The crystal structure of myoglobin
V. A low-resolution three-dimensional Fourier synthesis of sperm-whale myoglobin crystals

By G. Bodo, H. M. Dintzis, J. C. Kendrew and H. W. Wyckoff
Medical Research Council Unit for Molecular Biology, Cavendish Laboratory, University of Cambridge

(*Communicated by Sir Lawrence Bragg, F.R.S.—Received 22 April* 1959)

[Plate 8]

The study of type *A* crystals of sperm-whale has now been extended to three dimensions by using the method of isomorphous replacement to determine the phases of all the general X-ray reflexions having $d > 6$ Å, and a three-dimensional Fourier synthesis of the electron density in the unit cell has been computed. Data were obtained from the same derivatives which had been used in the previous two-dimensional study (Bluhm, Bodo, Dintzis & Kendrew 1958), in the course of which the x and z co-ordinates of the heavy atoms had been determined. Several methods were used to determine the y co-ordinates from the three-dimensional data; with a knowledge of all three co-ordinates of each heavy atom it was possible to establish the phases of nearly all the reflexions by a graphical method. The three-dimensional Fourier synthesis was evaluated on a high-speed computer from these phases and from the observed amplitudes of the reflexions.

A resolution of 6 Å was chosen because it should clearly reveal polypeptide chains having a compact configuration such as a helix. The electron-density map was in fact found to contain a large number of dense rod-like features which are considered to be polypeptide chains, probably helically coiled. In addition, a very dense flattened disk is believed to be the haem group with its central iron atom. Finally it was possible to identify the boundaries of the protein molecules by locating the intermolecular regions containing salt solution.

An isolated myoglobin molecule has dimensions about $45 \times 35 \times 25$ Å and within it the polypeptide chain is folded in a complex and irregular manner. For the most part the course of the chain can be followed, but there are some doubtful stretches, presumably where the helical configuration breaks down; a crude measurement of the total visible length of chain suggests that about 70 % of it may be in a helical or some similarly compact configuration. The haem group is near the surface of the molecule.

1. Introduction

In the previous paper of this series (Bluhm, Bodo, Dintzis & Kendrew 1958; hereafter referred to as IV) it was shown how the method of isomorphous replacement could be used to determine the signs of the *h*0*l* reflexions from monoclinic type *A* crystals of sperm-whale myoglobin. A Fourier projection was computed from all such reflexions having spacings greater than 4 Å, but this was found to be uninterpretable owing to the large degree of overlapping, the axis of projection being 31 Å, or some twenty atomic diameters. It was concluded that the analysis should be extended to three dimensions, which would involve the determination of the general phases of the *hkl* reflexions. For this purpose the method of isomorphous replacement would again be used, the ambiguities inherent in the determination of the phase of general reflexions being resolved by using several different heavy-atom derivatives. As a preliminary to this three-dimensional analysis it is necessary to determine as accurately as possible the x and z co-ordinates of the heavy atoms, already known

5. The collection of three-dimensional data

(*a*) *Choice of resolution*

The labour involved in collecting the data for a three-dimensional Fourier synthesis increases rapidly with the resolution desired, because the number of reflexions to which phases must be assigned is proportional to the cube of the radius of the sphere in reciprocal space over which the summation is to be carried out. Furthermore there are real (though slight) departures from strict isomorphism between the various derivatives studied, generally of the order 0·1–0·3 % of the cell axis, so that the accuracy of phase determination must decrease as one goes farther out in reciprocal space. For these reasons we decided to test the method of isomorphous replacement on a small scale, by working in the first instance at the lowest resolution which might be expected to give useful information about the structure of the myoglobin molecule.

We adopted the working hypothesis that a considerable proportion of the polypeptide chain has a configuration similar to the α-helix of Pauling, Corey & Branson (1951). There is no proof that this is the case, but there is some direct evidence of it in haemoglobin (Perutz 1951), as well as indirect evidence in myoglobin from studies of optical rotation (P. M. Doty, unpublished) and of rates of deuterium exchange (E. E. Benson & K. Linderstrøm-Lang, unpublished). At low resolution the α-helix would appear as a solid rod with axial electron density about 1·0 electrons/Å^3, embedded in a matrix of side chains of mean electron density about 0·3 electron/Å^3 (the mean overall electron density of the myoglobin molecule is about 0·4 electron/Å^3, and in type *A* crystals the electron density of the liquid regions has about the same value). Neighbouring α-helices would pack together with axial separations of 9 to 10 Å. We reached the conclusion that helices, if indeed they exist in myoglobin, would be clearly resolved if the Fourier synthesis included all terms having $d > 6$ Å; they should appear in such a synthesis as solid rods, since the region of reciprocal space being scanned includes only the first maximum of the Fourier transform of an α-helix. In fact there are about 400 reflexions having $d > 6$ Å, of which about 100 are *h*0*l* reflexions with real phases which had already been determined in the two-dimensional work.

The choice of a 6 Å limit has a further advantage which can be appreciated from a study of the radial distribution of intensities plotted against reciprocal spacing (figure 11). As in many other protein crystals this function has a minimum at about 6 Å^{-1}, so pronounced that if the data are cut off sharply at this point a Fourier synthesis can be calculated without fear of serious series termination errors even though no artificial temperature factor is applied to the data.

(*b*) *X-ray methods*

In all we measured the $|F|$ values of the *hkl* reflexions from six different types of myoglobin crystal, namely unsubstituted met-myoglobin, the PCMBS, $HgAm_2$ and Au derivatives, and the double derivatives PCMBS/$HgAm_2$ and PCMBS/Au. Nearly all these data were obtained with the Buerger precession camera and Cu$K\alpha$ radiation from the rotating-anode X-ray tube developed by Mr D. A. G. Broad. The method

a mean value in different unit cells. Later work has shown that in fact a different value of B should be used for each of the derivatives, but in the central region of reciprocal space the value can be varied widely without appreciably affecting the results.

(*b*) *Determination of phases*

In the method of isomorphous replacement the phases of the reflexions are found by solving a vector triangle whose sides are F, F_H and f_H (F = structure factor of protein; F_H = structure factor of protein + heavy atom; f_H = structure factor of heavy atom), given only the values of $|F|$ and $|F_H|$ but with a knowledge of both the amplitude and the phase of f_H. It is well known that in general two solutions satisfy a given set of data; in the method of multiple isomorphous replacement this ambiguity is resolved by repeating the process with several heavy atoms in different places in the unit cell. One of each pair of solutions should give an identical value of the phase of F for each derivative, and this common value is the correct one (Bokhoven, Schoone & Bijvoet 1951).

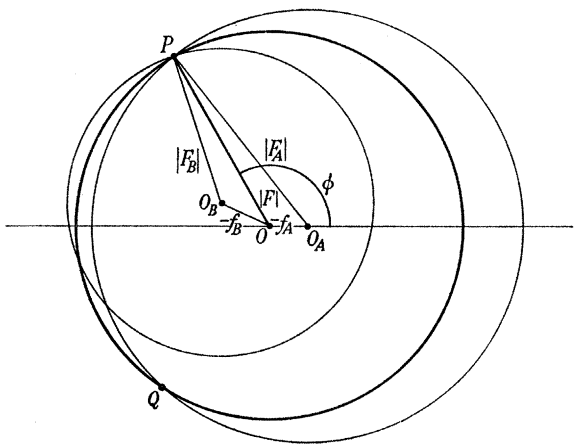


FIGURE 16. Graphical method of phase determination using two isomorphous derivatives. For explanation see text.

In the present instance there are five derivatives, PCMBS, $HgAm_2$, Au, PCMBS/$HgAm_2$ and PCMBS/Au. The most convenient way of discovering the correct phase angle is to adopt a method of graphical representation in which all the data for a given reflexion can be combined in one diagram. We used a scheme which has been described by Harker (1956), and which is illustrated in figure 16. A circle of radius $|F|$ is drawn with centre O. From the end of a vector OO_A, such that $OO_A = -f_A$, a circle of radius $|F_A|$ is drawn (if A = PCMBS, this vector always has phase angle 0 or π because the centre of symmetry of the PCMBS atoms has been chosen as origin), and its intersections with the first circle (P and Q) give two alternative values of ϕ. Next a circle of radius $|F_B|$ is drawn from the end of a vector OO_B equal to $-f_B$. If there are no experimental errors one of the intersections of this circle with the protein circle should coincide with one of the intersections of circle A with the protein circle. This intersection (P) establishes the direction OP of the protein vector

and gives the phase ϕ of the reflexion. Further circles can be drawn for the other derivatives, and ideally all of them should pass through the point B. In practice, owing to experimental errors and also, perhaps, to slight departures from true isomorphism, to the presence of some heavy atoms at subsidiary sites in the cell, and to incorrect assignment of heavy-atom co-ordinates, the circles do not all meet at a point—some of them, indeed, may not meet at all—but by this graphical method it is very easy to choose a 'best value' for the phase angle and to make a rough estimate of the reliability of the determination in any particular case.

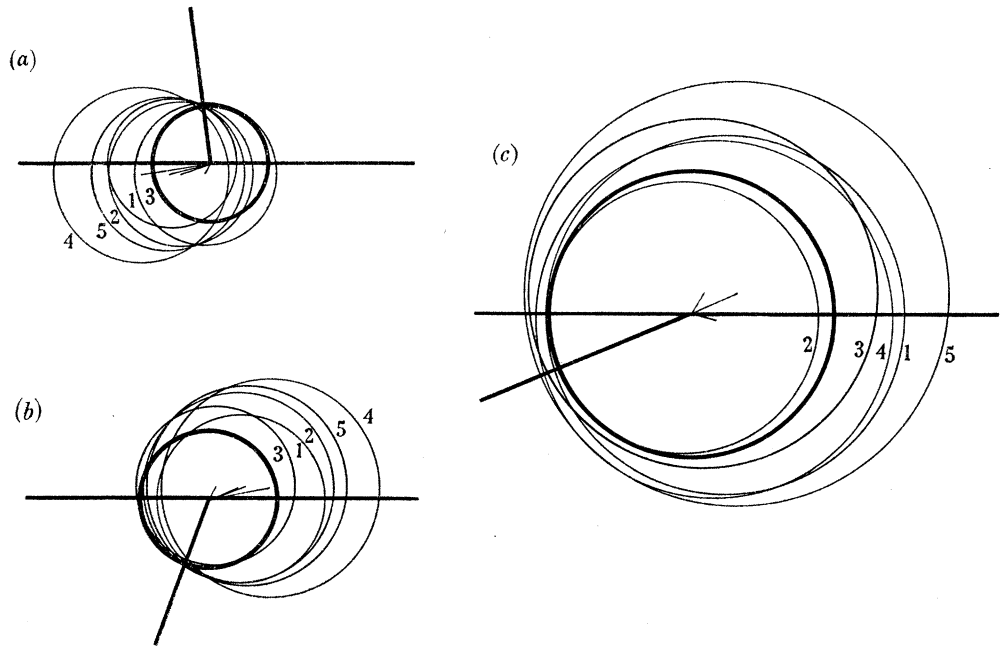


FIGURE 17. Examples of phase determination. The heavy circle represents the amplitude of the reflexion from unsubstituted protein, and the light circles those from the derivatives. 1, PCMBS; 2, $HgAm_2$; 3, Au; 4, PCMBS/$HgAm_2$; 5, PCMBS/Au. The short lines from the centres are the heavy-atom vectors; the heavy line indicates the phase angle eventually selected. (*a*) (411), (*b*) (91$\bar{1}$), (*c*) (212).

Some examples of actual phase determinations are given in figure 17. As an indication of the accuracy attained the following result may be quoted. Two independent sets of data were obtained for the *hk*0 reflexions, one from the main set of *hkl* data, the other from the precession pictures taken along z with $\mu = 17°$ which were used for determining y co-ordinates. Separate phase diagrams were drawn for each of these sets, and independent determinations of ϕ were made from each of the two diagrams corresponding to a given reflexion. In all there were 39 reflexions; in 2 the results were so ambiguous that a phase angle could not be assigned; in the other 37 the mean difference between the two values of the phase angle was only 11°. Again, the phases of five 0*k*0 reflexions were determined separately by one of the present authors (J. C. K.) from the z projection data, and by Sir Lawrence Bragg from x projection data; the mean phase difference between the two sets of results was 10°.

In all we examined 398 reflexions, of which 99 were *h*0*l*'s with phases 0 or π and 299 were *hkl*'s with general phases. Of these we omitted from the final Fourier synthesis two *h*0*l* reflexions whose signs were doubtful and six *hkl* reflexions whose phase could not be determined; and in addition we omitted either the A or B component of a further seven *hkl* reflexions for which the determination was partly ambiguous. A list of phases and amplitudes is given in table 3, which has been deposited in the Royal Society's archives.

(*c*) *Anomalous dispersion effects*

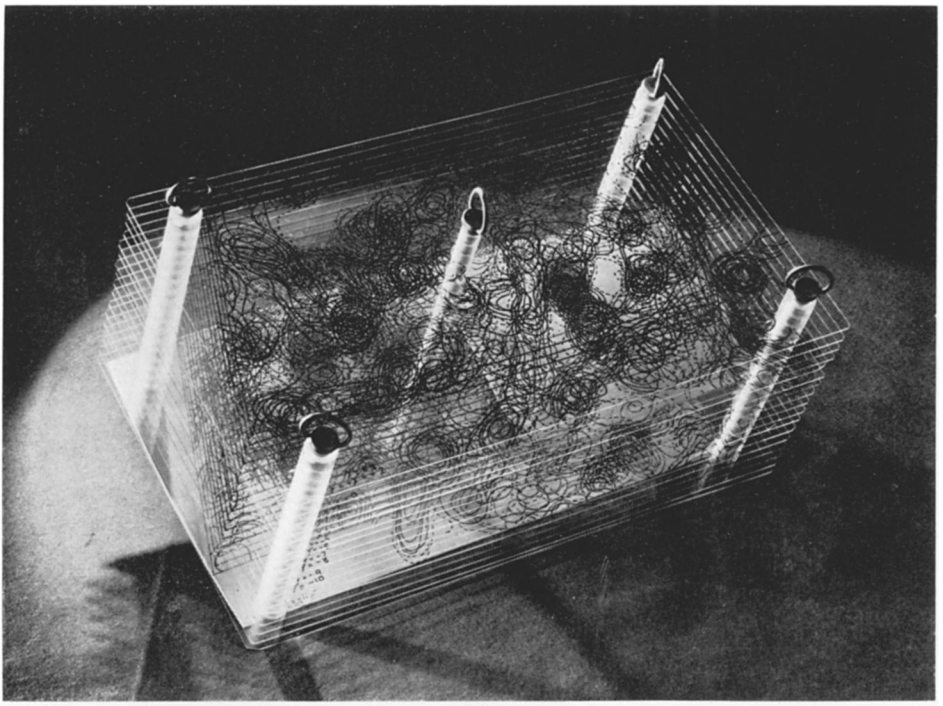
In the diffraction patterns of the heavy-atom derivatives we frequently observed pairs of 'equivalent' reflexions (*hkl*) and ($\bar{h}\bar{k}\bar{l}$) whose intensities were unequal, i.e. which did not obey Friedel's Law. Thus the intensities of the (020) and (0$\bar{2}$0) reflexions of the mercuri-iodide derivative differ by as much as 25 %. This was the largest difference observed, but similar discrepancies in excess of experimental error were noted in many reflexions from all the heavy-atom derivatives, and occasionally from unsubstituted myoglobin. We attribute these effects to anomalous dispersion by the heavy atoms, especially mercury and iodine; the non-real components of the structure factors of these atoms are as much as 10 to 15 % of the real component with Cu$K\alpha$ radiation (Dauben & Templeton 1955). In unsubstituted crystals the anomalies are probably due to the iron atom of the haem group, whose structure factor also has a large non-real component.

Similar anomalies have been observed in derivatives of haemoglobin by Blow (1958), who used the larger effects produced by chromium radiation for phase determination. In myoglobin no data have been obtained with chromium radiation; the effects observed with copper radiation are nearly always too small to give a decisive indication of phase, and we have made no use of them for this purpose. In a few cases where the effect is large we used it to confirm phases determined by isomorphous replacement, and found that its magnitude agreed with prediction.

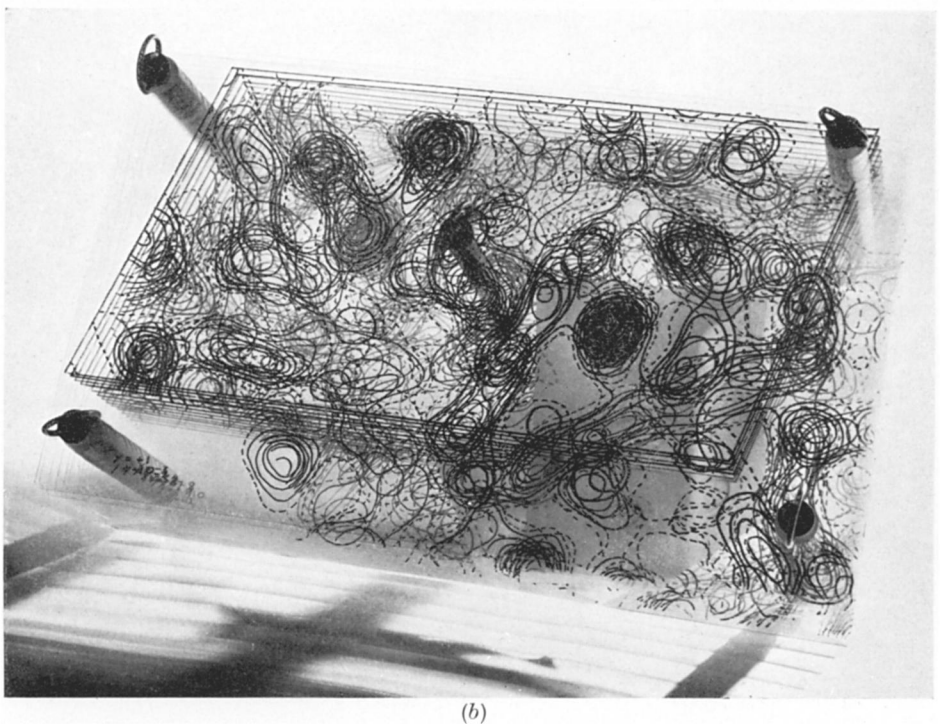
7. Computation of the three-dimensional Fourier synthesis

The three-dimensional Fourier synthesis was calculated on the EDSAC mark I computer of the Mathematical Laboratory, University of Cambridge, at intervals of $x/32$, $y/16$ and $z/16$ (i.e. about 2 Å in each direction). Owing to the symmetry of the space group it is only necessary to calculate the synthesis for half the unit cell, so the computation was arranged in the form of eight sections, perpendicular to y and ranging from $y = +\frac{1}{4}b$ to $y = -\frac{1}{4}b$. The program used was a modification of one prepared by Dr D. W. Green for computing two-dimensional Fourier syntheses; the real and imaginary components of all the structure factors were fed into the machine anew for each value of y, a one-dimensional synthesis along y being carried out during input; summations along x and z were then made by the two-dimensional program. Each section was computed in about 9 min, the total time for the whole summation being 72 min.

As a check, part of the calculation was repeated by Dr J. S. Rollett on the DEUCE computer at the National Physical Laboratory, using a program devised



(*a*)



(*b*)

FIGURE 19. Photographs of the three-dimensional Fourier synthesis of myoglobin, constructed from the sections illustrated in figure 18.

(*Facing p.* 93)

by himself. The results were identical apart from trivial differences nowhere exceeding, and generally much less than, one-fifth of the contour intervals used in figure 18; these discrepancies were probably due to minor errors in preparing the data for one or other computation, and are certainly not significant.

The eight sections of the Fourier synthesis are shown in figure 18. Contours have been drawn at convenient arbitrary intervals, which actually correspond to $9{\cdot}08 \times 10^{-2}$ electron/Å^3. Since no $F(000)$ term was included in the synthesis, the zero contour corresponds to the mean electron density of the unit cell, which is 0·395 electron/Å^3. To convert the contours shown to absolute electron densities it would be necessary to add 4·32 contour intervals at every point of the synthesis; this would have the effect of making the electron density almost everywhere positive, with only a few minor negative excursions. This result is a satisfactory check of the absolute scale of intensities used, though precise values of the electron densities could not be expected in a synthesis calculated from limited data with a sharp cut-off.

To study the synthesis a three-dimensional model was made by drawing the sections on thin Perspex sheets and piling them up at appropriate intervals; figure 19, plate 8, shows two photographs of this model. In reading the discussion of the synthesis in the next section it must be borne in mind that it is difficult to appreciate its features from the individual sections of figure 18; unfortunately it is impossible to give an adequate representation of the three-dimensional synthesis within the pages of a journal.

The diagrams used to determine the phase of each protein reflexion also indicate directly the phase angles appropriate to each of the heavy-atom derivatives; and from these, together with the observed structure amplitudes, Fourier syntheses were calculated (by Dr W. Hoppe) for some of the derivatives (PCMBS, $HgAm_2$, Au and PCMBS/$HgAm_2$). In general these syntheses resemble that of the unsubstituted protein very closely; there is no evidence that the protein molecules are appreciably displaced, and individual density values differ by less than one contour interval (on average by half a contour interval). The heavy atoms appear as high peaks in the expected positions; their co-ordinates are listed in table 1, and correspond within experimental error to those assumed in the structure factor calculations, a result which suggests that the latter were substantially correct for each of the heavy atoms, since the phases are based on an average of all the derivatives (a synthesis carried out with phases deduced from one derivative only would of course reproduce the co-ordinates assumed, however much these were in error). Examples of sections through the heavy-atom peaks are shown in figure 20, together with the corresponding difference-sections obtained by subtracting the Fourier synthesis of the protein from that of the heavy-atom derivative. The additional scattering matter in the heavy-atom peaks may be obtained by integrating the change in electron density over the whole peak; for PCMBS, $HgAm_2$ and Au the results are 81, 74 and 52 electrons respectively. These are all of the right order of magnitude, but higher than the assumed figures ($= \frac{1}{2}f(000)$), which were 65, 55 and 39 electrons respectively. However, the heavy-atom peaks are surrounded by diffraction fringes owing to the sharp termination of the series, and exact numerical

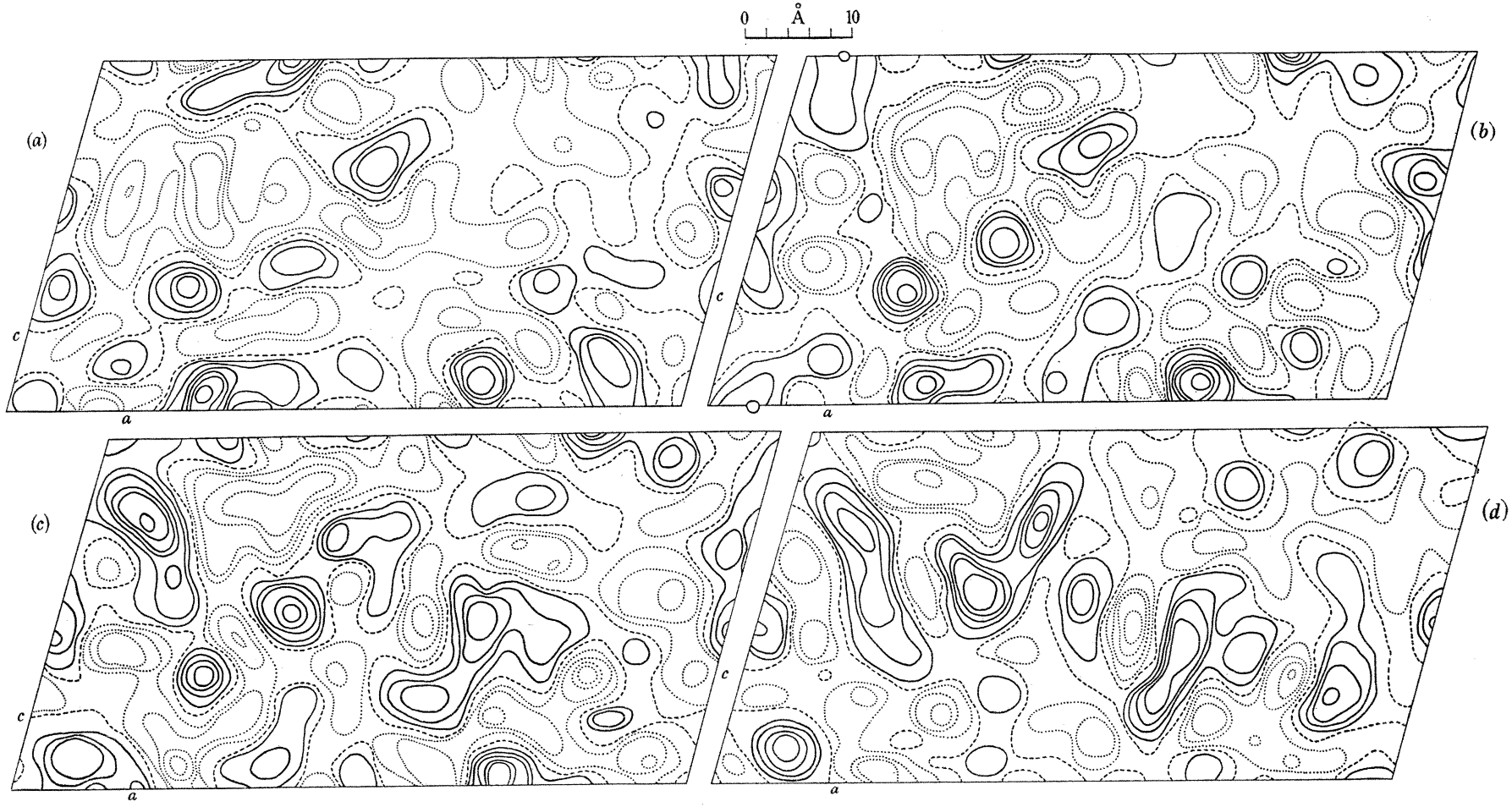


FIGURE 18. *a—d.*

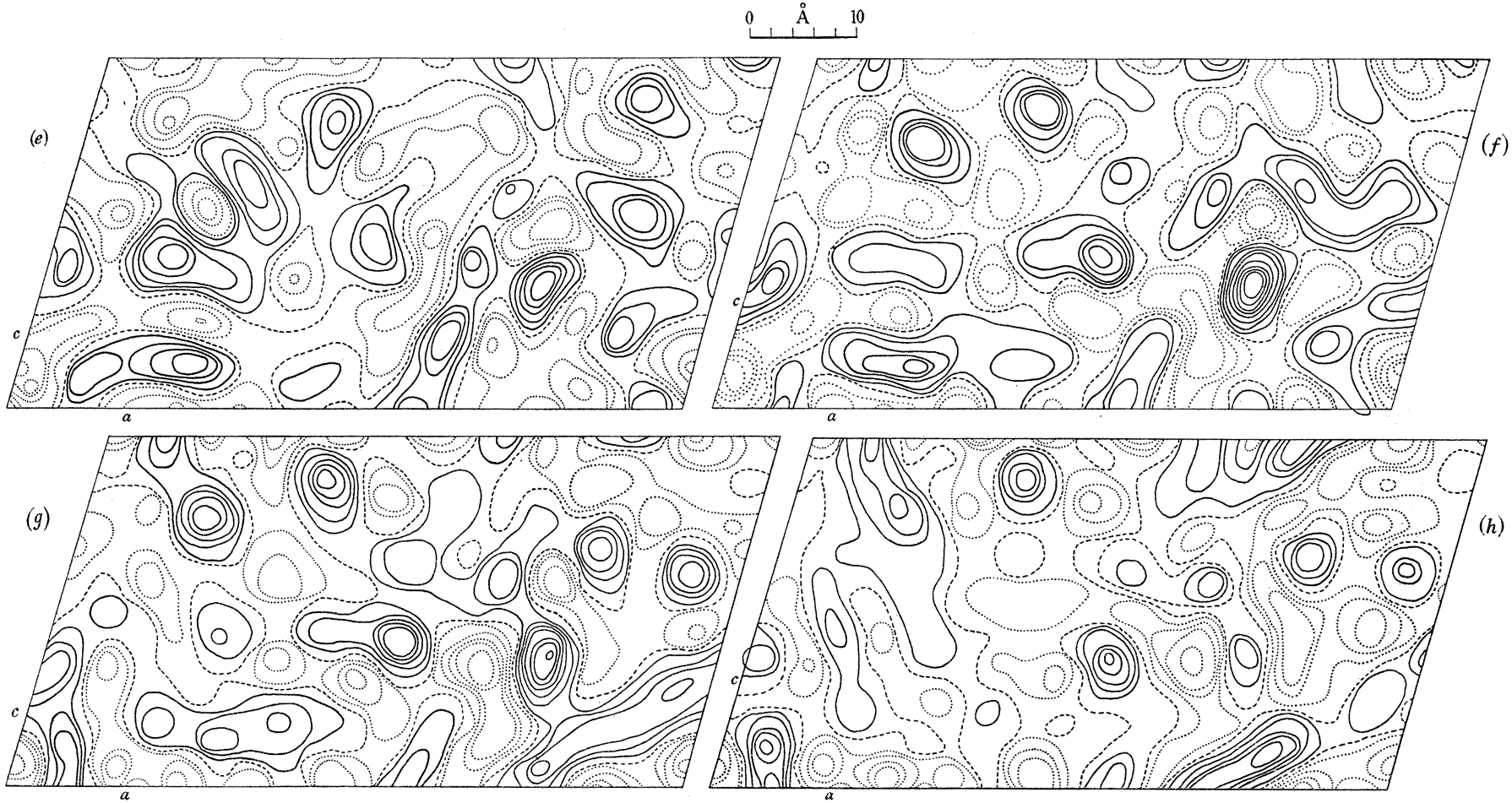


FIGURE 18. Sections perpendicular to y of the three-dimensional Fourier synthesis of myoglobin. (*a*) $y = +\frac{4}{16}b$ (*b*) $y = +\frac{3}{16}b$ (*c*) $y = +\frac{2}{16}b$ (*d*) $y = +\frac{1}{16}b$ (*e*) $y = 0$ (*f*) $y = -\frac{1}{16}b$ (*g*) $y = -\frac{2}{16}b$ (*h*) $y = -\frac{3}{16}b$. After rotation through 180° each section is identical with the section at a level $\frac{1}{2}b$ lower (or higher) in the unit cell. The zero contour is dashed; negative contours are dotted.

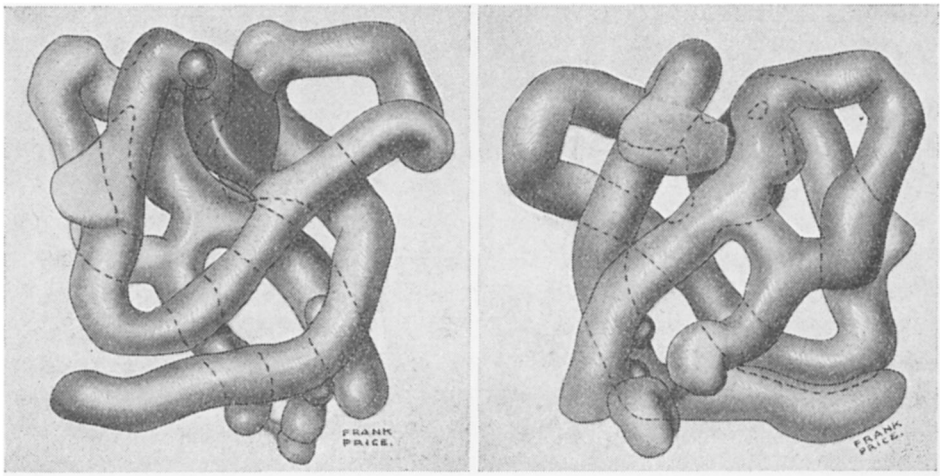


FIGURE 22. Drawings of a model of the myoglobin molecule. The haem group is shown in (*a*) as a foreshortened dark grey disk. The small spheres are the heavy atoms used in phase determination; in (*a*) the upper one is mercury of PCMBS. The two lower spheres visible in both drawings, but partly concealed by polypeptide chains, are gold of $AuCl_4'$ (upper) and mercury of $HgAm_2$ (lower).

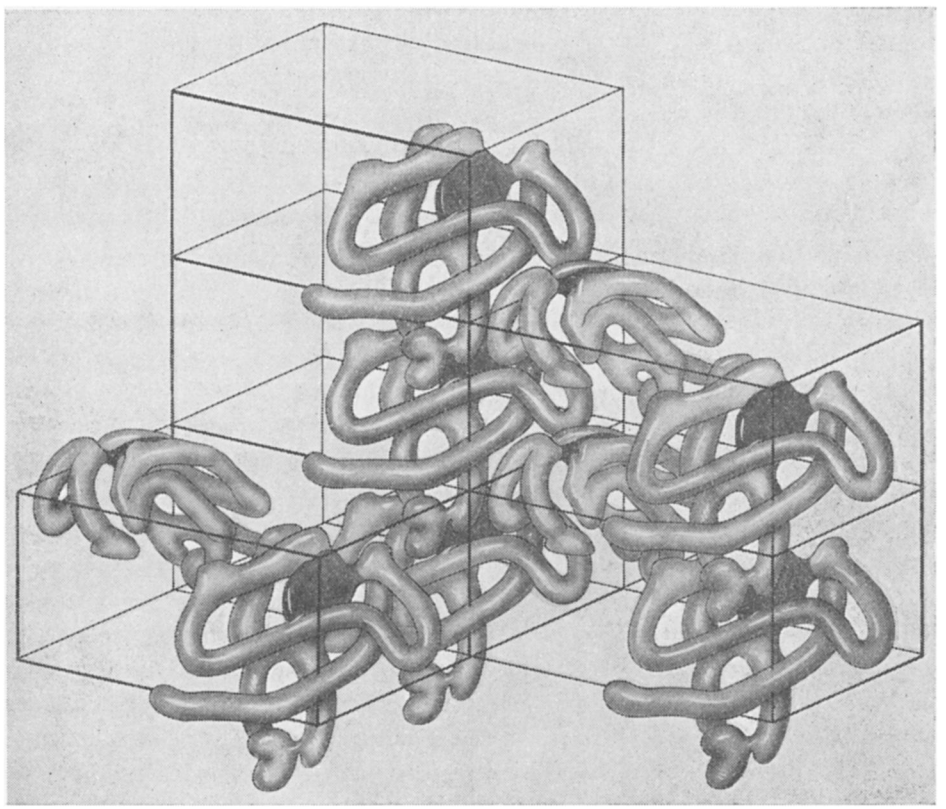


FIGURE 23. Drawing showing the arrangement of myoglobin molecules in the crystal lattice. The edges of the unit cell are indicated by lines, the y axis being vertical.