

## STRUCTURE OF HÆMOGLOBIN

A THREE-DIMENSIONAL FOURIER SYNTHESIS AT 5.5-Å. RESOLUTION, OBTAINED BY X-RAY ANALYSIS

By DR. M. F. PERUTZ, F.R.S., DR. M. G. ROSSMANN, ANN F. CULLIS, HILARY MUIRHEAD and DR. GEORG WILL

Medical Research Council Unit for Molecular Biology, Cavendish Laboratory, University of Cambridge AND

DR. A. C. T. NORTH

Medical Research Council External Staff, Davy Faraday Research Laboratory, Royal Institution, London, W.1

**V**ERTEBRATE hæmoglobin is a protein of molecular weight 67,000. Four of its 10,000 atoms are iron atoms which are combined with protoporphyrin to form four hæm groups. The remaining atoms are in four polypeptide chains of roughly equal size, which are identical in pairs<sup>1-3</sup>. Their amino-acid sequence is still largely unknown.

We have used horse oxy- or met-hæmoglobin because it crystallizes in a form especially suited for X-ray analysis, and employed the method of isomorphous replacement with heavy atoms to determine the phase angles of the diffracted rays<sup>4-7</sup>. The Fourier synthesis which we have calculated shows that hæmoglobin consists of four sub-units in a tetrahedral array and that each sub-unit closely resembles Kendrew's model of sperm whale myoglobin<sup>8</sup>. The four hæm groups lie in separate pockets on the surface of the molecule.

## Method of Analysis

Horse oxyhæmoglobin, crystallized from 1.9 *M* ammonium sulphate solution at pH 7, has the space group *C*2 with two molecules in the unit cell which lie on dyad axes<sup>9</sup>. In order to determine the phase angles of the 1,200 reflexions contained in the limiting sphere of 5.5 Å.<sup>-1</sup>, six different isomorphous heavy-atom compounds were used (ref. 9 and unpublished work). Intensities were measured photographically and by counter spectrometer (Arndt, U. W., and Phillips, D. C., unpublished work). The relative positions and shapes of the heavy-atom replacement groups were found by correlation functions based on Patterson methods<sup>10</sup> and refined by least squares

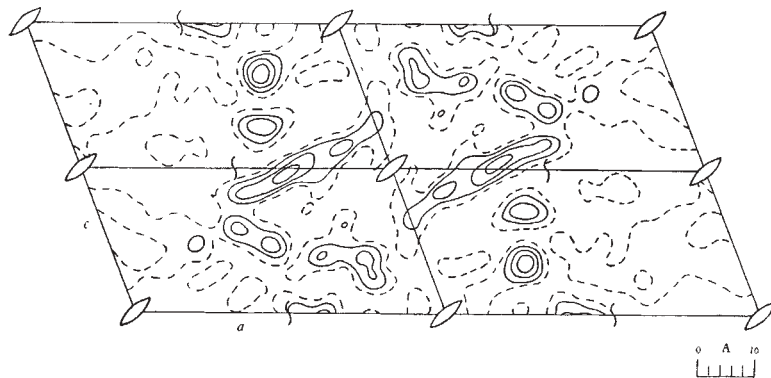


Fig. 1. Section at  $y = 1/32b$ . This cuts through the middle of the molecule on which the diagram is centred. 'Flat' areas indicating liquid appear on the left and right. Contours are drawn at intervals of 0.14 electron/Å<sup>3</sup>. The broken line marks 0.4 electron/Å<sup>3</sup>. Contours at lower levels are omitted.

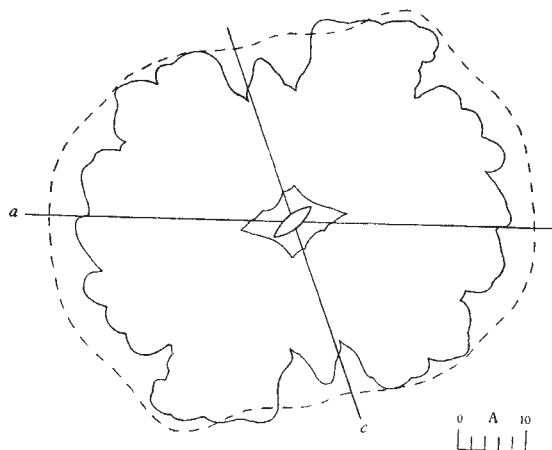
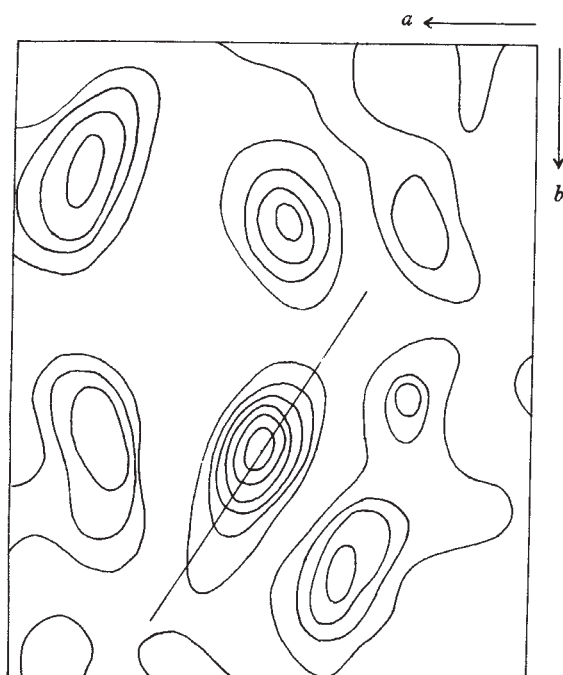


Fig. 2. External shape of the molecule. Full lines indicate the boundary derived from the contour at 0.54 electron/Å<sup>3</sup>. The broken line shows the boundary derived by Bragg and Perutz from two-dimensional data (ref. 13). Note the hole in the middle.

procedures. For each reflexion the structure amplitudes of all seven compounds were combined in an Argand diagram<sup>11</sup>, and the probability of the phase angle having a value  $\alpha$  was calculated for  $\alpha = 0, 5, 10, \dots, 355^\circ$ . The centroid of the probability distribution, plotted around a circle, was then chosen as the best vector  $F'$  in the Fourier synthesis<sup>12</sup>. The results were finally plotted on 32 contour maps showing the distribution of electron density in sections spaced 2 Å. apart normal to *b* (Fig. 1). The absolute configuration of the molecule was determined from anomalous dispersion<sup>4</sup>.

## External Shape of the Molecule

More than half the volume of the crystals is taken up by liquid of crystallization, which mainly fills the spaces between the molecules and shows up on the contour maps in the form of flat, featureless regions (Fig. 1). The outlines of the two molecules in the unit cell can be traced by following the boundaries between these regions and the continuous electron-dense regions described below. In Fig. 2 the outline of one molecule, traced from the periphery of the 0.54



Hæm 1 and 2

Fig. 3. Section at  $z = 1/4 c$  showing one of the two hæm groups in the asymmetric unit. The straight line indicates the hæm orientation derived from electron spin resonance (ref. 14). The lowest contour shown is at  $0.4 \text{ electron}/\text{Å}^3$

electron/ $\text{Å}^3$  contour, is seen in projection on the  $b$ -plane. To a first approximation it can be regarded as a spheroid with a length of  $64 \text{ Å}$ , a width of  $55 \text{ Å}$ , and a height of  $50 \text{ Å}$ , normal to the plane of the paper. Except for a slight shortening along  $a$ , this shape agrees with the earlier picture obtained from two-dimensional data, even including the dimple in the centre of the molecule (see Fig. 2 of ref. 12).

Positions of the Hæm and Sulphydryl Groups

Four peaks stand out from the rest, clearly representing the iron atoms with their surrounding porphyrin rings. A flattening of the peaks indicates the approximate orientation of the rings (Fig. 3). Fortunately, their exact orientation was already known from electron spin resonance<sup>14</sup>, and it only remained to assign the correct one of the four alternative orientations of the hæms to each of the electron density peaks. The results are shown in Fig. 4. The iron atoms lie at the corners of an irregular tetrahedron with distances of  $33.4$  and  $36.0 \text{ Å}$  between symmetrically related pairs. The closest approach between symmetrically unrelated iron atoms is  $25.2 \text{ Å}$ . (Table 1).

	Fe <sub>1</sub>	Fe <sub>2</sub>	S <sub>1</sub>
$x$	$-6.6 \text{ Å}$	$12.3 \text{ Å}$	$5 \text{ Å}$
$y$	$7.3 \text{ Å}$	$-10.7 \text{ Å}$	$10 \text{ Å}$
$z$	$13.1 \text{ Å}$	$18.2 \text{ Å}$	$16 \text{ Å}$
Fe <sub>1</sub> - Fe <sub>2</sub>	$33.4 \text{ Å}$		
Fe <sub>2</sub> - Fe <sub>4</sub>	$36.0 \text{ Å}$		
Fe <sub>1</sub> - Fe <sub>3</sub>	$25.2 \text{ Å}$		
Fe <sub>1</sub> - S <sub>1</sub>	$13 \text{ Å}$		
Fe <sub>2</sub> - S <sub>1</sub>	$21 \text{ Å}$		

Horse hæmoglobin contains four cysteine residues, but only two sulphydryls combine with mercury in the native protein<sup>5</sup>. From the positions of the mercury atoms we inferred that each of the two

sulphydryl groups is about  $13 \text{ Å}$  away from one iron atom and  $21 \text{ Å}$  away from another. (Fig. 4 and Table 1). The significance of this situation is discussed below.

Configuration of the Polypeptide Chains

The most prominent feature of the Fourier synthesis consists of more or less cylindrical clouds of high density, like the vapour trails of an aeroplane; they are curved to form intricate three-dimensional figures. Sections through various parts of these appear in Fig. 1. To build a model of the figures, we rolled out sheets of a thermo-setting plastic to the thickness of our sections on a scale of  $2 \text{ Å} = 1 \text{ cm}$ , and cut out the shape of each region on the contour maps where the density exceeds  $0.54 \text{ electron}/\text{Å}^3$  (this corresponds to the first full contour line in Fig. 1. From this section, for example, 14 shapes would be cut). The shapes were then assembled in accordance with their positions and heights in the different sections, and the hæm groups were attached in the appropriate orientation. The model was then baked to set it permanently. For comparison, a Fourier synthesis of sperm whale myoglobin was calculated at a resolution of  $5.5 \text{ Å}$ , using the new X-ray data of Kendrew *et al.*<sup>15</sup>, and a model of the electron density distribution was constructed by the method just described.

From the hæmoglobin Fourier synthesis there emerged four separate units which are identical in pairs. Fig. 5 shows one member of each pair together with the model of myoglobin on the left. In each unit the cloud of high density describes a complicated figure which, in the white unit (middle), can be traced from end to end by following the superimposed line. Except for two small gaps where the density sinks slightly below  $0.54 \text{ electron}/\text{Å}^3$ , the black unit (right) closely resembles the white one. There are several gaps interrupting the myoglobin model where, probably due to increased thermal motion, the electron density falls below the contour-level chosen for cutting the model. However, we know from the recent work of Kendrew and his collaborators<sup>15</sup> that

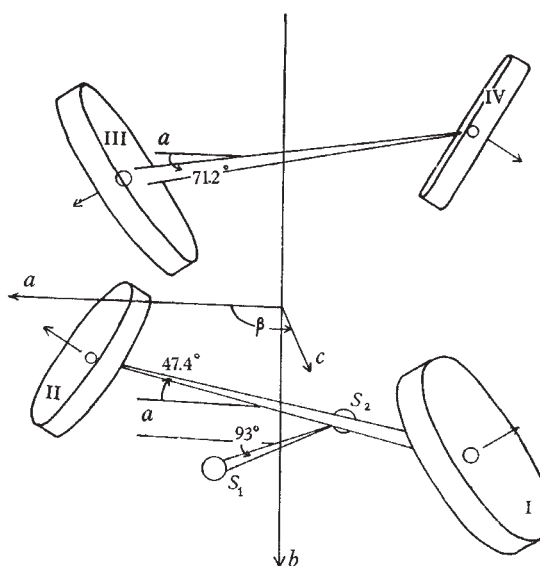


Fig. 4. Arrangement of hæm groups in hæmoglobin. Arrows indicate the reactive side of each group. The  $c$ -axis comes out of the paper towards the observer. This picture and all subsequent ones show the absolute configuration

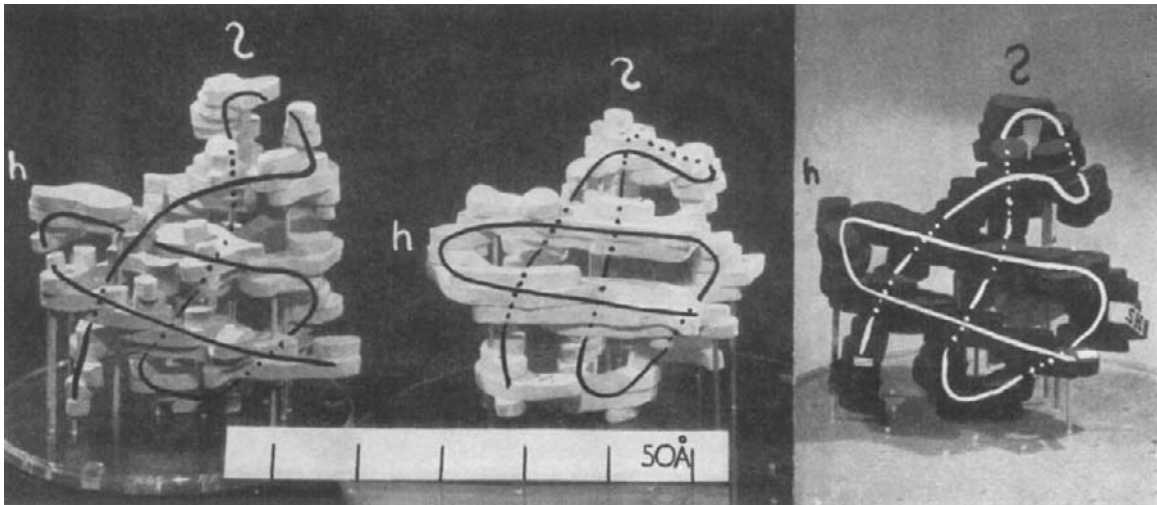


Fig. 5. Two different polypeptide chains in the asymmetric unit of hæmoglobin compared with myoglobin (left). The hæm groups are at the back of the chains

the gaps are bridged by a continuous polypeptide chain, and it is evident from Fig. 5 that apart from the gaps, the model has a configuration closely similar to the hæmoglobin units.

Clearly, the four tortuous clouds of high electron density in hæmoglobin represent the four polypeptide chains. The black and the white chains have similar, but not identical configurations. In the black chain the *S*-shaped bend at the top is more pronounced, the hæm group is lower, and the bend *h* sharper than in the white chain. These, however, are details. The

most important result is their resemblance to each other and to sperm whale myoglobin.

#### Arrangement of the Four Sub-units

The first step in the assembly of the molecule is the matching of each chain by its symmetrically related partner (Fig. 6). It will be noted that there is comparatively little contact between the members of each pair, suggesting rather tenuous linkages. In the next step the white pair is inverted and

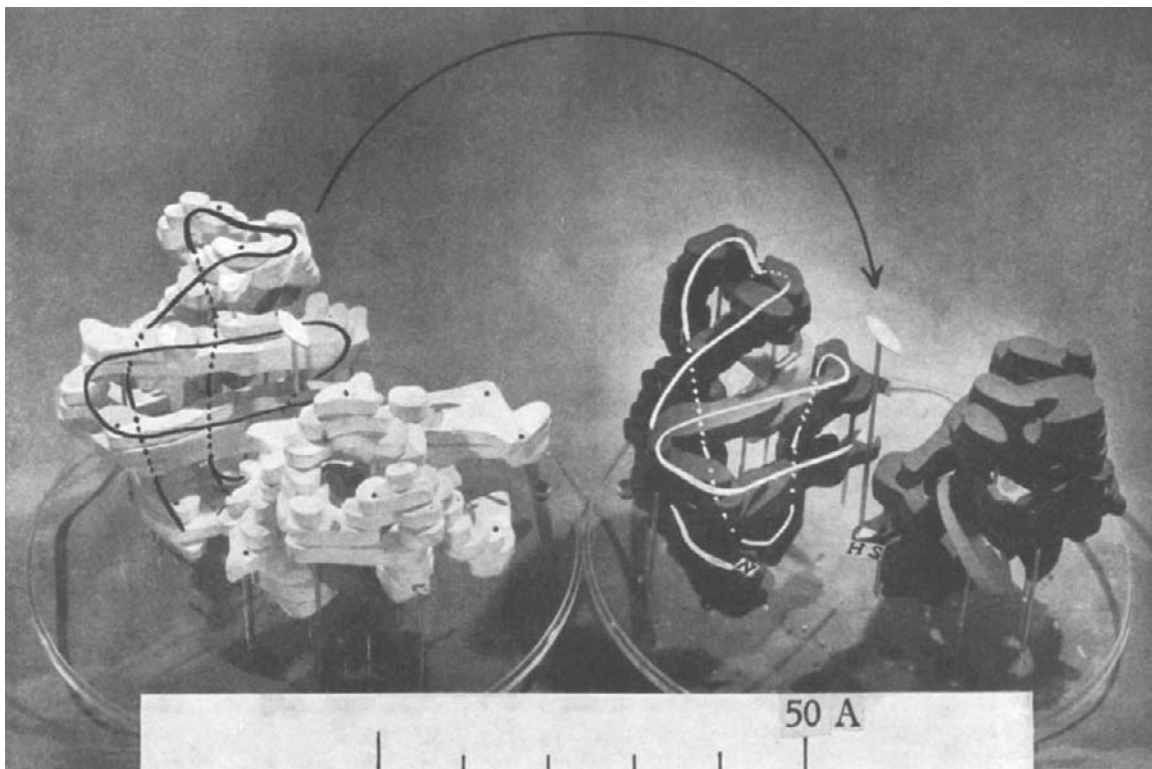


Fig. 6. Two pairs of chains symmetrically related by the dyad axis. The arrow shows how one pair is placed over the other to assemble the complete molecule

placed over the black pair as indicated by the arrow. Fig. 7 shows one white chain placed over the pair of black ones and Fig. 8 shows the molecule completely assembled. The resulting arrangement is tetrahedral and has almost, but not quite, the orthorhombic point group symmetry 222. It contains two 'pseudo dyads' which lie approximately at right angles to each other and to the true dyad, one emerging from the centre of Fig. 8 and the other from the centre of Fig. 9. This means that, to a first approximation, each sub-unit can be generated by a rotation of  $180^\circ$  from any of its neighbours. Figs. 7, 8 and 9 also show that the surface contours of the white chains exactly fit those of the black, so that there is a large area of contact between them. This structural complementarity is one of the most striking features of the molecule. Fig. 10 is a view down the true dyad axis and reveals a hole going right through the centre of the molecule, as was to be expected from the Fourier projection<sup>13</sup>. However, the van der Waals radii of the chains are much bigger than appears in the model, and little room may, in fact, be left for water or electrolytes to pass through. Fig. 10 also reveals a dimple at the top where the white chains meet. This is matched by a similar, but larger, hollow at the bottom where the black chains meet.

The hæm groups are seen to lie in separate pockets on the surface of the molecule (Fig. 8). Each pocket is formed by the folds in one of the polypeptide chains, which appears to make contact with the hæm group at four different points at least. The iron atoms in the neighbouring pockets formed by the black and the white chains are 25 Å. apart.

#### Information from the Fourier Synthesis of Myoglobin at 2-Å. Resolution (ref. 15)

Thanks to the similarity with myoglobin, the interpretation can be carried further than would have been possible on the basis of our results alone. Kendrew *et al.* have found that the straight stretches of rod indicated in Fig. 5 are  $\alpha$ -helices and that the N-terminal end of the chain is at the bottom left. The ends of the two hæmoglobin chains have been labelled *N* and *C* accordingly, as seen in Figs. 7 and 10. The myoglobin Fourier synthesis also reveals the hæm-linked amino-acid side-chain, probably histidine, on one side of the (ferric) iron atom and a small peak, probably representing a water molecule, on the other side. If this information is transferred to hæmoglobin, the reactive side of the hæm group is as indicated by the arrows in Fig. 4 and by the labels  $O_2$  in Figs. 7 and 8a.

Figs. 7 and 8b show the reactive sulphhydryl group (which is absent in myoglobin) to be attached to the portion of the black chain carrying the hæm-linked histidine. It is seen that the histidine and cysteine side-chains point in roughly opposite directions, one towards and the other away from the hæm group.

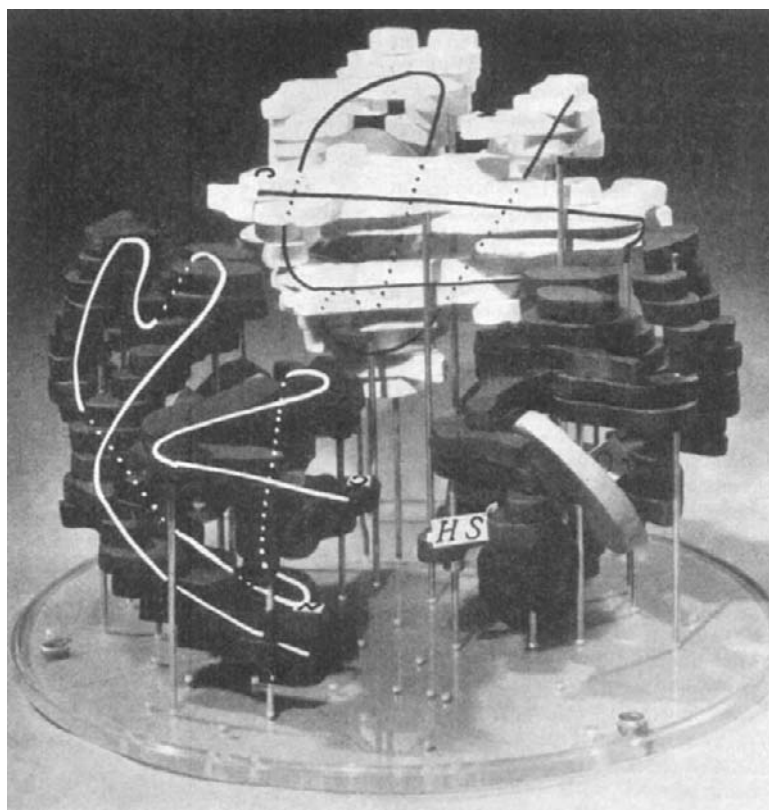


Fig. 7. Partially assembled molecule showing two black chains and one white

The sulphhydryl group may possibly be in contact with the loop of the white chain which lies below the hæm group on the left. This is the situation of one pair of sulphhydryl groups in the molecule. The second pair is probably attached to the white chains, but it is unreactive in the native protein and its position is still unknown.

#### Reliability of Results

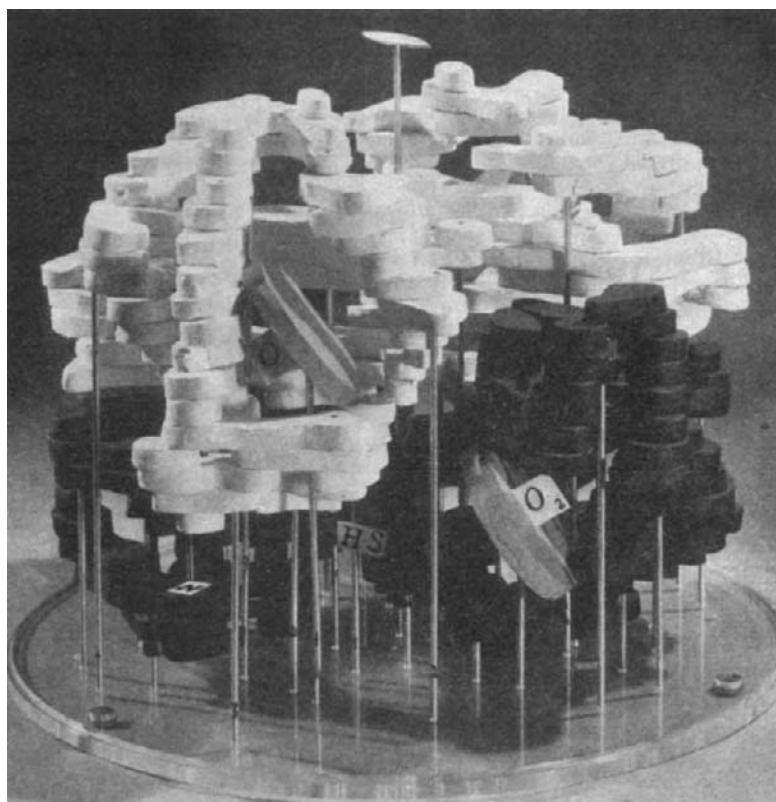
The method of isomorphous replacement makes no assumptions about the structure of the protein, and its results suffer from none of the ambiguities which bedevil the interpretation of vector maps. Ideally, the parent protein, in combination with two different isomorphous heavy-atom compounds, should give accurate phase angles and electron density maps which are free from all except series termination errors. In practice, errors and uncertainties arise from several sources and must be minimized by using more than two heavy-atom compounds. In our case it was possible to estimate the accuracy of the phase angle for each reflexion from the measure of agreement between the angles indicated by the six different compounds. From the standard error in the vector  $F$ , averaged over all reflexions, the standard error in the final electron density was calculated as 0.12 electron/Å.<sup>3</sup>, which amounts to 0.85 of the interval between successive contours in Figs. 1 and 3 (0.14 electron/Å.<sup>3</sup>), or 0.15 of the difference in density between 'peaks' and 'valleys' (0.7 electron/Å.<sup>3</sup>) (ref. 12). This calculated error is borne out by the observed fluctuations in the liquid regions between the molecules and by the difference in height between the two iron peaks (0.13 electron/Å.<sup>3</sup>).

An error of at least five times this magnitude would be needed to turn a peak into a valley and thus to simulate the backbone of a polypeptide chain in a region actually occupied by side-chains.

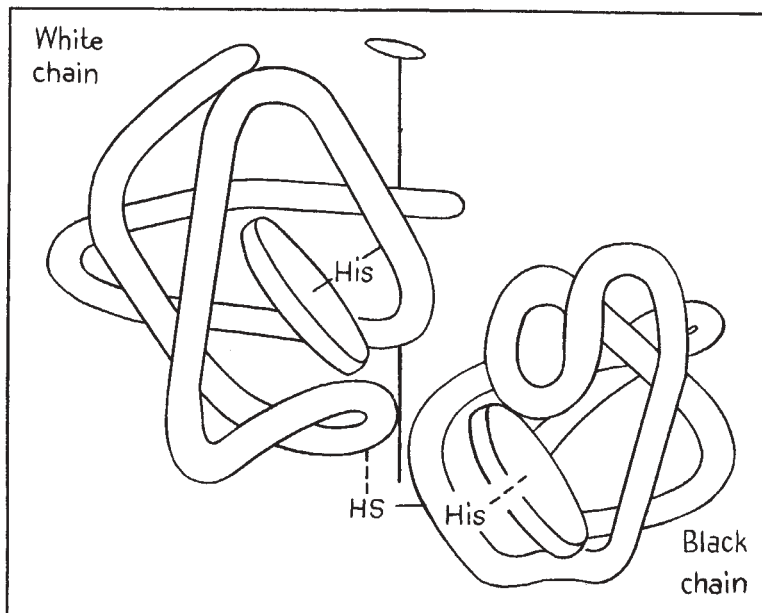
Normally, the accuracy of a structure is checked by comparison of calculated and observed intensities. In haemoglobin this is impossible so long as the light atoms are not resolved, but the positions of the iron atoms at least can be checked by their anomalous dispersion effect. Depending on the phase relationship between the structure factor of the four iron atoms by themselves and that of the entire molecule, the total intensity of any given reflexion  $I(hkl)$  may be larger or smaller than  $I(\bar{h}\bar{k}\bar{l})$ . We selected 50 reflexions for which the effect of anomalous dispersion was expected to be largest and measured the intensity of each reflexion in the four symmetrically related quadrants on the counter spectrometer. Statistically significant differences between  $I(hkl) + I(\bar{h}\bar{k}\bar{l})$  and  $I(h\bar{k}l) + I(\bar{h}kl)$  were found in 36 reflexions. In 34 of them the signs of the differences agreed with prediction, indicating that the phase angles of the reflexions as well as the positions of the iron atoms are correct.

As a further check on the positions of the iron atoms, an attempt was made to label them with heavy atoms. This was done by allowing crystals of methaemoglobin, which are isomorphous with oxyhaemoglobin, to react with *p*-iodophenyl-hydroxylamine, which attaches itself to the iron atoms. A difference Fourier projection on the *b*-plane showed four prominent peaks which lie well within the calculated distances between the iodine and the iron atoms.

Finally, to avoid any possible bias in the interpretation of the Fourier synthesis, we have tried to construct objective models containing only those features actually found in the electron density maps. Due to series termination errors and inaccuracies in many of the phase angles, these maps must certainly contain errors of detail. Large errors, on the other hand, are unlikely because, quite apart from the checks just described, the results really prove themselves. No combination of errors could have led to the appearance of four distinct chains of roughly equal length, in agreement with chemical evidence, to the similarity between the two pairs of chains which are not related by crystal symmetry, and to the resemblance between haemoglobin and myoglobin.



(a)



(b)

Fig. 8. (a) Haemoglobin model viewed normal to  $\alpha$ . The haem groups are indicated by grey disks. (b) Chain configuration in the two sub-units facing the observer. The other two chains are produced by the operation of the dyad axis

#### Discussion

The polypeptide chain-fold which Kendrew and his collaborators first discovered in sperm whale myoglobin has since been found also in seal myoglobin<sup>16</sup>.

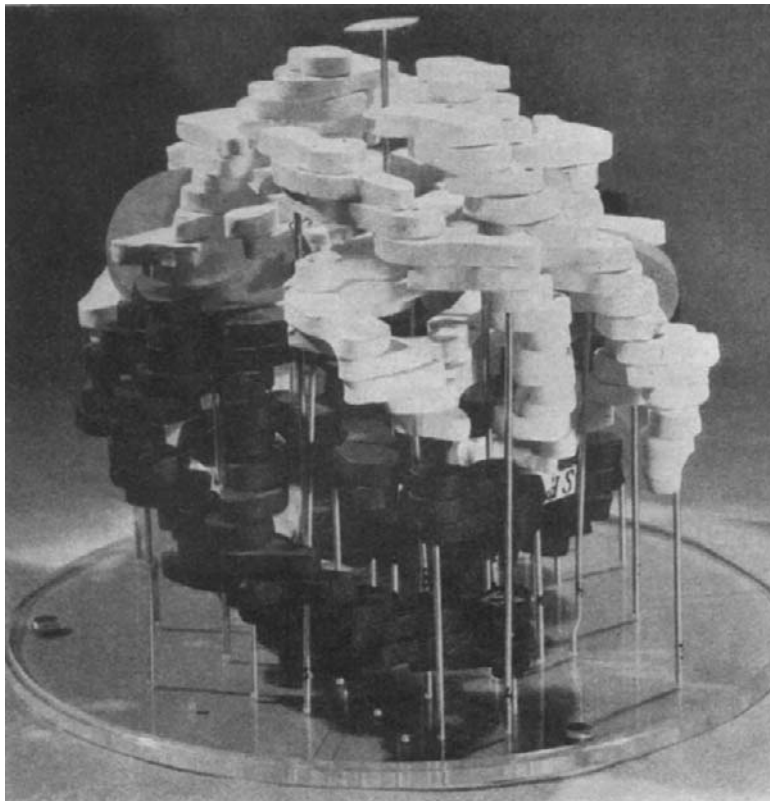


Fig. 9. Hæmoglobin model viewed normal to  $c$

Its appearance in horse hæmoglobin suggests that all hæmoglobins and myoglobins of vertebrates follow the same pattern. How does this arise? It is scarcely conceivable that a three-dimensional template forces the chain to take up this fold. More probably the chain, once it is synthesized and provided with a hæm group around which it can coil, takes up this configuration spontaneously, as the only one which satisfies the stereochemical requirements of its amino-acid sequence. This suggests the occurrence of similar sequences throughout this group of proteins, despite their marked differences in amino-acid content. This seems all the more likely, since their structural similarity suggests that they have developed from a common genetic precursor.

Little can be said as yet about the relation between structure and function. The hæm groups are much too far apart for the combination with oxygen of any one of them to affect the oxygen affinity of its neighbours directly. Whatever interaction between the hæm groups exists must be of a subtle and indirect kind that we cannot yet guess. A few observations of possible significance might be mentioned. Kendrew found that the combination of reduced myoglobin with oxygen involves no

structural changes detectable by X-ray analysis<sup>17</sup>, both forms being isomorphous with the metmyoglobin normally studied. In the hæmoglobin of horse and of man, on the other hand, the oxygenated and reduced forms are crystallographically different<sup>18-20</sup>. The structure of reduced hæmoglobin is still unknown, but it would not be surprising if loss of oxygen caused the four sub-units to rearrange themselves relative to each other, rather than to change their individual structure to a marked degree.

Riggs has shown that blocking the sulphhydryl groups reduces hæm-hæm interaction<sup>21</sup>. As Fig. 8 shows, these groups occupy key positions close to the hæm-linked histidines and to points of contact between two different sub-units. They may well play an important part in the transition between the oxygenated and reduced forms. Incidentally, the cysteine residue should provide a convenient marker for the peptide containing the hæm-linked histidine, and so help to determine the sequence in this important part of the chain.

A full account of this work will be published elsewhere.

We thank the Director and staff of the University of Cambridge Mathematics Laboratory for making their electronic computer *Edsac II* available to us for the many calculations involved in this work. We are

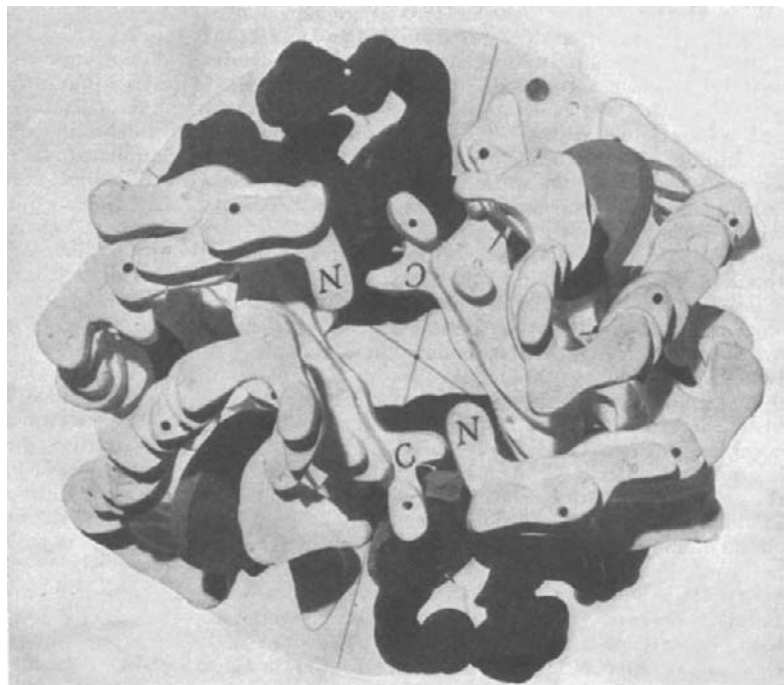


Fig. 10. A view down the  $b$ -axis. Note the proximity of the  $C$  and  $N$  terminal ends which could serve to form links between the two white chains

also deeply grateful to the Rockefeller Foundation for its long-continued financial support and to Sir Lawrence Bragg for his unfailing enthusiasm and encouragement. Finally, we wish to thank Dr. B. R. Baker, of the Stanford Research Institute, for a gift of organic mercurials, Dr. J. Chatt of Imperial Chemical Industries, Ltd., and Mr. A. R. Powell, of Johnson Matthey, Ltd., for gifts of heavy-atom compounds, and Mrs. Margaret Allen, Miss Ann Jury and Miss Brenda Davies for assistance.

<sup>1</sup> Rhinesmith, S. H., Schroeder, W. A., and Pauling, L., *J. Amer. Chem. Soc.*, **79**, 4682 (1957).

<sup>2</sup> Braunitzer, G., *Hoppe-Seyl. Z.*, **312**, 72 (1958).

<sup>3</sup> Wilson, S., and Smith, D. B., *Can. J. Biochem. Physiol.*, **37**, 405 (1959).

<sup>4</sup> Bokhoven, C., Schoone, J. C., and Bijvoet, J. B., *Acta Cryst.*, **4**, 275 (1951).

<sup>5</sup> Green, D. W., Ingram, V. M., and Perutz, M. F., *Proc. Roy. Soc., A*, **225**, 287 (1954).

<sup>6</sup> Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. W., and Phillips, D. C., *Nature*, **181**, 662 (1958).

<sup>7</sup> Blow, D. M., *Proc. Roy. Soc., A*, **247**, 302 (1958).

<sup>8</sup> Perutz, M. F., *Proc. Roy. Soc., A*, **225**, 264 (1954).

<sup>9</sup> Cullis, A. F., Dintzis, H. M., and Perutz, M. F., Conference on Haemoglobin, National Academy of Sciences, NAS-NRC Publication 557, p. 50 (Washington, 1958).

<sup>10</sup> Rossmann, M. G., *Acta Cryst.* (in the press).

<sup>11</sup> Bodo, G., Dintzis, H. M., Kendrew, J. C., and Wyckoff, H. W., *Proc. Roy. Soc., A*, **253**, 70 (1959).

<sup>12</sup> Blow, D. M., and Crick, F. H. C., *Acta Cryst.*, **12**, 794 (1959).

<sup>13</sup> Bragg, W. L., and Perutz, M. F., *Proc. Roy. Soc., A*, **225**, 315 (1954).

<sup>14</sup> Ingram, D. J. E., Gibson, J. F., and Perutz, M. F., *Nature*, **178**, 905 (1956).

<sup>15</sup> Kendrew, J. C., Dickerson, R. E., Strandberg, B., Hart, R. G., Davies, D. R., Phillips, D. C., and Shore, V. C. (see following article).

<sup>16</sup> Scouloudi, H., *Nature*, **183**, 374 (1959).

<sup>17</sup> Kendrew, J. C. (private communication).

<sup>18</sup> Haurowitz, F., *Hoppe-Seyl. Z.*, **254**, 266 (1938).

<sup>19</sup> Jope, H. M., and O'Brien, J. R. P., "Haemoglobin", 269 (Butterworths, London, 1949).

<sup>20</sup> Perutz, M. F., Trotter, I. F., Howells, E. R., and Green, D. W., *Acta Cryst.*, **8**, 241 (1955).

<sup>21</sup> Riggs, A. F., *J. Gen. Physiol.*, **36**, 1 (1952).

## STRUCTURE OF MYOGLOBIN

### A THREE-DIMENSIONAL FOURIER SYNTHESIS AT 2 Å. RESOLUTION

By Drs. J. C. KENDREW, R. E. DICKERSON, B. E. STRANDBERG, R. G. HART  
and D. R. DAVIES\*

Medical Research Council Unit for Molecular Biology, Cavendish Laboratory, Cambridge

AND

D. C. PHILLIPS and V. C. SHORE

Davy Faraday Laboratory, The Royal Institution, London

MYOGLOBIN is a conjugated protein consisting of a single polypeptide chain of about 153 amino-acid residues associated with an iron-porphyrin complex, the haem group; its molecular weight is about 18,000, and the molecule contains some 1,200 atoms (excluding hydrogen). Two years ago a preliminary report of the first stage of an investigation of the three-dimensional structure of sperm-whale myoglobin was published in *Nature*<sup>1</sup> (a detailed account of this work has appeared recently<sup>2</sup>). Several isomorphous crystalline derivatives of myoglobin containing heavy atoms (mercury or gold) at single sites on the molecule were prepared, and by comparing the X-ray diffraction patterns of these crystals with those of the unsubstituted protein, it was possible to deduce the phases of all the reflexions in the X-ray pattern having spacings greater than 6 Å. These phases, together with the observed amplitudes, were used to compute a three-dimensional Fourier synthesis of the electron density in the unit cell (which contains two molecules) at a resolution of 6 Å. In this synthesis the polypeptide chain was visible as a rod of high electron density, folded in a complex pattern (Fig. 5a); in addition, the single haem group with its iron atom, which is much more dense than any other atom in the molecule, could be identified as a disk of high electron density. The shape of the molecule could be determined with some confidence, as could most of the course of the single polypeptide chain, though there were several ambiguities where it turned through a large angle, so that the ends of the chain could not be located with certainty. Thus, the general nature of the tertiary

structure of the molecule was revealed, but not the secondary structure of the polypeptide chain, though the results were consistent with a helical configuration.

More recently, Scouloudi<sup>3</sup> used similar methods to obtain a two-dimensional Fourier projection of the unit cell of seal myoglobin, in which the molecular arrangement is entirely different from that in sperm-whale crystals; she was able to show that the myoglobins of these two species have essentially the same tertiary structure in spite of their different amino-acid compositions. Her work, by implication, confirmed the correctness of both analyses, as well as the deductions made about the shape of the molecule. In the accompanying article<sup>4</sup> Perutz *et al.* now describe a three-dimensional analysis of the related protein haemoglobin, at a slightly greater resolution, and show that each of the four sub-units of which this molecule is composed bears a close structural resemblance to myoglobin. It is apparent, therefore, that sperm-whale myoglobin possesses a structure the significance of which extends beyond a particular species and even beyond a particular protein.

We now present the results of a second stage in the analysis of sperm-whale myoglobin; in this the resolution has been increased to 2 Å., that is to say, not far short of atomic resolution. The resulting Fourier synthesis is very complicated, and a detailed study of it will take many months; in the meantime, our preliminary findings may be of interest.

#### Methods of X-ray Analysis

In this stage we have simply extended the methods which proved successful in the first stage of the analysis, comparing the diffraction pattern of unsub-

\* Visiting scientist from the National Institute of Mental Health, Bethesda, Md.