Acta Cryst. (1961). **14**, 1188

The Crystal Structure of Myoglobin: Phase Determination to a Resolution of 2 Å by the Method of Isomorphous Replacement

By R. E. Dickerson,* J. C. Kendrew and B. E. Strandberg†

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

(*Received* 2 *September* 1960 *and in revised form* 31 *January* 1961)

The method of multiple isomorphous replacement has been used to solve the structure of the protein myoglobin to a resolution of 2 Å. EDSAC II has been programmed to calculate the phase angles, and the statistical treatment of errors proposed by Blow & Crick has been used throughout, leading not to the most probable electron-density map but to the one with the least mean square error over the entire unit cell. Problems inherent in adapting the process for a digital computer are discussed, and examples are given of probability curves for typical phase determinations. A method of refining heavy-atom parameters in the course of phase determination is presented. Comparisons of the most-probable and least-error electron-density maps are presented in the region of the haem group, and it is shown that the latter map gives slightly better results. In an appendix a means of trial-and-error least-squares refinement of heavy-atom parameters using data from a centrosymmetric projection is discussed.

1. Introduction

Most methods of crystal-structure analysis involve the use of a certain amount of initial information about the structure, or the accumulation of knowledge in the course of iterative trial-and-error refinements. With proteins, which commonly contain thousands of atoms per molecule, such an approach is impracticable and more direct means of attack must be used. The method which has thus far proved successful is that of multiple isomorphous replacement. It is with the application of this method to sperm-whale myoglobin and with some of the computational problems encountered that this paper is concerned.

Green, Ingram & Perutz (1954) showed that the addition of only two mercury or silver atoms to a haemoglobin molecule of molecular weight 68,000 produced quite discernable changes in the diffracted X-ray intensities from the protein crystals. This demonstrated that isomorphous-replacement methods were practicable in proteins and hence that direct phase determination was possible. The first use of this method on proteins was the calculation of a centrosymmetric projection of horse haemoglobin by Bragg & Perutz (1954). Blow (1958) calculated the first projection involving complex structure factors, also of horse haemoglobin. In both cases the overlap of atoms in projection was so extensive that virtually no information could be obtained about the molecule.

The first three-dimensional application of multiple isomorphous replacement to a non-centrosymmetric structure was carried out by Kendrew and co-workers (Kendrew *et al.*, 1958; Bodo, Dintzis, Kendrew & Wyckoff, 1959). In this work the general features of the myoglobin molecule were determined using the four hundred reflections for which $2 \sin \theta$ was less than $\lambda/6$, or to '6 Å resolution'. The recent extension to 2 Å resolution (Kendrew *et al.*, 1960) involved a nearly twenty-seven-fold increase in data and necessitated abandonment of the previous graphical solution of phase angles in favour of a digital solution carried out on EDSAC II.

2. Geometrical solution for phase angles and treatment of errors

Harker (1956) proposed a graphical method of determining phases by isomorphous replacement in non-centrosymmetric structures. He assumes that there are available a parent compound and a series of at least two derivatives, each differing from the parent by the addition or substitution of a heavy atom or group. The parent and derivatives are required to crystallize isomorphously—that is, the heavy group cannot be so large that it appreciably perturbs the crystallization of the parent. In addition it is assumed that, by means which are summarized elsewhere (Dickerson, Kendrew & Strandberg, 1960), the coordinates of the atom or atoms of the heavy groups have been found, so that both magnitudes and phases of the scattering contribution of the heavy groups can be calculated. Scattering contributions from individual atoms are represented as vectors on the complex plane, and structure factors of molecules are obtained by vector addition of the individual atomic scattering contributions.

The following definitions will be needed:

$\mathbf{F}, \mathbf{f}_j, \mathbf{F}_j$ = vector scattering amplitudes for the parent compound, the heavy atom group j, and the

---

* Present address: Department of Chemistry, University of Illinois, U.S.A.

† Present address: Department of Chemistry, University of Uppsala, Sweden.

derivative made up of parent plus heavy-atom group, respectively;
F, f_j, F_j = magnitudes of the above vectors;
$\varphi, \alpha_j, \beta_j$ = phase angles of the above vectors.

Hence in complex-vector notation:

$$\mathbf{F}=F\exp i\varphi,\ \ \mathbf{f}_j=f_j\exp i\alpha_j,\ \ \mathbf{F}_j=\mathbf{F}+\mathbf{f}_j=F_j\exp i\beta_j. \quad (1)$$

Harker assumes that F and all the F_j's (magnitudes) are known from intensity measurements, and that all the f_j's and α_j's are calculable. With perfect isomorphism and no experimental errors, $\mathbf{F}$, $\mathbf{f}_j$ and $\mathbf{F}_j$ will form a closed triangle in the complex plane. If $-\mathbf{f}_j$ is plotted from the origin and if two circles are drawn, one of radius F centred at the origin and one of radius F_j centred at the head of vector $-\mathbf{f}_j$, then the two intersections of the circles (A and B, Fig. 1) will determine the two possible orientations of $\mathbf{F}$ which satisfy the vector equation (1). A second heavy-atom derivative is needed to resolve the ambiguity and determine the phase uniquely, as shown in Fig. 1.

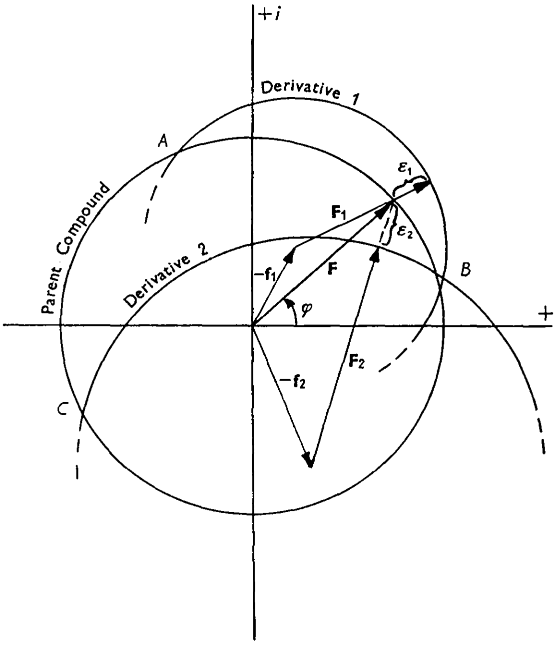


Fig. 1. Phase circle diagram of parent compound and two derivatives. Vector equation $\mathbf{F}_j=\mathbf{F}+\mathbf{f}_j$ satisfied for derivative 1 at points A and B; for derivative 2 at B and C. B is hence the correct phase angle. For general phase angle φ, lack of closure of vector triangles for derivatives 1 and 2 is ε_1 and ε_2 respectively.

These ideal conditions are never met in practice. Imperfect isomorphism and experimental errors prevent a perfect intersection of the three circles of Fig. 1, and more derivatives are needed to increase the accuracy of the phase determination. Blow & Crick have treated the question of errors (Blow, 1958; Blow & Crick, 1959) and have shown that all the errors may be treated as residing in the measurement of F_j as shown in Fig. 1, the 'lack of closure' error for a particular phase angle φ being $\varepsilon_j(\varphi)$. Assuming a Gaussian distribution of errors, the probability that a phase angle φ is correct is related to the lack of closure of the phase triangle for this angle, and is given by:

$$P_j(\varphi)=\exp -\varepsilon_j^2(\varphi)/2E_j^2. \quad (2)$$

This lack-of-closure error ε_j for heavy atom compound j is given by:

$$\varepsilon_j^2(\varphi)=(F_j-D_j)^2, \quad (3)$$

where $D_j(\varphi)$ is the third side of the phase triangle formed by the $-\mathbf{f}_j$ and $\mathbf{F}$ vectors, and is given by:

$$D_j^2(\varphi)=F^2+f_j^2+2Ff_j\cos(\varphi-\alpha_j)\,. \quad (4)$$

The root-mean-square error in the jth compound, E_j, may be found from a centrosymmetric projection where $\mathbf{F}$, $\mathbf{f}_j$, and $\mathbf{F}_j$ must be collinear:

$$E_j^2=\langle(|F-F_j|-f_j)^2\rangle=\langle(\Delta F_o-\Delta F_c)^2\rangle\,. \quad (5)$$

This implicitly assumes that the error E_j is the same for acentric and centric reflections. E_j, being derived from a comparison of a heavy-atom derivative with the parent compound, includes all experimental errors, errors from imperfect isomorphism and errors in location and characterization of the heavy group. When several heavy-atom derivatives are used simultaneously the total probability of a given phase angle φ_i is proportional to the product of the individual probabilities:

$$P_i=P_{(\varphi_i)}=\prod_j P_j(\varphi_i)=\exp -\sum_j \varepsilon_j^2(\varphi_i)/2E_j^2. \quad (6)$$

It would seem reasonable to take as the phase angle the most probable angle φ_M, the angle for which $\sum_j \varepsilon_j^2/2E_j^2$ is a minimum. However, Blow & Crick have shown that the 'best' Fourier synthesis, or that one for which the point-by-point sum of squares of the difference between the real structure and the computed Fourier synthesis is least, is that synthesis obtained by using the structure amplitude and phase of the *centre of gravity* of the probability distribution, plotted as a line density around the parent-compound phase circle. Thus in Fig. 2, P_i is plotted around the phase circle and the centre of gravity of the resulting line-probability density is at the end of vector $F\mathbf{m}$, with polar coordinates (mF, φ_B). (For clarity the diagram has been presented as scaled down by a factor of $1/F$ to unit radius, and the line density has been represented by radial distance from the phase circle.) If the probability distribution is sharp then vector $\mathbf{m}$ will fall very near to the unit circle, but if the probability is nearly uniform around the circle, $\mathbf{m}$ will be nearly of zero length. The magnitude of the vector $\mathbf{m}$ is hence a measure of the reliability of the phase determination, and functions as a weighting factor on the observed structure-factor magnitudes. (It will be shown later

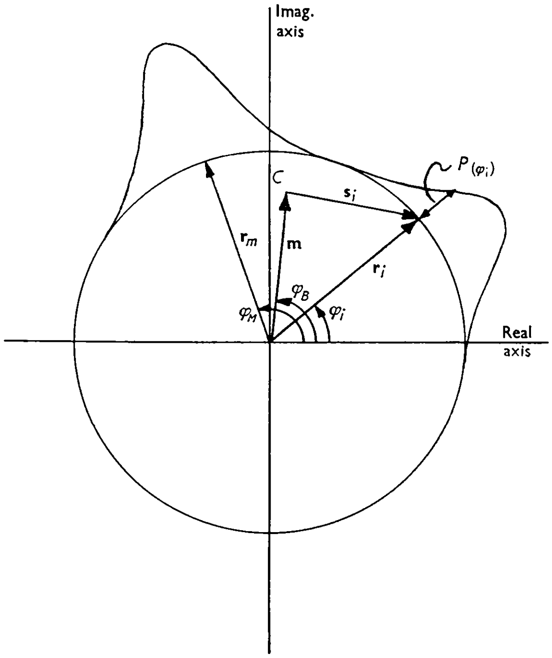


Fig. 2. Unit-radius phase circle with line probability density $P_{(\varphi)}$ represented radially outward from phase circle as base line. Vector $\mathbf{m}$ extends from origin to centre of gravity, C, of line density $P_{(\varphi)}$ taken around entire circle. Polar coordinates of C are (m, φ_B). $\mathbf{r}_i$ is general unit vector associated with phase angle φ_i, and vector structure factor for this phase angle is $F\mathbf{r}_i$. $\mathbf{r}_m$ is unit vector to point of maximum probability, with associated most probable phase φ_M. Vector $\mathbf{s}_i$ is defined as shown.

that m is the mean value of the cosine of the error in phase angle for the reflection in question.) A Fourier synthesis using mF and φ_B instead of the usual F and φ_M is 'best' in the sense that it takes into account the statistical probability that some phase angle other than the most probable angle is the true one.

3. Adaptation of the centroid method for digital computer

For the 6 Å Fourier synthesis of myoglobin, phase circles for the 300 acentric reflections were plotted graphically using five heavy-atom derivatives, and the phase angles were estimated visually (Bodo *et al.*, 1959). For the 176 reflections of the non-centrosymmetric projection of horse haemoglobin (Blow, 1958) the centroid was found by approximating the probability function by one or two Gaussian peaks and taking the centroid of this simplified system. For the 2 Å Fourier synthesis of myoglobin, phases were required for nearly 9,000 acentric reflections as well as the centric $h0l$ reflections, using four heavy-atom derivatives. The use of a digital computer permitted a more exact treatment of probabilities, and the methods employed should be of general applicability.

The centroid of a distribution has properties which make mechanization of its computation very simple. Let us assume that we have computed the total probability P_i for a number of equally spaced values of the phase angle φ_i around the circle. Then from the definition of the centroid:

$$\sum_i P_i(F\mathbf{s}_i) = 0$$

the summation being vectorial. The vector $\mathbf{s}_i$ as shown in Fig. 2 is the vector from the centroid of the probability distribution to the ith point on the phase circle, the circle being scaled down by $1/F$ to unit radius as mentioned earlier. Since $\mathbf{s}_i = \mathbf{r}_i - \mathbf{m}$, the following holds:

$$\sum_i P_i F(\mathbf{r}_i - \mathbf{m}) = 0\,,$$

$$\mathbf{m} = \sum_i P_i \mathbf{r}_i \Big/ \sum_i P_i = \sum_i P_i \exp i\varphi_i \Big/ \sum_i P_i\,. \qquad (7)$$

That is, the centroid vector $F\mathbf{m}$ is the weighted mean of all the possible scattering vectors $F\mathbf{r}_i$. The polar coordinates (m, φ_B) of $\mathbf{m}$ are given by:

$$m \cos\varphi_B = \sum_i P_i \cos\varphi_i \Big/ \sum_i P_i$$

$$m \sin\varphi_B = \sum_i P_i \sin\varphi_i \Big/ \sum_i P_i \qquad (8)$$

since $\mathbf{r}_i$ is a unit vector.

The error in phase angle at a given φ_i is defined as $\Delta\varphi_i = \varphi_B - \varphi_i$ (see Fig. 2). Then in equation (8) above, if we shift the origin line until it passes through the centroid, so that $\varphi_B = 0$ and $|\varphi_i| = |\Delta\varphi_i|$:

$$m = \sum_i P_i \cos\Delta\varphi_i \Big/ \sum_i P_i = \langle \cos\Delta\varphi_i \rangle\,. \qquad (9)$$

Hence m is the weighted mean of the cosine of the error in phase angle and is referred to as the 'figure of merit'.

The mean square error in electron density of the 'best' Fourier synthesis is simply related to the figures of merit:

$$\langle \Delta\varrho^2 \rangle = (2/V^2) \sum_{h=0}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F^2_{hkl}(1 - m^2_{hkl})\,. \qquad (10)^*$$

To prove this, note that equation (13) of Blow & Crick, giving the mean square error in electron density over the entire cell contributed by one reflection and its Friedel conjugate can be written in vector notation as:

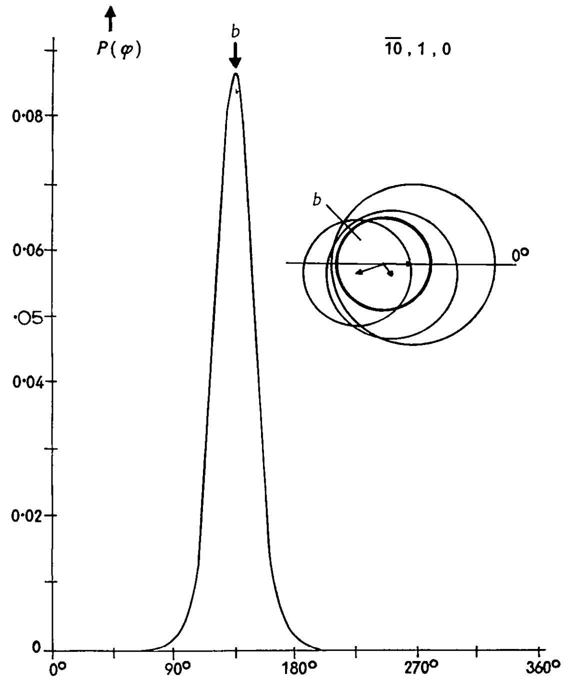
$$\langle \Delta\varrho^2_{hkl} \rangle = (2/V^2) \sum_i P_i(\mathscr{F} - F\mathbf{r}_i)^2 \Big/ \sum_i P_i\,.$$

$\mathscr{F}$ is the vector structure factor used in the Fourier synthesis whose accuracy we are examining, and $F\mathbf{r}_i$ is a choice of the true vector structure factor with a probability P_i of being correct. We have not assumed, as Blow & Crick did, that P_i is normalized.

Differentiating $\langle \Delta\varrho^2_{hkl} \rangle$ with respect to $\mathbf{r}$ and setting the differential to zero, we arrive at an equation

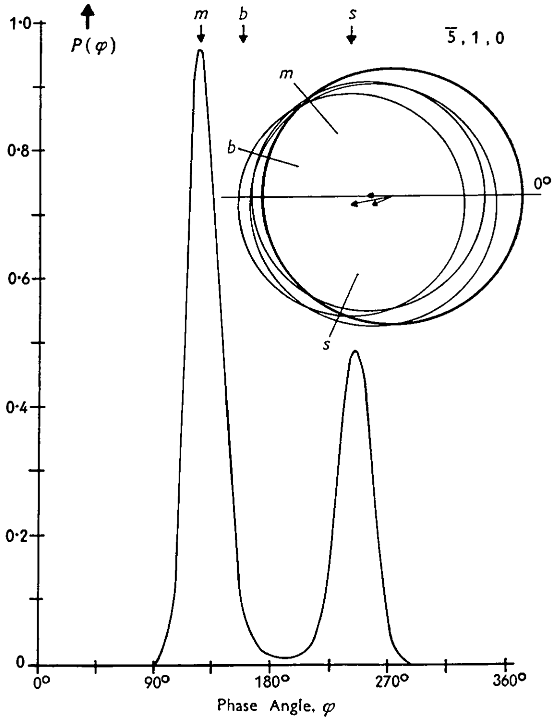
---

* The symbol $\langle \Delta\varrho^2 \rangle$ is used as a convention to denote $\langle (\Delta\varrho)^2 \rangle$, the mean square value of $\Delta\varrho$.



(a)

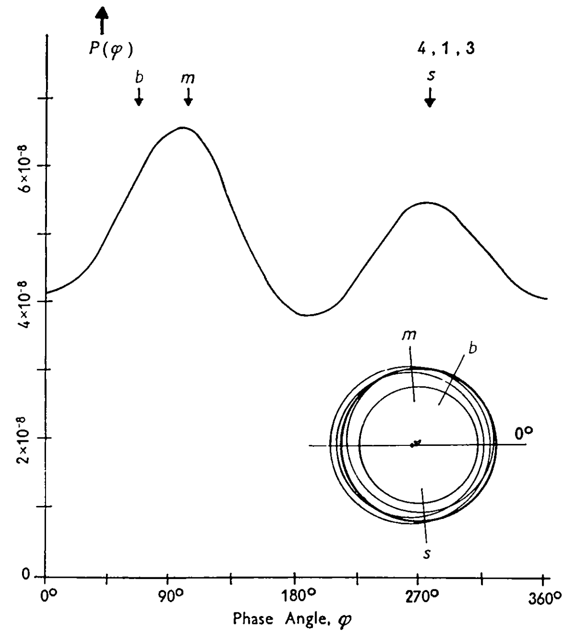
Fig. 3.



(b)

Fig. 3.

equivalent to equation (14) of Blow & Crick, giving the value of $\mathscr{F}$ which will minimize the mean square error:



(c)

Fig. 3. Phase-circle diagrams for three reflections from 6 Å myoglobin data, and probability function plots from equation (6), text. Heavy-atom vectors shown in phase circle diagrams by small arrows; parent compound circle darker than others. Letter *b* indicates 'best' phase, or that of centroid of probability distribution; *m* indicates most probable phase if different from *b*; *s* indicates second most probable phase. Figures of merit or radial components of centroid are 0·96, 0·59 and 0·048, respectively.

(*a*) Typical unimodal function.
(*b*) Typical bimodal distribution.
(*c*) Poorest determination of entire 2 Å work.

Note that although Fig. 3(*b*) is bimodal, the greater sharpness of intersection as compared with 3(*a*) is reflected in a tenfold greater probability density $P_{(\varphi)}$ (before normalization).

$$\mathscr{F} = F\sum_i P_i\mathbf{r}_i/\sum_i P_i = F\mathbf{m} \tag{12}$$

by equation (7) above. Hence for the 'best' Fourier:

$$\begin{aligned}\langle\Delta\varrho^2_{hkl}\rangle &= 2F^2\sum_i P_i(\mathbf{m}-\mathbf{r}_i)^2/V^2\sum_i P_i\\ &= 2F^2\sum_i P_i(\mathbf{m}-\mathbf{r}_i).(\mathbf{m}-\mathbf{r}_i)/V^2\sum_i P_i\\ &= 2F^2\sum_i P_i(m^2-2m\cos\Delta\varphi_i+1)/V^2\sum_i P_i\\ &= (2F^2/V^2)\{m^2-2m(\sum_i P_i\cos\Delta\varphi_i/\sum_i P_i)+1\}\\ \langle\Delta\varrho^2_{hkl}\rangle &= (2F^2/V^2)(1-m^2)\,, \end{aligned} \tag{13}$$

where as before, $\Delta\varphi_i$ is the angle between $\mathbf{m}$ and $\mathbf{r}_i$. Finally, the total mean square error from all reflections is as given in equation (10).

In the EDSAC II phase programme, P_i is computed in five-degree intervals around the phase circle, and P_i, $P_i\cos\varphi_i$ and $P_i\sin\varphi_i$ are added into three storage locations, while a record is kept of the angle with greatest probability, φ_M. At the end of the sweep around the circle, m and φ_B are calculated from equations (8). In addition to the h, k and l values necessary to identify the reflections, F, φ_M, mF, φ_B and m are punched out on paper tape for each reflection in a form suitable for use as an input tape for a Fourier synthesis. Running sums of m and $F^2(1-m^2)$ can be built up so that at the end the mean figure of merit and mean square electron-density error can be computed.

In Figs. 3(*a*)–(*c*) are shown the Harker phase-circle diagrams for three reflections selected from the earlier myoglobin data to 6 Å resolution. The figures of merit for the three reflections are 0·96, 0·59 and 0·048 respectively. In a centroid Fourier synthesis the (4, 1, 3) reflection would therefore be entered with only 5% of its full contribution. This reflection, the worst in the entire 6 Å (and 2 Å) work, was in fact omitted from the 6 Å synthesis as untrustworthy. The advantage of the centroid method is that it automatically takes such considerations into account without the necessity of a long manual search through thousands of phase determinations for bad reflections. It is thus particularly suitable for use with a digital computer or with large quantities of data.

4. Refinement of relative *y* coordinates

So far it has been assumed that all parameters of the heavy atoms are known. It may be worthwhile to describe a method designed for space group $P2_1$ but capable of extension to other symmetries, by which certain coordinates may be easily be refined in the course of phase determination. In space group $P2_1$ the position of the origin is fixed in the x and z directions by space-group symmetry while in the y direction it is not. What are important are not the absolute y coordinates of the heavy atoms but relative y coordinates between different heavy atoms. In this work all coordinates were known roughly, having been obtained by methods explained in an earlier paper by Bodo *et al.* (1959). All parameters of the heavy atoms except the relative y coordinates were refined by least-squares methods using the centric $h0l$ projection data (see appendix). This section outlines the refinement of heavy-atom relative y coordinates by a least-squares procedure which seeks the best set of phase circles.

As was mentioned in section 2, the most probable phase angle φ_M is that for which the following quantity is minimized:

$$\sum_j \varepsilon_j^2(\varphi)/2E_j^2 = \sum_j w_j(F_j-D_j)^2 \tag{14}$$

with the weighting factor $w_j=1/2E_j^2$. The sum j is taken over all heavy-atom compounds for one reflection. A similar sum taken over all reflections, η, for one heavy-atom derivative provides a quantity whose minimization with respect to the parameters of the heavy atom leads to the best values of these parameters:

$$\sum_\eta W_\eta(F_{j,\eta}-D_{j,\eta})^2\,, \tag{15}$$

where W_η is an as yet unspecified weighting factor. It is possible in principle to set up the full set of normal equations and refine all the heavy-atom parameters at once (Dickerson, Kendrew & Strandberg, 1960), and such a procedure may be justified in the absence of faster refinement methods, or with faster computers. This was felt to be unnecessary with myoglobin since all but the y coordinates had been refined by two other methods, the centrosymmetric projection method and the Rossmann approximate least-squares method (see appendix). With only one variable y_j per heavy atom, the normal equation for minimizing expression (15) is:

$$\{\sum_\eta W_\eta(dD/dy_j)^2\}\,\Delta y_j = \sum_\eta W_\eta(dD/dy_j)(F_{j,\eta}-D_{j,\eta})\,. \tag{16}$$

With an approximate set of phases calculated from the initial y coordinates, a set of shifts can be computed which would best improve agreement of the individual derivatives. From these a new set of phases can be calculated and refinement of phases and y coordinates can be carried out alternately until the desired degree of convergence is reached.

For space group $P2_1$ and for several other monoclinic space groups, the calculation of dD/dy is particularly simple if the heavy-atom group can be approximated by a single isotropic atom. In this case the magnitude of f_j depends solely upon x_j and z_j while the phase, α_j, depends solely upon y_j (if f_j is allowed to be a positive or negative scalar). For space group $P2_1$ under these conditions:

$$\begin{aligned}\mathbf{f}_j &= 2f_{oj}t_j\cos 2\pi(hx_j+lz_j-\tfrac{1}{4}n)\exp 2\pi i(ky_j+\tfrac{1}{4}n)\\ &= f_j\exp 2\pi i(ky_j+\tfrac{1}{4}n)\,,\end{aligned} \tag{17}$$

where $n=0$ for k even and $n=1$ for k odd. f_{oj} and t_j are the atomic scattering and temperature factors. Then with $D_{j,\eta}$ defined as in equation (4):

$$(dD_{j,\eta}/dy_j) = (-Ff_j/D_{j,\eta}).2\pi k\sin(\varphi-\alpha_j)\,. \tag{18}$$

The calculation of dD/dy_j and the terms of equation (16) can be carried out as the phase angles are computed with virtually no loss of time, and the shifts in y_j are punched out at the end of the phase determination. Convergence can be followed by noting the size of shifts in successive cycles, the mean change in phase angles from one cycle to the next, and the gradual increase in mean figure of merit, in about this order of decreasing sensitivity.

Several weighting factors W_η were tried, but the one which gave the best convergence behavior was $W_\eta=m^2$, which is attractive on intuitive grounds since the function being minimized has the dimensions of the square of the structure amplitude.

5. Application to the myoglobin synthesis to 2 Å resolution

With four heavy-atom compounds the time required for one phase determination on EDSAC II was $3\frac{1}{2}$ sec., of which about 1 sec. was occupied in punching results. The full set of 9600 reflections required 10 hr. For this reason, relative y refinement was carried out on pseudo-random tenths of the data, on all reflections with the same specified low-order digit in the observed F's, followed by three final cycles of over one third of the data each. The limit of the refinement method seemed to be around 0·005 in relative y, or 0·15 Å, as evidenced by the tendency of different one-tenth samples to converge to slightly different values. The mean figure of merit remained around 0·61 throughout, and was not a sensitive indicator of the progress of refinement. A better indicator was the mean change in phase angles between successive cycles. For one case in which the indicated shifts in relative y's were 0·007 or 0·21 Å, the mean change in phase angles as a result of application of the shifts was about 6°. Near the end of the refinement when indicated shifts were 0·0015, or 0·045 Å, the phase angles showed a mean change of about $2\frac{1}{2}°$.

The mean figure of merit was found to decrease with increasing θ, and a plot of $\ln \langle m \rangle$ against $(2 \sin \theta)^2$ was very close to linear with a slope of $-1·52$. This fall-off is to be expected, since the outermost reflections, being both small and of short wavelength in the Fourier summation, will be most sensitive both to experimental errors and to imperfect isomorphism. Since it is these outer reflections which are the most informative about details of structure, it was decided to calculate one Fourier synthesis with an artificial modifying factor of $\exp 1·52(2 \sin \theta)^2$, which would make the radial fall-off of structure factors the same as that of the unweighted data. For two Fourier syntheses, additional sharpening factors of $\exp 2·2(2 \sin \theta)^2$ were imposed, removing all radial fall-off and giving essentially point-atom structures.

Five Fourier syntheses were calculated for restricted regions of the unit cell in order to decide which function should be used for the full electron-density map. These syntheses were calculated using coefficients with magnitudes and phases as follows:

(1) F; φ_M.
(2) $F \exp 2·2(2 \sin \varphi)^2$; φ_M.
(3) mF; φ_B.
(4) $mF \exp 1·52(2 \sin \varphi)^2$; φ_B.
(5) $mF \exp 3·72(2 \sin \varphi)^2$; φ_B.

φ_M = the most probable phase, and φ_B = the 'best' phase (see Blow & Crick).

Sections containing the haem group are shown for syntheses (1), (4) and (5) in Figs. 4(*a*)–(*c*). The unsharpened 'best' Fourier (3) showed a blurring of detail ascribable to the damping effect of the figure

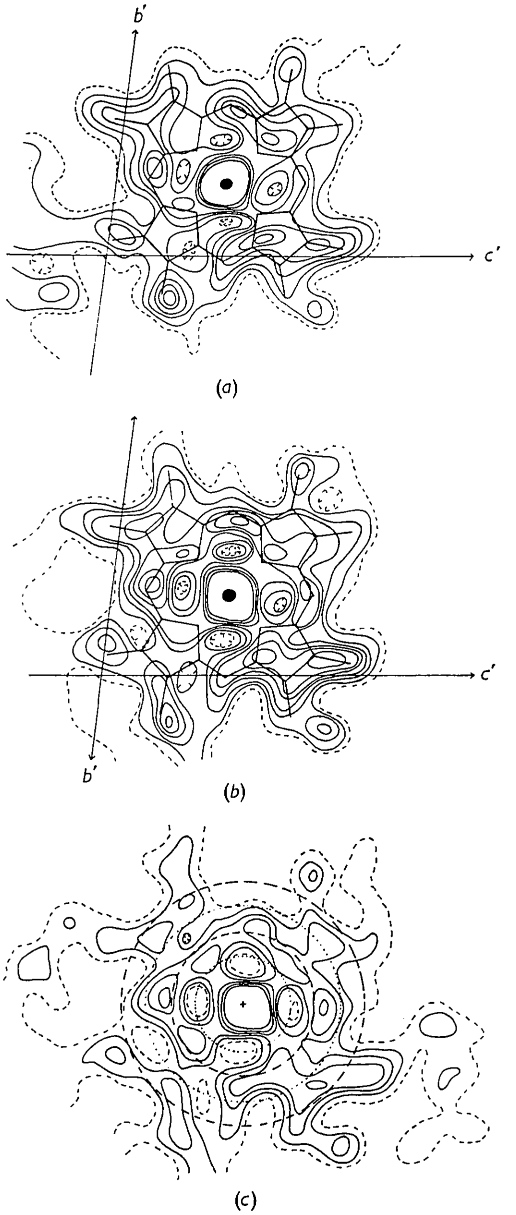


Fig. 4. Observed and theoretical electron-density distributions in plane of haem group:

(*a*) Normal Fourier synthesis, coefficients (F, φ_M), no sharpening.
(*b*) 'Best' Fourier synthesis, coefficients (mF, φ_B), sharpening function $\exp 1·52\ (2 \sin \theta)^2$.

The above two diagrams have the skeleton of an ideal porphyrin ring superimposed for comparison.

(*c*) Oversharpened point-atom 'best' Fourier-synthesis coefficients (mF, φ_B), sharpening function $\exp 3·72(2 \sin \theta)^2$. The four innermost maxima or minima of diffraction ripples produced by an abrupt cut-off of data at 2 Å have been superimposed. Dashed lines are maxima, dotted lines, minima.

Higher contours of the central iron atom are omitted.

of merit. The partially sharpened 'best' synthesis (4) and the unsharpened normal synthesis (1) were very much alike, with the former being marginally better over some parts of the unit cell. This is illustrated in Figs. 4(*a*) and (*b*); there is slightly less distortion of the electron density in Fig. 4(*b*), particularly in the upper left part of the diagram. Point-atom sharpening, as might be expected, produces objectionable diffraction ripples. In Fig. 4(*c*) the pattern of trough-crest-trough expected from a sharp cut-off at 2 Å is apparent, and this second trough produces very serious distortion in the lower left corner of the diagram. The 'best' synthesis sharpened to have the same radial structure factor fall-off as the observed data was judged to be the most satisfactory and was used for the full electron-density calculation.*

6. Conclusions

The Blow–Crick formulation of the solution of phases with the multiple isomorphous-replacement method has proved quite adaptable for use on a digital computer. Their method of selecting the proper phase angle, detecting poorly determined reflections, and weighting the Fourier synthesis accordingly, has produced results which are slightly better than an indiscriminate treatment of all phases alike, good or bad. Where the errors are greater, either experimental errors or those arising from imperfect isomorphism, the difference between equal-weight Fourier syntheses and centroid Fourier syntheses will probably be even more pronounced. The computer time required to calculate φ_M and φ_B is the same, and the 'best' Fourier technique has the very great advantage of eliminating the need for a reflection-by-reflection inspection of the phase output in order to detect poor phase determinations.

APPENDIX

Refinement of heavy atom parameters other than relative *y*'s

By R. G. Hart

Department of Zoology, Washington University,
St. Louis, Missouri, U.S.A.

Refinement of *x* and *z* coordinates, scale and temperature factors and effective atomic numbers of each heavy-atom derivative were carried out using centric *h*0*l* data, and as an independent check on the refinement, all parameters were refined by an approximate method developed by Rossmann (1960). Without preliminary phase information an exact least-squares refinement is impossible, but in the *h*0*l* zone, where there are only four possible combinations of signs of *F* and F_j, trial-and-error least-squares refinement is quite practicable.

In the EDSAC II programme developed for this purpose the function to be minimized is:

$$E = \sum_{\eta} (\Delta F_c - \Delta F_o)^2$$
$$= \sum_{\eta} \{Z_j \hat{f}_{oj} t_j \cos 2\pi(hx_j + lz_j) \pm |F| \pm k|F_j|\}^2 . \quad (19)$$

In this expression the index η refers to one *h*0*l* reflection, Z_j is an effective atomic number, $\hat{f}_{oj}$ is the unitary structure factor of heavy atom *j*, the temperature factor t_j is given by $t_j = \exp -B_j \sin^2 \theta/\lambda^2$, x_j and z_j are heavy-atom coordinates and *k* is the scale factor between parent and derivative data. It was found possible to approximate the unitary scattering factor at this resolution by an exponential function, and to collect this and the temperature factor into one term:

$$\hat{f}_{oj} t_j = \exp -B'_j (2 \sin \theta)^2 . \quad (20)$$

Starting from specified trial values of the parameters x, z, Z, B' and *k*, the parameters were varied one at a time by -2, -1, $+1$, and $+2$ units, the size of the unit shift being specified in advance for each parameter. Error sums, *E*, were built up for these twenty different combinations of parameter values plus the unshifted set. Two sign ambiguities exist in each term of the error sum of equation (19), but in each term that combination of signs was considered correct which gave the smallest value of the term. Each refinement of a heavy-atom derivative against the parent compound therefore gave a set of signs for the parent compound, whose values could be cross checked with those of other heavy-atom refinements. The agreement was found to be very good, which served to verify the validity of the refinement process.

The fact that only one parameter was varied at a time was compensated for by use of partial shifts and a method of steepest descents with five variables to a minimum *E* value. Let E_o be the error sum from the original unshifted parameter values, E_m be the minimum error sum produced by shifts in value of a parameter ξ, and ξ_m be the value of the parameter which produces this minimum error sum. Finally, let E_M be the minimum error sum produced by *any* of the given parameters. The unweighted shift indicated for this parameter would be $\Delta\xi = \xi_m - \xi_0$. The shift applied in this programme, however, was:

$$\Delta\xi' = (\xi_m - \xi_0) . (E_m - E_0)/(E_M - E_0) . \quad (21)$$

Hence a full shift was applied only to that parameter which was most influential in decreasing the error sum. Whenever all five parameters refined to within the specified shift units, these units were quartered and refinement continued without interruption. With five parameters and 700 centric reflections, each cycle required about three to four minutes computing time,

---

* A quantitative comparison of both the normal haem map and the partially sharpened 'best' map (syntheses 1 and 4) with an ideal haem group, carried out by Dr D. C. Phillips, has shown that the 'best' haem gives slightly better agreement. Details will be reported elsewhere.

and refinement was usually complete after ten to fifteen cycles.

The refined values of the parameters compared well with the Rossmann programme values, except for the effective atomic numbers, which were uniformly smaller than indicated by the Rossmann programme. The Hart $h0l$ values were used in the 2 Å Fourier synthesis, although there may be some question remaining about the proper Z values to use, as evidenced by negative regions in the myoglobin Fourier synthesis at the three heavy-atom sites and some blurring of detail in the immediate neighbourhood. This problem is currently being investigated.

Thanks are due to Dr R. G. Hart, whose heavy-atom refinement programme proved invaluable, and to Dr M. G. Rossmann, whose least-squares refinement provided an independent check on the parameters. We should also like to express our appreciation to Miss Mary Pinkerton, whose assistance was invaluable at all stages of the work.

References

Blow, D. M. (1958). *Proc. Roy. Soc.* A, **247**, 302.

Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794.

Bodo, G., Dintzis, H. M., Kendrew, J. C. & Wyckoff, H. W. (1959). *Proc. Roy. Soc.* A, **253**, 70.

Bragg, W. L. & Perutz, M. F. (1954). *Proc. Roy. Soc.* A, **225**, 315.

Dickerson, R. E., Kendrew, J. C. & Strandberg, B. E. (1960). *Symposium on Computer Methods and the Phase Problem*, p. 84. Glasgow: Pergamon Press.

Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). *Proc. Roy. Soc.* A, **225**, 287.

Harker, D. (1956). *Acta Cryst.* **9**, 1.

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. & Phillips, D. C. (1958). *Nature, Lond.* **181**, 662.

Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C. & Shore, V. C. (1960). *Nature, Lond.* **185**, 422.

Rossmann, M. G. (1960). *Acta Cryst.* **13**, 221.

Acta Cryst. (1961). **14**, 1195

The Single Isomorphous Replacement Method

By D. M. Blow and Michael G. Rossmann

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

(*Received* 8 *September* 1960 *and in revised form* 31 *October* 1960)

Despite the phase ambiguity which arises when one isomorphous pair is used to determine phases in a non-centrosymmetric structure, a single pair of compounds can be used to give an interpretable Fourier synthesis. Two conditions must be satisfied: the replacing atoms must themselves form a non-centrosymmetric array, and a sufficient number of terms must be available.

The method has been applied to the crystalline proteins haemoglobin and myoglobin. Examples are given which show the improvements which can be made by the use of weighting functions and by the introduction of anomalous-dispersion data.

1. Introduction

The isomorphous-replacement method, when applied in its conventional form to a non-centrosymmetric structure, leads to an ambiguous result for the phase angle (Bokhoven *et al.*, 1951). A general method of removing the ambiguity is to employ a series of compounds with isomorphous replacements at different sites, and this method has been used successfully with the proteins myoglobin and haemoglobin (Kendrew *et al.*, 1960; Perutz *et al.*, 1960). However, the preparation of suitable isomorphous protein derivatives has been a matter of great difficulty, and is likely to remain the most time-consuming step in a protein structure determination. It is therefore important to find methods which use the minimum number of isomorphous derivatives.

Rogers (1951) suggested a procedure applicable when only one isomorphous pair is available. Although in the form proposed it only applies to different atoms substituting at the same site, it may readily be generalized. A synthesis is calculated in which each term is given the phase of the scattering of the replacing electrons, with a sign chosen according as the intensity is increased or decreased by the replacement. When the replacing electrons form a non-centrosymmetric arrangement, this synthesis tends to be similar to the electron-density function, but with background superimposed. Rogers' function has the unfortunate property of giving strong weight to terms where the isomorphous replacement has little effect on the intensity.

Kartha & Ramachandran (1955) showed how the minimum function (Buerger, 1951) applied to the difference Patterson could in principle reveal a non-centrosymmetric structure under the same circum-