Acta Cryst. (1960). **13**, 221

The Accurate Determination of the Position and Shape of Heavy-Atom Replacement Groups in Proteins

By Michael G. Rossmann

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

(*Received* 29 *June* 1959)

If two isomorphous heavy-atom derivatives of a protein are available with structure factors F_1 and F_2 respectively, then a Patterson synthesis with coefficients $(|F_1|-|F_2|)^2$ will give accurate information about the relative position and shape of the heavy-atom peaks in the two compounds.

Introduction

Green, Ingram & Perutz (1954) introduced the method of isomorphous replacement with heavy atoms to the X-ray analysis of the centro-symmetric zone of horse haemoglobin. In order to determine the phase angle of general reflections multiple isomorphous replacement has to be used, involving the preparation of a series of isomorphous compounds each bearing one or more heavy atoms placed on different sites (Kendrew *et al.*, 1958). The 'heavy atoms' in proteins may often consist of sizeable organic residues to which one or more heavy atoms are attached, producing irregular 'lumps' on an electron-density map.

Patterson coefficients $(|F_{\text{Heavy}}|-|F|)^2$ can be used to determine the position of the heavy atoms in centrosymmetric projections. However, for non-centrosymmetric crystals there exists the problem of determining the position of one atom relative to another without a fixed symmetry element as origin. The first solution of this problem is due to Harker (1956) who considers trial-and-error techniques based on selected reflections. Bragg (1958) suggested a method of fitting sinusoidal curves to the envelope of the differences plotted as a function of $\sin\theta$, which can then be shown to correspond to the structure-factor graph of the heavy-atom content of the unit cell. Perutz (1956) proposed two correlation functions, later modified and improved by Blow (1958), which depend on calculating certain Fourier summations from which the heavy-atom positions can be deduced. The great advantage of the Perutz–Blow approach is that an overall picture of the cell is obtained, so that subsidiary sites can be discovered. The present paper deals with another Fourier technique which is believed to possess the advantage of giving less background effect, thus leaving no doubt in placing the atomic centre within about 0·4 Å from data with 6Å resolution.

Theory

Let the vectors $\mathbf{F}(OP)$, $\mathbf{F}_1(OQ)$, $\mathbf{F}_2(OR)$ be the representation on the Argand diagram of the structure factors of the plane (hkl) for the unsubstituted compound, and for the isomorphous heavy atom derivatives 1 and 2 respectively (Fig. 1). Thus if $\mathbf{f}_1$ and $\mathbf{f}_2$ are the heavy-atom contributions alone in compounds 1 and 2, then $\mathbf{F}_1=\mathbf{F}+\mathbf{f}_1$ and $\mathbf{F}_2=\mathbf{F}+\mathbf{f}_2$. Let QR be the vector $\mathbf{f}$, let $\angle\, ROQ=\varphi$ and let $\angle\, RPQ=\psi$.

It will be shown that a Patterson synthesis with $(|F_1|-|F_2|)^2$ as coefficients is approximately equivalent to the self-Patterson of the heavy atoms in compound 1, plus the self-Patterson of the heavy atoms in compound 2, minus the cross-Patterson between the heavy atoms in compounds 1 and 2. That is to say, the Patterson will give positive peaks at the end of vectors between atoms in the same compound, but negative peaks at the end of vectors between atoms in the two different compounds.

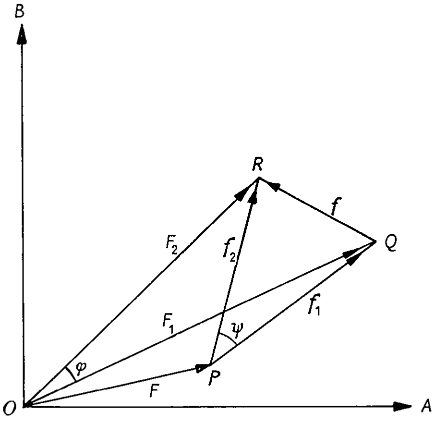

Fig. 1. Vector diagram showing relationship between structure factors in different isomorphous compounds.

In ΔROQ

$$f^2=F_1^2+F_2^2-2F_1F_2\cos\varphi\;.$$

The difference in phase, φ, between the phase angles of the heavy-atom derivatives 1 and 2 will be small if the heavy-atom contribution in each of the compounds is only a small fraction of the total scattering material, as must inevitably occur in the majority of reflexions in a protein structure. Thus to a first order approximation $\cos\varphi=1$. Hence

$$f^2 \simeq F_1^2+F_2^2-2F_1F_2$$
$$\simeq (F_1-F_2)^2,$$

where F_1 and F_2 are the magnitudes of the measured structure factors.

I am indebted to Sir Lawrence Bragg for pointing out that the above approximation is not strictly true because $\cos\varphi$ tends to unity only as a consequence of **f** being small. Nevertheless the terms which have greatest effect on the Patterson summation, namely those for which $(F_1-F_2)^2$ is large, are the terms for which the approximation is most nearly true since $(F_1-F_2)^2$ can only be large if $\mathbf{F}_1$, $\mathbf{F}_2$ and **f** are all collinear, which implies that $f^2=(F_1-F_2)^2$. Further the mean value of f is greater than $\frac{1}{2}(f_1+f_2)$ since the mean value of the modulus of the sum of sinusoidal waves is always greater than $\frac{1}{2}$, although the average approaches $\frac{1}{2}$ as the number of waves (atoms) increase. Thus, to summarize, the large terms in the Patterson summation contain least approximation, and there is a slightly greater probability of occurrence of large terms rather than small terms.

It is therefore possible to determine the quantity f^2 to a first-order approximation from measured data. If we now calculate a Patterson synthesis with coefficients $(F_1-F_2)^2$, it is equivalent to a Patterson synthesis of the structure given by the structure factors **f**.

Let us consider the heavy-atom structure 1 in the unit cell of the protein by itself. This will give rise to a structure factor $\mathbf{f}_1$, for a particular reflection. Similarly the heavy-atom structure 2 by itself in the identical cell gives rise to the structure-factor vector $\mathbf{f}_2$.

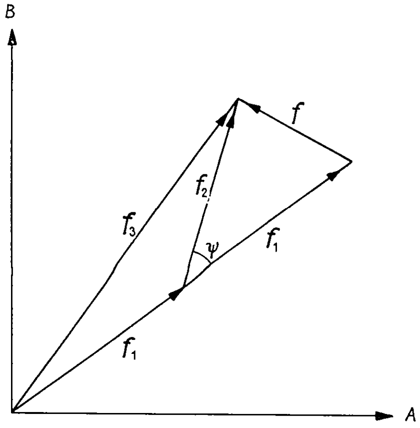

Fig. 2. Vector diagram of heavy atom contributions in the protein unit cell.

Hence if both heavy-atom structure 1 and 2 are present in the unit cell simultaneously, we would obtain a structure-factor vector $\mathbf{f}_3=\mathbf{f}_1+\mathbf{f}_2$. Inspection of Fig. 2 shows that

$$f_3^2=f_1^2+f_2^2+2f_1f_2\cos\psi\,.$$

Thus a Patterson synthesis of the structure equivalent to $\mathbf{f}_3$ is equal to the sum of the Patterson synthesis of heavy-atom structure 1 by itself (f_1^2), plus that of heavy-atom structure 2 by itself (f_2^2), plus that with coefficients $2f_1f_2\cos\psi$. Hence the last of these can only represent the vectors relating the heavy atoms of structure 1 to the heavy atoms of structure 2 inside the cell. This we shall term a 'cross-Patterson'. Again by inspection of Fig. 2 we see that:

$$f^2=f_1^2+f_2^2-2f_1f_2\cos\psi\,.$$

Hence a Patterson synthesis with coefficients f^2 is equivalent to the sum of

(i) The f_1^2 Patterson, or 'self-Patterson' of heavy atoms in structure 1,
(ii) The f_2^2 Patterson, or 'self-Patterson' of heavy atoms in structure 2,
(iii) The $-2f_1f_2\cos\psi$ Patterson, or minus the 'cross-Patterson' between heavy atoms in structures 1 and 2.

However, an f^2 synthesis has been shown to be roughly equivalent to an $(F_1-F_2)^2$ synthesis; thus a Fourier summation with coefficients $(F_1-F_2)^2$ will have negative peaks which determine the relative position of the heavy atoms in compounds 1 and 2.

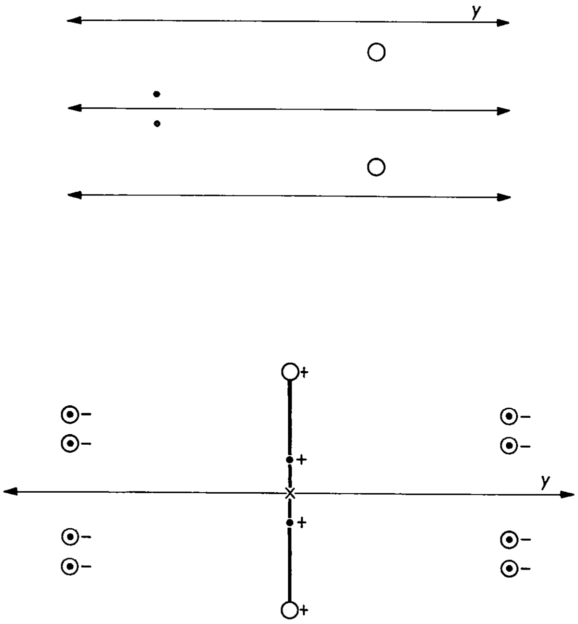

Fig. 3. (*a*) Point diagram of a hypothetical structure 1 (dots) and structure 2 (circles) in space group $P2$. (*b*) Point position of peaks in the $(F_1-F_2)^2$ correlation function between structure 1 and 2.

Let us consider the space group $P2$. Let us say that the first derivative has a heavy atom at x_1, y_1, z_1, with a symmetry related heavy atom at $\bar{x}_1, y_1, \bar{z}_1$. Similarly the second derivative will have heavy atoms at x_2, y_2, z_2 and $\bar{x}_2, y_2, \bar{z}_2$ (Fig. 3(*a*)). Thus single positive 'self-Patterson' peaks should appear at $2x_1, 0, 2z_1$ and $2x_2, 0, 2z_2$ plus equivalent peaks related by the centre of symmetry at the origin (Fig. 3(*b*)). These are in fact Harker peaks, but if there is more than one atom in the asymmetric unit positive Harker as well as positive non-Harker peaks would appear. Single negative, cross-Patterson, peaks will appear at (x_1+x_2), (y_1-y_2), (z_1+z_2) and at (x_1-x_2), (y_1-y_2), (z_1-z_2) and equivalent positions related by the Patterson symmetry of space group $P2$, namely $P2/m$ (Fig. 3(*b*)). It is now apparent that, provided each of the derivatives has only one heavy atom per asymmetric unit, the positive self-Patterson peaks appear in the plane $y=0$, while the negative cross-Patterson peaks appear in the plane $y=y_1-y_2$.

Effect of scaling

One difficulty of the isomorphous-replacement technique is the determination of the absolute scale of intensities for each compound. From the point of view of a correlation function as described in this paper it is only necessary to put the two compounds on the same relative scale. Thus we could write our Patterson coefficients as $(F_1-KF_2)^2$ where K determines the scaling between compounds 1 and 2. Hence the density, ϱ, at any point in the summation is given by:

$$\varrho(x, y, z) = \sum_h \sum_k \sum_l (F_1-KF_2)^2 \cos 2\pi(hx+ky+lz)$$

$$\therefore \quad \frac{\partial \varrho}{\partial K} = -2\sum_h \sum_k \sum_l (F_1-KF_2)F_2 \cos 2\pi(hx+ky+lz)$$

$$= -2\sum_h \sum_k \sum_l \delta F \,.\, F_2 \cos 2\pi(hx+ky+lz)\;,$$

where δF is the difference in magnitude of the structure factors between the two isomorphous compounds, and is small. Therefore, $\partial\varrho/\partial K$ is also small, and hence an accurate knowledge of scale factor is not required. Consider in contrast both the Perutz type correlation functions which can be summarized as having coefficients of the type $\{F_1^2-(KF_2)^2\}$. Here

$$\varrho'(x, y, z) = \sum_h \sum_k \sum_l (F_1^2-K^2F_2^2) \cos 2\pi(hx+ky+lz)$$

$$\therefore \quad \frac{\partial \varrho'}{\partial K} = -2\sum_h \sum_k \sum_l (KF_2)F_2 \cos 2\pi(hx+ky+lz)\;.$$

Now in general $KF_2 \gg \delta F$; thus $\partial\varrho'/\partial K \gg \partial\varrho/\partial K$, which implies that the Perutz type of correlation function is $(KF_2/\delta F)_{\text{mean}}$ times as sensitive to the exact value of relative scale factor as the function proposed in this paper. For horse haemoglobin this factor would be approximately five. Indeed Blow (1958) and Bodo, Dintzis, Kendrew & Wyckoff (1959) did have much trouble in determining accurate scale factors.

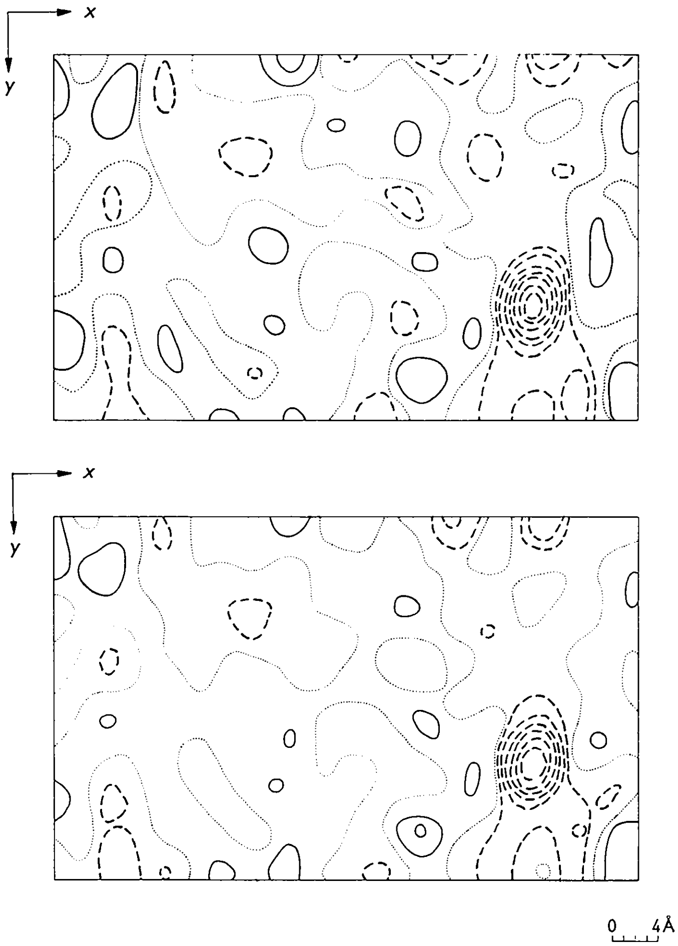

Fig. 4. Section $z=c/8$ through a negative correlation peak in an example taken from horse haemoglobin. The top figure assumes a relative scale factor $K=1{\cdot}000$, and the lower figure uses $K=0.955$.

Fig. 4 shows the effect of a change of 4·5% on the scale factor K for the $(F_1-KF_2)^2$ function. It will be seen that there is little difference in the sections taken through one of the negative correlation peaks, in an example taken from horse haemoglobin, calculated with the two different scale factors.

Application to horse haemoglobin

A number of different heavy-atom derivatives of horse haemoglobin have been made (Cullis, Dintzis & Perutz, 1957) which crystallize in the monoclinic cell of space group $C2$ with dimensions

$$a=109{\cdot}2,\; b=63{\cdot}2,\; c=54{\cdot}7\ \text{Å};\; \beta=110{\cdot}7°.$$

Three-dimensional $(F_1-F_2)^2$ correlation functions relating the different heavy-atom derivatives were calculated using about 1,200 independent terms, representing a sphere in reciprocal space of radius corresponding to about 5·5 Å. A number of examples are given here to illustrate the new method. Full details of the results will be presented in a later paper.

Figs. 5 and 6 show a composite view of the Patterson syntheses correlating three different heavy-atom compounds of haemoglobin looking down through one quarter of the unit cell. Fig. 5 represents the correlation between 'HgAc$_2$' (one Hg atom per asymmetric unit at $x=0{\cdot}16$, $y=0{\cdot}00$, $z=0{\cdot}18$) against '*PCMB*' (one Hg atom per asymmetric unit at $x=0{\cdot}07$, $y=0{\cdot}16$, $z=0{\cdot}29$), while Fig. 6 shows the correlation between HgAc$_2$ and '*DMA*' (two mercury atoms per asymmetric unit 3·5 Å apart with mean co-ordinates $x=0{\cdot}06$, $y=0{\cdot}14$, $z=0{\cdot}29$). The latter example is interesting because the orientation of the line joining the two mercury atoms of *DMA* can be deduced from the shape of the negative peaks. The presence of the larger *DMA* group can also be seen in the positive self-Patterson peak.

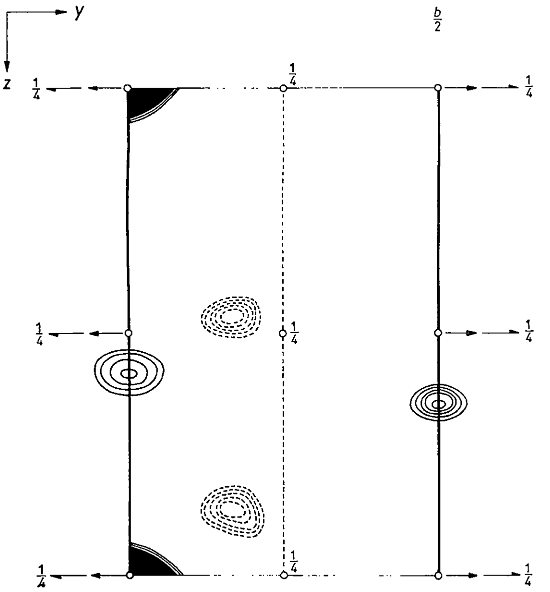

Fig. 5. Composite view of the $(F_{\mathrm{HgAc_2}}-F_{PCMB})^2$ correlation function unit cell between the $x=0$ and $x=a/4$ layers. Contours are at arbitrary levels. Continuous lines are positive, dashed lines are negative contours. The zero and ± first contours are omitted.

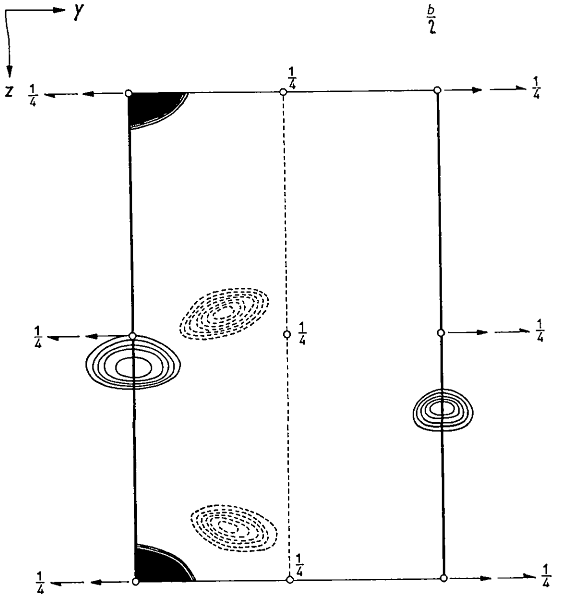

Fig. 6. Composite view of the $(F_{\mathrm{HgAc_2}}-F_{DMA})^2$ correlation function unit cell between the $x=0$ and $x=a/4$ layers. Contours are at the same levels as in Fig. 5. The '*DMA*' compound differs from the '*PCMB*' compound in that it has two mercury atoms in close proximity, instead of only a single one in about the same position.

Fig. 7(*a*) shows the section $y=\frac{5}{32}b$ on which the cross-Patterson peaks are expected to be situated. It shows the correlation of HgAc$_2$ against *DMA*, thus giving negative peaks, while Fig. 7(*b*) shows the correlation of a compound containing both HgAc$_2$ and *DMA* attached to the same haemoglobin molecule against the unsubstituted protein, which therefore gives positive peaks in the same section.

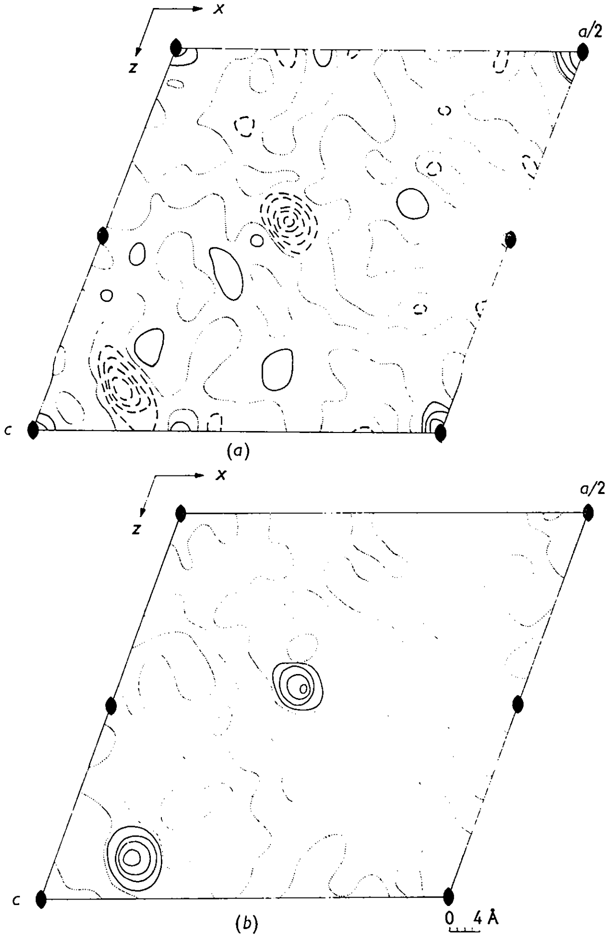

Fig. 7. (*a*) Section $y=\frac{5}{32}b$ through $(F_{\mathrm{HgAc_2}}-F_{DMA})^2$ correlation function. (*b*) Section $y=\frac{5}{32}b$ through $[F_{(DMA+\mathrm{HgAc_2})}-F_{\mathrm{unsubstituted}}]^2$ correlation function. Contours at arbitrary levels. Continuous lines are positive and dashed lines are negative contours. Contours are at the same levels in (*a*) and (*b*).

The exact positions of the maxima of all peaks were determined by graphical interpolation. The positive and negative peaks of the correlation functions give two independent estimates for each of the co-ordinates of every atom. These co-ordinates can be further checked by correlating each compound against each of the two others. Thus four estimates for each of the co-ordinates of the heavy atom in any one compound can be made. The standard deviation between the different estimates of any one co-ordinate for all the heavy atom derivatives we have worked on was never greater than 0·40 Å, but generally was less than 0·25 Å, which is a very reasonable achievement when compared with the kind of accuracy obtained in the crystallography of smaller compounds. For instance if observations are cut off at the limit of the copper sphere, giving a minimum spacing of 0·65 Å, a standard deviation of about 0·02 Å in atomic co-ordinates can be expected. In our present study we have a cut-off of just under $10 \times 0{\cdot}65$ Å and a standard deviation of just over $10 \times 0{\cdot}02$ Å in atomic positions.

Suppose that we correlate a derivative containing one heavy atom of equivalent atomic number Z_1 against a derivative of equivalent atomic number Z_2. Then the two self-Patterson peaks will have heights which are roughly proportional to Z_1^2 and Z_2^2, while the cross-Patterson peaks will have heights of roughly $-Z_1Z_2$. This relationship was found to hold fairly accurately. The origins were very roughly of size 2·5 $(Z_1^2+Z_2^2)$. The largest subsidiary peak was found to be along the y-axis of the Patterson in certain of the functions, but it was always less than $\frac{2}{3}$ in height of the actual correlation peaks. In general, however, subsidiary peaks were less than half of the correlation peak heights.

Determination of the shape of the heavy atoms

The self-Patterson peaks of the heavy atoms in each compound had always had the same shape, irrespective of the compound against which the correlation was made. In addition the self-Patterson peaks of the $(F_1-F_2)^2$ functions in three dimensions had the same cross-section as the peaks in $(F_1-F_2)^2$ Pattersons of the centric $h0l$ reflexions. Thus it is possible to state that, provided sufficient terms are introduced into the summation, the atomic shapes are reproduced fairly faithfully by the $(F_1-F_2)^2$ correlation function.

The peaks for horse haemoglobin are by no means spherical, perhaps due to a certain amount of randomness of the heavy-atom positions. In the case of one of the compounds (Fig. 6) the elliptical shape of the peak is due to two mercury atoms in close proximity. As a first approximation it was assumed that the heavy atoms in all the compounds can be regarded as ellipsoids, so that the effective scattering factor, f_e, of an atom may be represented by

$$f_e = Z \exp\left[-(\beta_1h^2+\beta_2k^2+\beta_3l^2+\beta_4hk+\beta_5kl+\beta_6lh)\right]$$

for the reflexion hkl. Here Z is the equivalent atomic number of the atom and $\beta_1\ldots\beta_6$ are the constants which determine the shape and orientation of an ellipsoidal atom. Hence every atom is completely determined by ten parameters $Z, \beta_1\ldots\beta_6$ and x, y, z.

A least-squares analysis, based on the validity of the $(F_1-F_2)^2$ correlation function, was used to determine all ten parameters for each of the heavy atoms. For this purpose the function

$$E = \sum_{\text{all reflexions}} w[(F_1-KF_2)^2-f^2]^2$$

was minimized by means of cyclic iterations.

Now $\mathbf{f}=\mathbf{f}_2-\mathbf{f}_1$. But since the atomic parameters determine $\mathbf{f}_2$ and $\mathbf{f}_1$, it follows that f^2 is completely determined by the ten parameters per atom. The negative sign between $\mathbf{f}_2$ and $\mathbf{f}_1$ signifies that all atoms in structure 1 have to be entered with a negative atomic number.

K, the relative scale factor between the two isomorphous compounds, was treated as an additional variable. It was assumed that all parameters are independent variables, and hence all but the diagonal terms of the least-squares metrix were neglected. It soon appeared that this assumption was far too rough, causing considerable difficulty in choosing a reasonable weighting factor, w, because of the interaction between the shape parameters $\beta_1\ldots\beta_6$ and the equivalent atomic number, Z. After many experiments with other functions, the weighting factor $w(h, k, l) = (F_1-KF_2)^2$ seemed the most reasonable. Such a scheme emphasizes those reflections with large differences. However, the greater the difference between $|F_1|$ and $|F_2|$ the smaller must be the phase angle φ (Fig. 1), between these vectors, for $\mathbf{f}$ has a maximum size given by the sum of the scattering factors for the particular reflexion over all the heavy atoms in the unit cell. Hence w is largest when the basic approximation $f^2=(F_1-F_2)^2$ is most nearly true.

Convergence was improved by means of separate damping factors (Rossmann, Jacobson, Hirshfeld & Lipscomb, 1959) on positional, thermal, size (Z), and scaling parameters. One least-squares cycle on EDSAC 2 for the 1,200 independent differences required four minutes. About twenty cycles were required to achieve convergence. Reasonable agreement was obtained for atomic parameters of a particular compound, refined against different derivatives. The precise analysis of the results of these parameters will be given elsewhere.

All the data on which the results of this paper are based were collected by Dr M. F. Perutz and Miss A. F. Cullis. The observed structure factors have an accuracy of greater than $0{\cdot}05F$, which is about twice as accurate as is usual in crystallographic work, yet the very fragile crystals have to be mounted in vapour, in sealed capillaries, at accurately defined conditions of the mother liquor. I am also much indebted to Dr M. F. Perutz for his encouragement of this work, as well as to Mr M. Wells for writing a Fourier programme, and the University of Cambridge Mathematical Laboratory for making available the electronic digital computer EDSAC 2.

References

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

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

Bragg, W. L. (1958). *Acta Cryst.* **11**, 70.

Cullis, A. F., Dintzis, H. M. & Perutz, M. F. (1957). *Conference on Haemoglobin*, p. 50. NAS-NRC Publication 557, Washington.

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.

Perutz, M. F. (1956). *Acta Cryst.* **9**, 867.

Rossmann, M. G., Jacobson, R. A., Hirshfeld, F. L. & Lipscomb, W. N. (1959). *Acta Cryst.* **12**, 530.

Acta Cryst. (1960). **13**, 226

X-ray, Dielectric, and Optical Study of Ferroelectric Lead Metatantalate and Related Compounds

E. C. Subbarao, G. Shirane and F. Jona

Westinghouse Research Laboratories, Pittsburgh, Pennsylvania, U.S.A.

(*Received* 24 *April* 1959 *and in revised form* 5 *June* 1959)

Single crystals of ferroelectric lead metatantalate, $PbTa_2O_6$, have been grown using $Pb_2V_2O_7$ as a flux. An X-ray and optical study shows that $PbTa_2O_6$ has orthorhombic symmetry with

$$a_0 = 17{\cdot}68,\ c_0 = 7{\cdot}754\ \text{Å},\ b_0/a_0 = 1{\cdot}002$$

at room temperature. The ferroelectric axis is perpendicular to the [001] direction. Dielectric constant and birefringence along the three directions have been measured as a function of temperature through the Curie point at 265 °C. The crystals remain orthorhombic in the paraelectric region.

A dielectric and X-ray study of the solid solution systems $Pb(Ta,Nb)_2O_6$, $(Pb,Ba)Nb_2O_6$, and $(Pb,Sr)Ta_2O_6$ shows the existence of a phase boundary in the first two systems, separating different ferroelectric modifications. Besides the ferroelectric modification, $PbTa_2O_6$ has a non-ferroelectric, rhombohedral form.

1. Introduction

Investigations following the discovery of ferroelectricity in barium titanate, $BaTiO_3$, have revealed ferroelectric properties in a number of double oxides. These include lead metatantalate, $PbTa_2O_6$, and lead metaniobate, $PbNb_2O_6$, which are ferroelectric with Curie temperatures at 260 °C. (Smolenskii & Agranovskaya, 1954) and 570 °C. (Goodman, 1953), respectively. $PbNb_2O_6$ belongs to the orthorhombic system with

$$a_0 = 17{\cdot}65,\ b_0 = 17{\cdot}91,\ c_0 = 7{\cdot}736\ \text{Å}$$

at room temperature. The large unit cell made it difficult to understand the structural implications of the occurrence of ferroelectricity in this crystal. Recently Francombe & Lewis (1958) have carried out dielectric, optical, and X-ray studies on single crystals of $PbNb_2O_6$. These authors have drawn attention to the close structural resemblance between the paraelectric phase of $PbNb_2O_6$ and some tetragonal tungsten bronzes. These structures bear a similarity to the perovskite structure in possessing BO_6 octahedra surrounding an A ion. Looking at the structure of $PbNb_2O_6$ in this manner, it may be possible to understand the ferroelectric behavior of this crystal.

Compared with the rather detailed studies made on $PbNb_2O_6$, relatively little information is available on the related compound, $PbTa_2O_6$, since the latter has been studied only in the polycrystalline form. The purpose of the present investigation is to carry out X-ray, dielectric, and optical studies on single crystals of $PbTa_2O_6$. This study has been extended to the solid solution system $PbTa_2O_6$–$PbNb_2O_6$, in order to elucidate any structural distinction between these two ferroelectric compounds.

2. Polymorphism and crystal growth

Studies on the polymorphism of $PbNb_2O_6$ had revealed the existence of a non-ferroelectric modification in addition to the ferroelectric form. In the case of $PbTa_2O_6$, only the ferroelectric modification had been reported (Smolenskii & Agranovskaya, 1954; Francombe & Lewis, 1958). However, an X-ray examina-