

below, for the atom O_2 is at $y = \frac{1}{4}$, and it would therefore lie over a trough if the second layer were undisplaced with respect to the first. In these circumstances the $h06$ reflexions would indicate a spacing of 7.3 Å instead of twice this. If O_1 lies over the troughs, however, alternate layers must suffer a relative displacement of $b/12$ so that a c_2 type of lattice will be produced, in accordance with the observations. The further alternative, that the O_1 and O_2 atoms might both lie at the same level and at positions midway between the crests and troughs of the layer below, can also be ruled out. This would introduce the possibility of equivalent relative shifts of $\pm b/24$ in each layer, and would therefore lead to extinction of the $hk0$ reflexions except for $k = 12n$.

Since the chrysotile layers may be expected to have a plane of symmetry perpendicular to the b axis,

there is no possibility of a further polymorph related to para-chrysotile in the same way as clino-chrysotile is related to ortho-chrysotile.

I wish to thank the Directors of Ferodo Ltd for permission to publish this paper.

References

- ARUJA, E. (1943). Ph.D. Thesis, Cambridge.
 HARGREAVES, A. & TAYLOR, W. H. (1945). *Miner. Mag.* **27**, 204.
 WHITTAKER, E. J. W. (1951). *Acta Cryst.* **4**, 187.
 WHITTAKER, E. J. W. (1953). *Acta Cryst.* **6**, 93.
 WHITTAKER, E. J. W. (1955). *Acta Cryst.* **8**, 571.
 WHITTAKER, E. J. W. (1956a). *Acta Cryst.* **9**, 855.
 WHITTAKER, E. J. W. (1956b). *Acta Cryst.* **9**, 862.
 WHITTAKER, E. J. W. & ZUSSMAN, J. (1956). *Miner. Mag.* **31**, 107.

Acta Cryst. (1956). **9**, 867

Isomorphous Replacement and Phase Determination in Non-centrosymmetric Space Groups

BY M. F. PERUTZ

*Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems,
Cavendish Laboratory, University of Cambridge, England*

(Received 24 April 1956)

This paper deals with problems arising from the method of isomorphous replacement with heavy atoms in crystalline proteins. It is assumed that a series of isomorphous compounds is available, with the heavy atoms occupying different positions in each compound. The problems which arise are twofold. The first is the determination of the position of the heavy atoms in each compound separately, and the second consists in finding their relative positions in different compounds. If centrosymmetric projections are available, both problems can be solved with the help of Fourier series using $(|F_H| - |F|)^2$ as coefficients, F_H and F being the structure factors of the protein-heavy-atom compound and of the protein respectively. In the absence of centrosymmetric projections the positions of the heavy atoms can be found in each compound separately by means of Fourier series using $(|F_H|^2 - |F|^2)$ as coefficients and employing a special device to remove the ambiguities inherent in such syntheses. When this has been done, the relative positions of the heavy atoms in different compounds can be found with the help of one of two possible correlation functions. One is a Fourier series using the products $A_1 A_2$ as coefficients, A_1 and A_2 being the real parts of the structure factors of the heavy-atom-free compound, referred to the different heavy-atom positions, or the centres between them, as origins. The other function is a Fourier series using the products

$$[(|F_{H_1}|^2 - |F|^2)(|F_{H_2}|^2 - |F|^2)]$$

as coefficients, F_{H_1} , F_{H_2} and F referring respectively to two different heavy-atom compounds and to the pure protein. The correlation functions are tested on a hypothetical case and are shown to give satisfactory results.

1. Introduction

Isomorphous replacement with heavy atoms is a method of phase determination commonly used in the structure analysis of organic compounds. In proteins

this method was first applied by Green, Ingram & Perutz (1954), but was confined to the determination of the signs of reflexions with real structure amplitudes. On account of their optically active nature all proteins crystallize in non-centrosymmetric space groups, so

that real structure amplitudes are encountered only in centrosymmetric projections. Experience with haemoglobin (Bragg & Perutz, 1954) and myoglobin (Kendrew, Dintzis & Bodo, 1956) has shown that such projections do not give sufficient information, and that three-dimensional electron-density maps will have to be prepared, involving the measurement of the phase angles of complex structure amplitudes by direct methods.

Bokhoven, Schoone & Bijvoet (1951) have envisaged the possibility of solving non-centrosymmetric structures directly by using a series of three isomorphous crystals, one crystal containing the compound under investigation by itself, the second containing the same compound with heavy atoms H_1 attached at x_1, y_1, z_1 and symmetry-related positions, and the third having heavy atoms H_2 attached at different sites x_2, y_2, z_2 and symmetry-related positions. In order to apply this method of analysis the positions of the heavy atoms must first be found relative to some arbitrarily chosen origin in each of the isomorphous crystals, and then the positions of these origins relative to each other must be determined.

In the structure analysis of organic compounds of medium molecular weight the heavy atoms can usually be found by straightforward Patterson methods, and the question of origin is unlikely to arise, because some parts of the molecular structure will usually have been solved as a result of the first isomorphous replacement, thus providing suitable reference points for locating the heavy atoms in the second replacement. In proteins this is unlikely to happen and special methods therefore have to be devised to cope with these problems.

In centrosymmetric projections of proteins the vectors between heavy atoms can be found by using difference Pattersons with $(|F_H| - |F|)^2$ as coefficients, F_H and F being the structure amplitudes of the protein-heavy-atom compound and of the protein respectively (Green *et al.*, 1954). In non-centrosymmetric projections or in three dimensions the only kind of difference Patterson available is one with $|F_H|^2 - |F|^2$ as coefficients; this gives ambiguous results due to the appearance of two kinds of vector peaks: those due to interaction between the heavy atoms themselves, and those due to interaction between the heavy atoms on the one hand and electron-density peaks in the protein on the other hand. This paper contains a suggestion for distinguishing between the two kinds of peaks.

However, even when the vectors between the heavy atoms have been found in each of the isomorphous compounds, the absence of intersecting symmetry axes may leave the origin of the vectors indeterminate, so that the relative positions of the heavy atoms in the two isomorphous compounds remain unknown. Harker (1956) has discussed a way of solving this problem by trial, using the observed intensity changes in a number of selected reflexions. In actual practice,

however, accurate parameters might be difficult to obtain by any method in which the number of reflexions used is necessarily small, on account of the substantial errors to which individual intensity measurements tend to be subject, due to imprecise isomorphism and other causes. These considerations led the author to search for some method in which the accuracy of the heavy-atom parameters would be assured by the large number of terms used, regardless of random errors in individual intensity measurements. Moreover, there is the possibility of employing groups of heavy atoms for isomorphous replacement in proteins. The relative positions of two such groups in different members of an isomorphous series would be easier to find by a vector method employing many terms, rather than a method of trial employing only a few. The following gives a description of two kinds of correlation function which should serve to determine the relative co-ordinates of heavy atoms in different crystals with a high degree of accuracy.

2. Isomorphous replacement by a single heavy atom in the asymmetric unit

Consider a triclinic crystal belonging to the space group $P1$ and having one protein molecule in the unit cell. The protein molecule contains the electron-density peaks P . In addition we have the two isomorphous compounds, one with a heavy metal H_1 at x_1, y_1, z_1 and the other with H_2 at x_2, y_2, z_2 . Their structures are shown in Fig. 1(a) and (b). We now call the structure factors of the metal-free protein F , and those of the two heavy-metal compounds F_{H_1} and F_{H_2} . Let the structure factor of the *metal-free protein*, referred to H_1 as origin, be

$$F = A_1 + iB_1 = |F| (\cos \alpha_1 + i \sin \alpha_1);$$

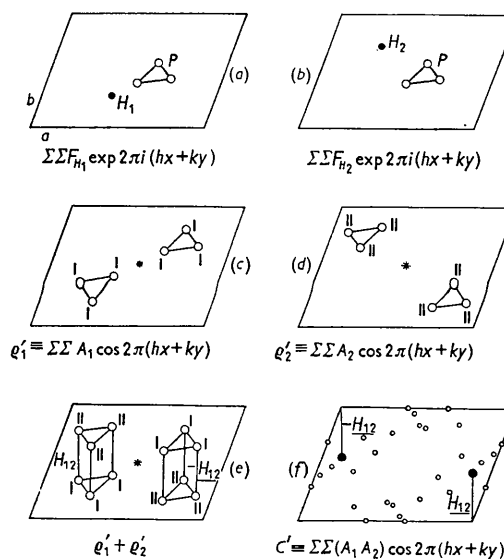


Fig. 1.

a similar expression would refer to the metal-free protein with H_2 as origin. We now wish to find the vector H_{12} , defined by

$$u_{12} = x_2 - x_1, \quad v_{12} = y_2 - y_1, \quad w_{12} = z_2 - z_1.$$

The structure factor of H_1 and H_2 are referred to as f_1 and f_2 .

$$|F_{H_1}|^2 = (A_1 + f_1)^2 + B_1^2 = A_1^2 + 2A_1f_1 + f_1^2 + B_1^2, \quad (1)$$

$$|F|^2 = A_1^2 + B_1^2. \quad (2)$$

Subtracting (2) from (1), we obtain

$$A_1 = \{|F_{H_1}|^2 - |F|^2 - f_1^2\} / 2f_1; \quad (3)$$

similarly

$$A_2 = \{|F_{H_2}|^2 - |F|^2 - f_2^2\} / 2f_2. \quad (4)$$

We now calculate a Fourier series with the products $(A_1A_2)_{hkl}$ as coefficients. For reasons to be explained below, the result is a correlation function which will be called C' and which shows prominent peaks at $\pm(u, v, w)_{12}$, thus giving us the vector H_{12} :

$$C'(u, v, w) = \frac{1}{V} \sum_h \sum_k \sum_l (A_1A_2)_{hkl} \cos 2\pi(hu + kv + lw). \quad (5)$$

Alternatively a Fourier series may be calculated using the products $[(|F_{H_1}|^2 - |F|^2)(|F_{H_2}|^2 - |F|^2)]_{hkl}$ as coefficients:

$$C(u, v, w) = \frac{1}{V} \sum_h \sum_k \sum_l [(|F_{H_1}|^2 - |F|^2)(|F_{H_2}|^2 - |F|^2)]_{hkl} \times \cos 2\pi(hu + kv + lw). \quad (6)$$

In the particular case of the structures illustrated in Fig. 1 the result of (6) is almost identical with that of (5), at least as far as the prominent peaks at $\pm(u, v, w)_{12}$ are concerned.

The origin of the vector peaks in these correlation functions will now be explained with the help of diagrams. Take first the function C' , and consider the structure of the two isomorphous compounds shown in Fig. 1(a) and (b). Suppose now we determine A_1 and A_2 as defined by equations (3) and (4) and then calculate the two Fourier series

$$\varrho'_1 = \frac{1}{ab \sin \gamma} \sum_h \sum_k A_1 \cos 2\pi(hx + ky), \quad (7)^*$$

and

$$\varrho'_2 = \frac{1}{ab \sin \gamma} \sum_h \sum_k A_2 \cos 2\pi(hx + ky). \quad (8)$$

The results would be two electron-density maps, each showing the real structure together with its inversion image at the position of the heavy atom as centre of symmetry (Fig. 1(c) and (d)). Suppose ϱ'_1 and ϱ'_2 are now added together, as is done graphically

* Equations (7), (8), (9), (10) and (15) are written in the form of two-dimensional series to correspond with the projections shown in the figures.

in Fig. 1(e), then the resulting density map will contain two sets of peaks separated by the vector H_{12} . In order to find this vector, the author first thought of calculating a Patterson of $(\varrho'_1 + \varrho'_2)$, i.e. calculating a Fourier series with $(A_1 + A_2)^2$ as coefficients. However, among the three elements contained in such a series, i.e. $A_1^2 + 2A_1A_2 + A_2^2$, the terms A_1^2 and A_2^2 merely represent the self-Patterns of ϱ'_1 and ϱ'_2 , and are of no interest. The important terms are those containing the product A_1A_2 . A Fourier series calculated with these products as coefficients gives the correlation function

$$C'(u, v) = ab \sin \gamma \iint \varrho'_1(xy) \varrho'_2(x+u, y+v) dx dy, \quad (9)$$

which is represented diagrammatically in Fig. 1(f). This contains all possible vectors between peaks marked I and II in Fig. 1(e), but none of them occurs more than twice, with the only exception of the vector H_{12} , which occurs three times, i.e. once for each of the three protein peaks P . It is obvious from this figure that if the protein contained n electron-density peaks P , the vector H_{12} would recur n times.

Though the terms involved in the function C at first sight look quite different from those in C' , the arguments relating to the two functions are similar. Consider again the structure of Fig. 1(a) and (b), and suppose we now calculate two difference Pattersons called Π_1 and Π_2 , with $(|F_{H_1}|^2 - |F|^2)$ and $(|F_{H_2}|^2 - |F|^2)$ as coefficients. Such Pattersons contain vector peaks HH due to interaction between the heavy atoms, in addition to those marked HP , due to interaction between the heavy atoms and the protein. In the present case the peak HH coincides with the origin, and all non-origin peaks are of the type HP . Thus Π_1 and Π_2 form a pair of maps (Fig. 2(a) and (b)) differing from

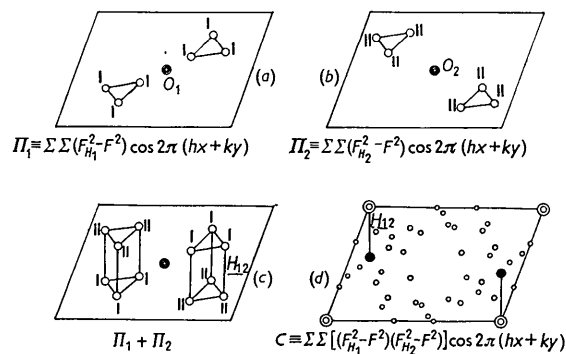


Fig. 2.

ϱ'_1 and ϱ'_2 only in the appearance of peaks at the origin. Suppose Π_1 and Π_2 are now added together, then the vector H_{12} again appears once for each of the three protein peaks P (Fig. 2(c)). In analogy with equation (9), a Fourier series using $[(|F_{H_1}|^2 - |F|^2)(|F_{H_2}|^2 - |F|^2)]$ as coefficients will give a correlation function of the form

$$C(u, v) = ab \sin \gamma \iint \Pi_1(xy) \Pi_2(x+u, y+v) dx dy, \quad (10)$$

in which H_{12} appears as the only prominent peak (Fig. 2(d)).

The function C contains two kinds of peak not present in C' : one is a slightly larger origin peak corresponding to interaction of the origin peaks in Π_1 and Π_2 ; the others correspond to interaction between the origin peak in Π_1 and the non-origin peaks in Π_2 , and vice versa. No vector other than H_{12} occurs more than twice.

When the vector between the heavy atoms has been found by either of the correlation functions C' or C the way is open for determining the phase angles $\alpha(hkl)$. We now define ω as

$$\omega = \pm 2\pi(hu_{12} + kv_{12} + lw_{12}). \quad (11)$$

The sign chosen for ω is arbitrary and decides whether the final structure will be a right-handed or left-handed enantiomorph. The phase angles of the metal-free protein referred to H_1 and H_2 as origin now work out as follows:

$$\cos \alpha_1 = \{ |F_{H_1}|^2 - |F|^2 - f_1^2 \} / 2|F|f_1, \quad (12)$$

$$\begin{aligned} \cos \alpha_2 &= \cos(\alpha_1 - \omega) = \cos \alpha_1 \cos \omega + \sin \alpha_1 \sin \omega \\ &= \{ |F_{H_2}|^2 - |F|^2 - f_1^2 \} / 2|F|f_1, \end{aligned} \quad (13)$$

$$\sin \alpha_1 = \{ \cos \alpha_2 - \cos \alpha_1 \cos \omega \} / \sin \omega. \quad (14)$$

This is all the information required to determine α . The equations show, however, that $\sin \alpha$ will be indeterminate if ω is either 0 or π .

3. Isomorphous replacement by acentric constellations of heavy atoms

If the protein is very large, isomorphous replacement by a single heavy atom attached to each protein molecule may not produce intensity changes sufficiently big to determine phase angles, in which case it may become necessary to introduce groups of heavy atoms into the protein. In other cases the chemistry of the protein molecule may be such that several sites have an equal affinity for a given heavy atom. It is important therefore to consider isomorphous replacement by non-centrosymmetric constellations of heavy atoms. This case is shown diagrammatically in Fig. 3(a) and (b), where the protein molecule, represented by the same three peaks P as before, has attached to it two different constellations of heavy atoms H_i and H_j .

The first step consists in finding the heavy-atom positions in each compound separately. This can be done with the help of Fourier series using $(|F_{H_i}|^2 - |F|^2)$ as coefficients (Fig. 3(c) and (d)) and employing the following device for distinguishing between the two kinds of vector peak HH and HP . Two kinds of crystal

are prepared of each of the heavy-atom compounds H_iP and H_jP : one in which all the protein molecules in the crystal have a group of heavy atoms attached to it, and another in which only half the protein

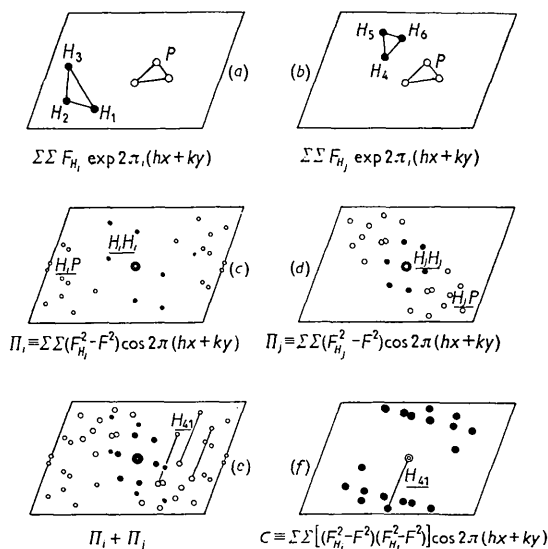


Fig. 3.

molecules are combined with that group, the other half, distributed at random through the crystal, containing no heavy atoms. Judged by the experience gained in the X-ray analysis of organic compounds of moderate molecular weight, the possibility of preparing a series of such crystals of varying heavy-atom content might appear remote. In proteins, however, this problem presents little difficulty. So long as the crystals of P , H_iP and H_jP are strictly isomorphous, those of intermediate composition will also be isomorphous. In effect these will be mixed crystals whose structure factor will be called $F_{H_{ij}}$. Two kinds of Fourier series are now calculated, using respectively $(|F_H|^2 - |F|^2)$ and $(|F_{H_{ij}}|^2 - |F|^2)$ as coefficients. The resulting contour maps will allow a distinction to be made between vector peaks HH and HP , for the reason that the height of HH varies as the square of the heavy atom content of the crystal, whereas that of HP varies as a linear function of it. When the positions of H_i and H_j have thus been found individually and relative to some arbitrarily chosen origin in each crystal, the relative positions of H_i and H_j may be found from equation (15):

$$C(u, v) = \frac{1}{ab \sin \gamma} \sum_h \sum_k [(|F_{H_i}|^2 - |F|^2)(|F_{H_j}|^2 - |F|^2)] \cos 2\pi(hu + kv). \quad (15)$$

This function gives a contour map containing the vectors H_{ij} and H_{ji} as the only prominent peaks. The explanation for the prominent appearance of these vectors, and the comparative insignificance of all vectors peaks of the type HP , may be seen with the

help of the same kind of argument as was used in § 2 above. Consider again the two difference syntheses Π_1 and Π_2 in Fig. 3(c) and (d). The vectors HP arise from the superposition of three sets of triangles P , displaced from each other by the vectors H_{ii} and H_{ij} . If Fig. 3(c) and 3(d) are now superimposed, as in Fig. 3(e), the vectors between triangles H_iP and H_jP are seen to occur in sets of three, corresponding to the three electron-density peaks P of the protein. Altogether there are 18 such sets of three parallel vectors, corresponding to all the interactions H_{ii} and H_{ji} . Thus, if the protein contained n electron-density peaks P , any one vector of the type H_{ij} would recur n times. The number of different vectors of other kinds is now very large, but except for accidental coincidences none of them occurs more than twice; these other vectors would therefore be expected to form a more or less uniform background for the prominent vector peaks H_{ij} . The correlation function (15) contains all the information on the relative positions of H_i and H_j which the intensities differences between the three isomorphous compounds can give. If the configuration of the individual constellations H_i and H_j is known, then the interpretation of the correlation function should not be difficult. The cross-Patterson of the two constellations H_i and H_j is drawn; this consists of a pattern of vector peaks which is determined by the known relative orientation of H_i and H_j and is independent of their relative positions. A search is then made in the correlation function C for this pattern of peaks. The position of this pattern in relation to the origin peak gives all the vectors required.

The phase angles α for the metal-free compound can now be determined as follows. First the structure factors $|f_i|$ and $|f_j|$, and the phase angles φ_i and φ_j , of the heavy-atom constellations H_i and H_j have to be calculated relative to some arbitrarily selected origin. The value of $\cos \alpha$ and $\sin \alpha$ can then be found from equations (16) and (17), using either graphical or analytical methods:

$$\begin{aligned} \cos(\alpha - \varphi_i) &= \cos \alpha \cos \varphi_i + \sin \alpha \sin \varphi_i \\ &= \frac{|F_{H_i}|^2 - |F|^2 - |f_i|^2}{2|F||f_i|}, \quad (16) \end{aligned}$$

$$\begin{aligned} \cos(\alpha - \varphi_j) &= \cos \alpha \cos \varphi_j + \sin \alpha \sin \varphi_j \\ &= \frac{|F_{H_j}|^2 - |F|^2 - |f_j|^2}{2|F||f_j|}. \quad (17) \end{aligned}$$

The enantiomorphous form of the protein determined by the phase angles will depend on the way H_i and H_j are placed relative to the chosen origin. No other ambiguity arises in the interpretation.

4. Application to different crystal symmetries

4.1. Space group $P2$

Suppose the unit cells of two of the three isomorphous compounds contain the heavy atoms H_1

and H_2 in general positions $x_1, y_1, z_1; \bar{x}_1, y_1, \bar{z}_1$ and $x_2, y_2, z_2; \bar{x}_2, y_2, \bar{z}_2$ (Fig. 4(a) and (b)). The x and z

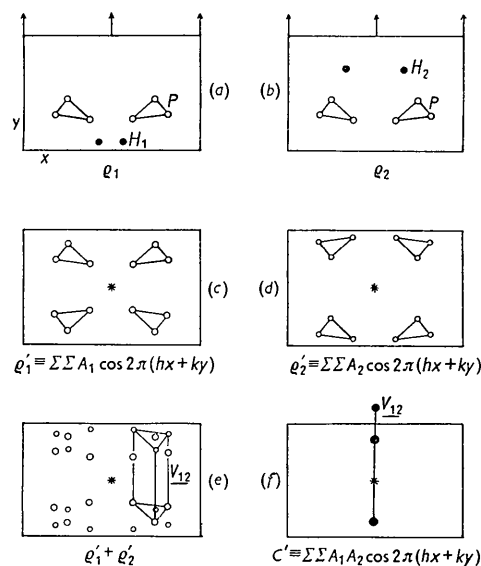


Fig. 4.

co-ordinates of the heavy atoms can be found from $(|F_H| - |F|)^2$ syntheses on the b plane. $v = y_2 - y_1$ may be found by means of either of the two correlation functions C or C' , but whereas the results of these functions are almost identical in the space group $P1$, they are different in $P2$.

Using the expression

$$C'(0, v, 0) = \frac{1}{b} \sum_k \left[\sum_h \sum_l (A_1 A_2) \right] \cos 2\pi kv, \quad (18)$$

the vector between the heavy atoms is found from a linear section along v through the three-dimensional correlation function C' . The linear section is sufficient, because A_1 and A_2 are calculated for the points $0, y_1, 0$ and $0, y_2, 0$ respectively, i.e. for the centres between the heavy atoms as origins (Fig. 4(c) and (d)). The vector V_{12} in Fig. 4(e) and (f) therefore represents the distance between these two centres, and not the vector between the heavy atoms themselves.

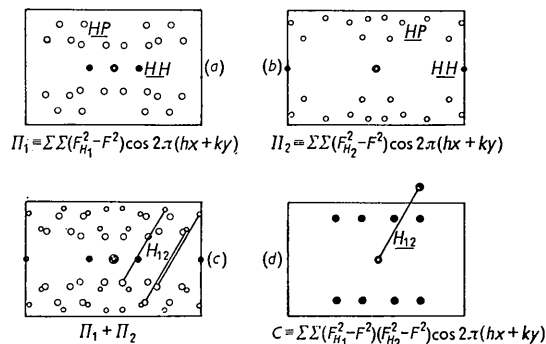


Fig. 5.

The correlation function C , on the other hand, does give the vectors between the heavy atoms themselves, as shown in Fig. 5. The positions of the vector peaks can be found from one-dimensional sections along v at known co-ordinates u and w :

$$C(u, v, w) = \frac{1}{V} \sum_h \sum_k \sum_l [(|F_{H_1}|^2 - |F|^2)(|F_{H_1}|^2 - |F|^2)] \cos 2\pi(hu + kv + lw). \quad (19)$$

If the unit cell contains more than one heavy atom per asymmetric unit, then these atoms are likely to form non-centrosymmetric constellations, in which case their positions have to be found by the procedure outlined in the preceding section. The relative positions of the heavy atoms in the two isomorphous crystal can be found by a series of one-dimensional sections of the form of equation (19). The problems arising in the other enantiomorphous space groups in the monoclinic system are very similar to those in $P2$ and need not be specially discussed.

4.2. Space group $R3$

The case is similar to that of the non-centrosymmetric constellation in $P1$, in so far as no centrosymmetric projection is available and heavy atoms placed in general positions will necessarily form non-centrosymmetric constellations. The procedures which could be used to find their positions would be the same as those outlined in § 3 above, except that approximate co-ordinates of the heavy atoms could be obtained from projections of the correlation functions on the c plane and along the c axis.

5. A practical test

At the time of writing, experimental material for a practical test of the correlation functions described in this paper was not yet available. It was decided, therefore, to use a hypothetical case adapted as far as possible to conditions which might actually be encountered in proteins. This consisted in attaching two pairs of mercury atoms to haemoglobin at different sites, calculating the structure factors of the mercury compounds, modifying them by the addition of random experimental errors of an order of magnitude rather larger than would be expected in practice and using the inaccurate intensities of the mercury compounds thus obtained to find the vector H_{12} . The accuracy of the phase angles determined by this method was also tested.

The procedure adopted was as follows. Measured absolute values of $F^2(0kl)$ of horse methaemoglobin (space group $C2$) were used, and each reflexion was given an arbitrary phase angle chosen at random. The choice of random phase angles is not quite satisfactory, since the phase angles in a real structure are inter-related, depending on the height and distribution of

the electron-density peaks; but without knowing the structure this choice was the best that could be adopted. Atoms of atomic number 68 were then placed at the following positions:

$$H_1 \text{ at } y_1 = 0, z_1 = \pm 0.29; y_1 = \frac{1}{2}, z_1 = \pm 0.29;$$

$$H_2 \text{ at } y_2 = 0.13, z_2 = \pm 0.33; y_2 = 0.63, z_2 = \pm 0.33.$$

$F_{H_1}^2$ and $F_{H_2}^2$ were now calculated for these parameters. Next a random experimental error (r.m.s. $\partial F = 10\%$) was added to or subtracted from each F_H , and a set of 'experimentally measured' F_H^2 values was calculated and used to work out A_1 and A_2 . Of a total of 58 reflexions, 12 had to be discarded because either H_1 or H_2 made too small a contribution to determine A . This left 46 products $A_1 A_2$ of which 17 were zero, thus leaving 29 terms in all for the calculation of $C'(v, 0)$ according to equation (18). The plain curve in Fig. 6 shows the result, which takes the

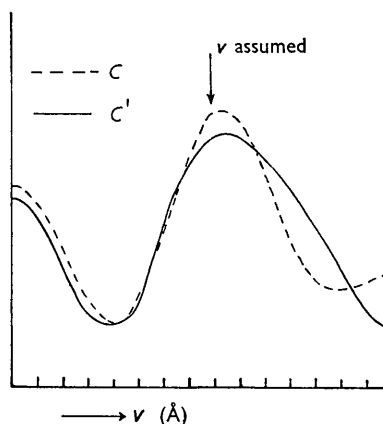


Fig. 6.

form of a linear section along v through the projection on the a plane. The curve has a dominant peak close to the expected position, the error in v being 0.5 \AA . This is satisfactory in view of the small number of terms used and the omission of terms below 6 \AA spacing. Many more terms, including $0kl$ as well as $h k 0$ reflexions, could of course have been included in the Fourier series, but this would have made the amount of labour involved in the test unreasonably large.

The difference between the values of z_1 and z_2 (0.04) chosen for this test is very small. In consequence the function $C(v, w)$ would be expected to exhibit two peaks H_{12} close to the y axis. These would overlap sufficiently to produce a peak in a linear section along y .

The function

$$C(v, 0) = \frac{1}{b} \sum_k \sum_l [\sum_l (|F_{H_1}|^2 - |F|^2)(|F_{H_2}|^2 - |F|^2)] \cos 2\pi kv \quad (20)$$

was therefore calculated and is shown as the broken curve in Fig. 6. It has a peak at the same point as $C'(v, 0)$.

Having found H_{12} , it was decided to calculate the phases with the given 'experimental' values of $|F_{H_1}|^2$ and $|F_{H_2}|^2$. This work showed that the r.m.s. error of 10% in F_H resulted in an r.m.s. error in the phase angles of 30° . This seemed large at first sight, but closer examination of the individual reflexions showed the usual error for medium and strong reflexions to be of the order of only 10° . The r.m.s. error of 30° for all reflexions arises from a few large errors occurring in several very weak reflexions, and from a single error of 148° in a reflexion where a wrong sign would have been given to α through misinterpretation of the data. Out of a total of 58 reflexions, 10 had real structure amplitudes. Of the remaining 48, the phases of 14 reflexions were left in doubt and one was wrongly determined, thus leaving 33, or roughly two thirds of the reflexions, correctly measured. To measure the remaining third, it would be necessary to prepare a fourth member of the isomorphous series, with heavy atoms H_3 at positions different from H_1 and H_2 . The amount of labour involved in such an analysis would be very great, but it does offer a feasible method of solving the structure of a protein directly.

It is difficult to judge at this early stage whether sufficient numbers of different isomorphs can be prepared to apply this method, and whether their degree of isomorphism will be sufficiently good to allow accurate determination of phase angles. On the whole, there is good ground for optimism. The molecules in wet protein crystals are loosely packed and have only few points of contact. The interstices between them thus contain room even for large heavy-metal complexes to be attached without affecting the unit-cell dimensions. Recent experience with haemoglobin (Dintzis, 1956) and myoglobin (Kendrew *et al.*, 1956) indicates that a great variety of crystalline heavy-metal complexes can be prepared. In monoclinic crystals a good criterion of isomorphism is to be found in the reliability factor $R(h0l) = \frac{\sum(|\partial F_o| - |\partial F_c|)}{\sum|\partial F_o|}$, where $|\partial F| = |F_H| - |F|$. This factor has been found to vary greatly in different heavy-metal complexes and in different proteins. In some derivatives it is

0.3 or less, while in others it is much higher, or rises with increasing angle θ , even in the absence of any changes in unit-cell dimensions; this effect shows that the presence of metal complexes may cause slight changes in the structure of protein molecules without affecting their packing in the crystal lattice.

For three-dimensional analysis a series of heavy-metal derivatives will have to be selected in which $R(h0l)$ is sufficiently small to promise accurate results. So far, no relationship has been worked out between the magnitude of $R(h0l)$ and the accuracy of the phase angles of the general reflexions, or between possible errors in the phase angles and the resulting errors in the electron-density distribution. It is hoped that further development of the work will clarify these problems.

The development of this work from the germ of an idea to a practical method of analysis owes much to two suggestions made by Dr F. H. C. Crick and Dr W. Cochran. Dr Crick suggested that I should try the product A_1A_2 at a time when I was experimenting with coefficients of the form $(A_1 + A_2)^2$, and Dr Cochran remarked at a colloquium that coefficients of the form $[(|F_{H_1}|^2 - |F|^2)(|F_{H_2}|^2 - |F|^2)]$ would also be worth trying. Finally Sir Lawrence Bragg suggested the trial on a hypothetical case. I am most grateful for all this help and for the great interest shown in this work by all my colleagues.

References

- BOKHOVEN, C., SCHOONE, J. C. & BIJVOET, J. M. (1951). *Acta Cryst.* **4**, 275.
 BRAGG, W. L. & PERUTZ, M. F. (1954). *Proc. Roy. Soc. A*, **225**, 315.
 DINTZIS, H. M. (1956). Unpublished.
 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., DINTZIS, H. M. & BODO, G. (1956). Unpublished.