslick, K. S. (1980). *J. Amer. Chem. Soc.* **102**, 3224–3237.

Kahn, R., Fourme, R., Gadet, A., Janin, J., Dumas, C. & Andre, D. (1982). *J. Appl. Crystallogr.* **15**, 330–337.

Luzzati, V. (1952). *Acta Crystallogr.* **5**, 802–810.

Mallett, J. F. W., Champness, J. N., Faruqi, A. R. & Gossling, T. H. (1977). *J. Phys. E. Sci.* **10**, 351.

Monod, J., Wyman, J. & Changeux, J. P. (1965). *J. Mol. Biol.* **12**, 88–118.
Nagai, K. & Kitagawa, T. (1980). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 2033–2037.
Phillips, S. E. V. (1980). *J. Mol. Biol.* **142**, 531–554.
Perutz, M. F. (1968). *J. Crystal Growth*, **2**, 54–56.
Perutz, M. F. (1970). *Nature (London)*, **228**, 726–739.
Perutz, M. F. (1979). *Annu. Rev. Biochem.* **48**, 327–386.
Perutz, M. F., Hasnain, S. S., Duke, P. J., Sessler, J. L. & Hahn, J. B. (1982). *Nature (London)*, **295**, 535–538.
Shaanan, B. (1983). *J. Mol. Biol.* **171**, 31–59.
Takano, T. (1977). *J. Mol. Biol.* **110**, 569–584.
Ten Eyck, L. & Arnone, A. (1976). *J. Mol. Biol.* **100**, 3–11.
Warshell, A. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 1789–1793.

Edited by J. C. Kendrew

Note added in proof: The 3 co-ordinate sets (final, symmetry-averaged, and unrestrained), observed and calculated structure amplitudes, and phases for Hb have been deposited with the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, N.Y. 11973, U.S.A., from which copies are available.