Acta Cryst. (1965). 18, 212

The Combination of Isomorphous Replacement and Anomalous Scattering Data in Phase Determination of Non-Centrosymmetric Reflexions

By A. C. T. North

Medical Research Council External Staff, Davy Faraday Research Laboratory, 21 Albemarle Street, London, W.1, England

(*Received* 30 *January* 1964)

Earlier attempts to combine anomalous scattering data and isomorphous replacement data in phase determination are discussed. The relative weights given to the two types of observation have not been on a satisfactory basis; either an arbitrary method has been used or the intrinsic accuracy of the anomalous scattering data has not been realized. A new method is proposed, which allows appropriate weighting to be applied.

Many of the heavy atoms used in preparing isomorphous derivatives of proteins give rise to appreciable anomalous scattering effects. For instance, with Cu $K\alpha$ radiation the imaginary part of the atomic scattering factor of mercury is about 0·12 of the real part and of uranium 0·18. When an isomorphous series of crystals is available, the intensity differences between substituted and unsubstituted crystals are a more powerful aid to phase determination than the smaller intensity differences between Friedel pairs of reflexions within one substituted crystal, but the disparity is not as great as would be suggested by the above figure of 0·12 or 0·18; for, in comparing intensity data from different crystals, errors arise from scaling the sets of data together and from inaccuracies in the absolute values of absorption effects (or errors in the absolute values of absorption corrections); in comparing Friedel pairs from a single crystal, scaling problems are less acute or non-existent and only relative absorption factors are important. Indeed, if the crystal morphology is favourable, absorption may be neglected entirely. Thus in horse haemoglobin, space group $C2$, four reflexions would be equivalent if Friedel's law held: hkl, $\bar{h}\bar{k}l$, $h\bar{k}l$, $\bar{h}k\bar{l}$. The pair hkl, $\bar{h}\bar{k}\bar{l}$ have absorption factors that are equal to each other but different from those of the pair $h\bar{k}l$, $\bar{h}k\bar{l}$. Anomalous scattering makes the two members of each pair unequal and the intensity difference due to anomalous scattering, free of crystal absorption errors, is given by $(I_{hkl}-I_{\bar{h}\bar{k}\bar{l}})$ or $(I_{h\bar{k}l}-I_{\bar{h}k\bar{l}})$ or, better, their mean.

When anomalous scattering data are combined with data from isomorphous replacement for the purpose of phase determination, it is important to be able to allow for the intrinsically greater accuracy of the anomalous scattering intensity differences. This paper describes a method for doing this. It will be assumed that the unsubstituted crystals do not give rise to appreciable anomalous scattering.

Determination of phases by the multiple isomorphous replacement method

It is well known that a single isomorphous replacement leads to two possible values of the phase angle. A second isomorphous replacement also leads to two possible values, one of which should be the same as one of the first pair if experimental errors are negligible. Unfortunately, errors are not usually negligible and Blow (1958) and Blow & Crick (1959) have shown that it is useful to construct a phase-probability distribution for each reflexion. The method is illustrated by Fig. 1. Let the observed values of the

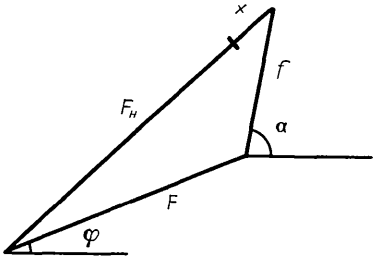

Fig. 1. x represents the discrepancy between the observed amplitude F_H and the resultant of the vectors **F** and **f**.

structure amplitudes of the unsubstituted and the substituted crystals be F, F_H. Let the calculated contribution of the additional atoms in the substituted compound be $\mathbf{f}=f\exp 2\pi i\alpha=\mathbf{a}+i\mathbf{b}$. Consider the probability of the phase associated with F being φ. The amplitude of the vector sum of **F** and **f**, F_c, will not in general be equal to F_H but will be equal to $F_H+x(\varphi)$, where $x(\varphi)$ represents the 'lack of closure' of the vector triangle. $x(\varphi)$ is given by F_c-F_H where

$$F_c^2 = F^2 + f^2 + 2Ff\cos(\varphi - \alpha)$$
$$= F^2 + f^2 + 2F(a\cos\varphi + b\sin\varphi)$$
$$\text{i.e. } x(\varphi) = -F_H + [F^2 + f^2 + 2F(a\cos\varphi + b\sin\varphi)]^{\frac{1}{2}}. \quad (1)$$

Blow & Crick show that, if E represents the r.m.s. error associated with the measurements, the relative probability $P(\varphi)$ of any particular φ being correct is given by

$$P(\varphi) = \exp[-x(\varphi)^2/2E^2]. \quad (2)$$

A curve showing the relative probability of all phase angles can thus be derived, and where two or more isomorphous replacements are available the joint probability is given by multiplying the separate distributions together. According to Blow (1958) *the value of E, being a measure of the difference between theory and observation, includes errors from all sources: non-isomorphism, inaccurate weighting and placing of the heavy atoms, wrong scale factors, and so on, as well as observational error in F and F_H. E* can be estimated from the discrepancies between the observed and calculated heavy atom contributions of centrosymmetric reflexions, for which of course **F**, $\mathbf{F}_H$ and **f** must be collinear. Given the phase probability distribution for each reflexion, Blow & Crick show that it is possible to calculate the electron-density Fourier map of the structure in two ways. Either the most probable value of each phase angle may be used, giving rise to the 'most probable Fourier' or the centroid of each probability distribution may be used to give the 'best Fourier', a weighted Fourier synthesis which is expected to give the minimum mean square difference in electron density from the true Fourier synthesis. This latter approach, discussed in more detail by Blow & Crick, has been found the more satisfactory in the determination of the structure of the proteins myoglobin (Kendrew, Dickerson, Strandberg, Hart, Davies, Phillips & Shore, 1960) and haemoglobin (Cullis, Muirhead, Perutz, Rossmann & North, 1961, 1962). 'Best' Fourier syntheses have also been used in the structure determinations of lysozyme (Blake, Fenn, North, Phillips & Poljak, 1962) and chymotrypsinogen (Kraut, Sieker, High & Freer, 1962).

Combination of anomalous scattering with a single isomorphous replacement

The case of a single isomorphous substitution, where the anomalous scattering may be used to indicate which of the two possible solutions is the correct one, was described by Bijvoet (1954). Fig. 2 shows vector diagrams for a Friedel pair of reflexions. The diagrams are mirror images in respect of the **F** vector and the **f** vector, the real part of the heavy atom contribution. The imaginary part, δ, has a similar sense in the two diagrams and the resultant $\mathbf{F}_H$ vectors, $\mathbf{F}_H^+$ and $\mathbf{F}_H^-$ are of different lengths. As Blow & Rossmann (1961) pointed out, the situation can conveniently be represented (Fig. 3) by superposing the mirror image of

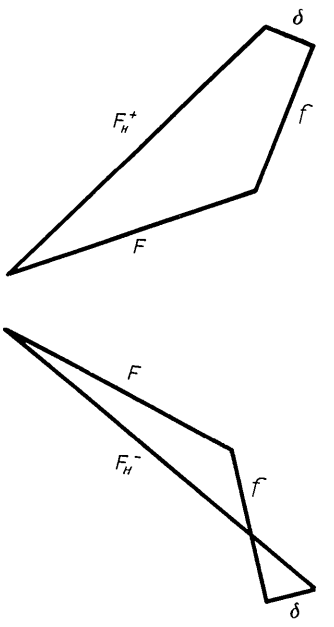

Fig. 2. Vector diagrams for a Friedel pair of reflexions.

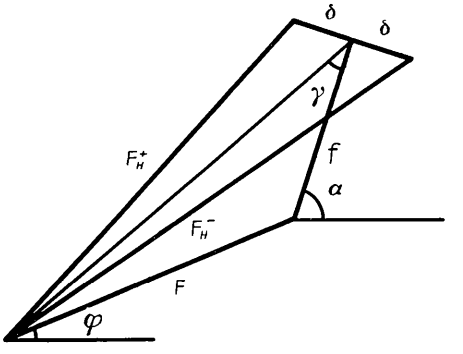

Fig. 3. Mirror image of the vector diagram of the $\bar{h}\bar{k}\bar{l}$ reflexion superposed on the vector diagram of the *hkl* reflexion.

the $\bar{h}\bar{k}\bar{l}$ diagram on to the *hkl* diagram, in which case the anomalous scattering component can be thought of as giving rise to a retardation in phase angle for the $\bar{h}\bar{k}\bar{l}$, and an advance for the *hkl* reflexion. If $\mathbf{F}_H$, the resultant vector in the absence of anomalous scattering, makes an angle γ with **f**, then

$$F_H^{+2} = F_H^2 + \delta^2 - 2\delta F_H \sin\gamma$$

and

$$F_{\bar{H}}^{-2} = F_H^2 + \delta^2 + 2\delta F_H \sin\gamma$$

so that

$$F_H^{+2} - F_{\bar{H}}^{-2} = -4\delta F_H \sin\gamma. \quad (3)$$

As δ is normally small compared with F_H,*

* If δ is not small compared with F_H, then f will be of comparable magnitude to or greater than F_H. Then, unless F also is small, the two solutions from isomorphous replacement are very close together and it is not very important to distinguish between them. If F is small, a lower accuracy in phase determination is acceptable.

$$F_H^+ - F_{\bar{H}}^- = \Delta_H = -2\delta \sin\gamma \,. \quad (4)$$

Thus, examination of the sign of Δ_H indicates whether γ lies in the range 0–π or π–2π, thereby often permitting one of the solutions of the phase diagram to be preferred. The reliability of the test is a maximum when γ is near to $\pi/2$ or $3\pi/2$ and becomes lower as γ approaches zero or π. When γ is very close to zero or π, *i.e.* $\mathbf{F}_H$ and $\mathbf{f}$ are nearly collinear, the two solutions of the phase triangle are close together and it is not important to distinguish between them, but for many reflexions with intermediate values of γ, Δ_H may not be sufficiently significant compared to the experimental errors in the intensity data to permit one of the solutions to be preferred with any confidence. Thus, although a single pair of isomorphous crystals, differing by an anomalous scatterer, can in principle be used for phase determination, experimental errors may allow only a proportion of phases to be found in this way.

Combination of anomalous scattering with multiple isomorphous replacement data

In his determination of the phase angles of the non-centrosymmetric [100] zone of horse haemoglobin, Blow (1958) used anomalous scattering data to supplement the phase probability curves derived from multiple isomorphous replacement, assessing the significance of the anomalous scattering differences in terms of the observational errors. Highly significant differences indicated that some ranges of phase angles could be ruled out and others permitted. In this way it was possible in a number of cases to exclude one of the peaks of a bimodal probability distribution and to find the most probable phase angle unambiguously. Blow himself pointed out that this treatment was rather arbitrary.

In the determination of the three-dimensional structure of horse haemoglobin (Cullis, Muirhead, Perutz, Rossmann & North, 1961, 1962) both 'most probable' and 'best' Fourier electron-density maps were calculated. Phase determination was again carried out by constructing probability curves from the multiple isomorphous replacement data. When the most probable phase angle had been deduced, the anomalous scattering data were examined to see whether they conflicted with it. In cases of conflict, the two types of data were carefully assessed to decide which was the more reliable; again the result was obtained in a non-rigorous and rather arbitrary way. For the 'best' Fourier synthesis, it was not considered feasible to make use of the anomalous scattering data in calculating the centroid of the probability distribution, so that the phases were calculated from the isomorphous differences alone.

A method of combining anomalous scattering and isomorphous replacement data in order to derive the centroid of the probability distribution in a more rigorous manner was described by Blow & Rossmann (1961) for the single isomorphous replacement case but has also been used in the determination of the structure of lysozyme by multiple isomorphous replacement (Blake, Fenn, North, Phillips & Poljak, 1962). The method is based on the fact that the mirror image of the Argand diagram of the $\bar{h}\bar{k}\bar{l}$ reflexion is similar to the Argand diagram of the hkl reflexion but for the reversal of the sense of the imaginary part of the heavy atom contribution. The data for the $\bar{h}\bar{k}\bar{l}$ reflexions may therefore be treated as if they came from a separate isomorphous compound with parameters identical with those of the original compound but with the opposite sign for the imaginary component of the atomic scattering factor. In the lysozyme phase determination, intensities of the Friedel pairs of reflexions were measured for each of the three heavy-atom compounds and the problem was treated as if there had been a total of six heavy-atom compounds. The method was found to be satisfactory, but analysis of the phases showed that the anomalous scattering data had played comparatively little part in determining the positions of the centroids of the probability distributions. In order to determine the absolute configuration of the crystal, two separate sets of phases had been calculated, based on the two possible enantiomorphous space groups, $P4_12_12$ and $P4_32_12$; it had been anticipated that the inclusion of the anomalous scattering data

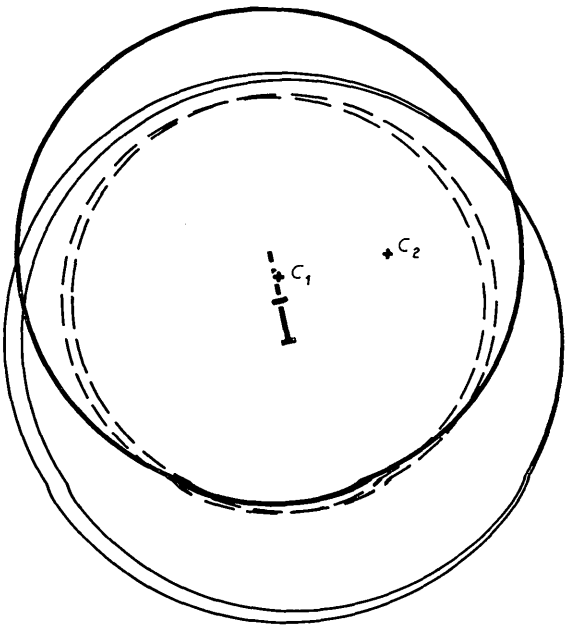

Fig. 4. Phase circles for a lysozyme reflexion. The **f** vectors of the two substituted compounds are fortuitously nearly collinear. The short perpendicular line at the end of each **f** vector represents the positive and negative δ vectors. Broken lines are used for the **f** vectors and phase circles of compound 1, solid lines for compound 2. A heavy circle is used for the unsubstituted compound.

in the way described would have led to markedly sharper phase probability curves for the correct space group, but the difference was found to be extremely marginal in general, although some reflexions were affected considerably, whereas the highly significant anomalous scattering changes consistently indicated the space group $P4_32_12$.

The situation is illustrated by data for one of the reflexions. Fig. 4 shows a vector diagram for the unsubstituted crystal and two substituted compounds isomorphous with it. Two circles, corresponding to the Friedel pair of reflexions, have been drawn for each of the substituted compounds, in the way described by Blow & Rossmann (1961). The intersections of the circles for compound 1 (broken circles) with that for the unsubstituted crystal give rise to a single very broad probability maximum. Each circle for compound 2 (solid circles) gives rise to two probability maxima that are well defined but about 180° apart; the probability curves for compound 2 are shown in Fig. 5 (broken lines) and it can be seen that one pair

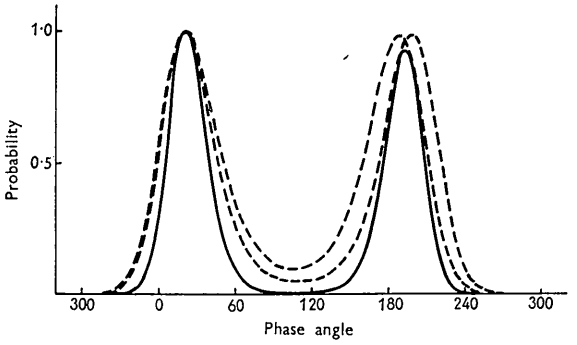

Fig. 5. The broken lines are the phase probability curves for the Friedel pair of reflexions from compound 2, derived by the Blow & Rossmann method. The solid line is the joint probability curve. The curves are not normalized.

of maxima nearly coincide but the other pair are separated by about 10°. The anomalous scattering data therefore appear to favour one maximum decisively, but in fact the two maxima of the joint probability curve (solid line, Fig. 5) differ little in size because the widths of the individual maxima are of the same order as their separation. The reason for this is that the anomalous scattering change is small compared with the value E used for the r.m.s. error in structure amplitude. As described above, the E's were estimated from the difference between observed and calculated heavy atom contributions. They are of a suitable magnitude to use in the comparison of structure amplitudes from isomorphous crystals but they are overestimates of the much smaller errors involved in the comparison of Friedel reflexions from one crystal. Examination of the lysozyme data (Blake, Fenn, North, Phillips & Poljak, unpublished data) shows that the r.m.s. errors appropriate to the anomalous scattering differences from one crystal are as little as a quarter of the values appropriate to heavy atom differences, so that it is clear that the full potentialities of the anomalous scattering data are not realized if the hkl and $\bar{h}\bar{k}\bar{l}$ data are treated as if they belong to different crystals. An alternative method of combining anomalous scattering and isomorphous replacement data is now proposed.

Improved method for combining anomalous scattering and multiple isomorphous replacement data

The phase probability distribution function for each compound containing an anomalously scattering heavy atom may be considered to be the product of two distribution functions, the first relating to the structure amplitude differences between the substituted and unsubstituted compounds and the second to the differences, due to anomalous scattering, between Friedel pairs of reflexions from the substituted compound. Structure amplitude differences should be compared to r.m.s. errors E for the first function, determined as previously described, and E' for the second function, determined for example by comparison of Friedel pairs of reflexions from centrosymmetric zones of the substituted crystal (or alternatively from Friedel pairs of reflexions from the unsubstituted crystal, which is assumed to contain no anomalous scatterers). The first function is calculated as previously, from equations 1 and 2, with f, a and b obtained from the real part of the atomic scattering factor and F_H given, accurately enough for our purpose, by the mean of F_H^+ and F_H^-; this function will give a probability distribution, in general bimodal, symmetrical about the **f** vector.

The second function may be derived analogously to the first. For a given phase angle φ, the difference $x'(\varphi)$ between the observed anomalous scattering component Δ_H and the calculated component, given by equation (4), is

$$x'(\varphi)=\Delta_H+2\delta\sin\gamma\,. \qquad (5)$$

Now,

$$\sin\gamma=(F/F_c)\sin(\alpha-\varphi)=(F/F_cf)(b\cos\varphi-a\sin\varphi)\,.$$

Therefore,

$$x'(\varphi)=\Delta_H+2(F\delta/F_cf)(b\cos\varphi-a\sin\varphi)\,. \qquad (6)$$

The values of $x'(\varphi)$ and E' may be used to calculate the second function by use of an equation similar to equation (2).

This second distribution function will be symmetrical about the δ vector and will tend to reduce the symmetry of the first function.

In this way, the joint probability distribution may be obtained rigorously and with appropriate weighting for the two types of data.

Equation (6) has been derived on the assumption that the phase triangle is perfectly closed at phase

angle φ, so that $F_c = F_H$. In other cases closure can only be achieved by assuming errors in one or all of F, F_H and f and it is not obvious whether it is more appropriate to use F_H for F_c or to calculate F_c from $\mathbf{F}_c = \mathbf{F} + \mathbf{f}$. However, the main requirement of the method is to weight appropriately the peaks of the heavy atom difference function, and since F_c and F_H are approximately equal in the region of the peaks, it appears to be valid to use the constant value F_H for F_c in calculating the probability function from (6).

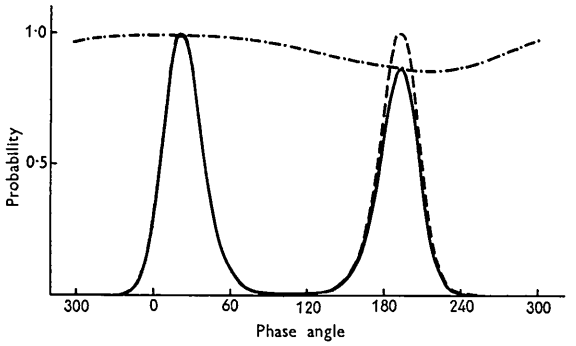

Fig. 6. The broken line is the isomorphous replacement phase probability curve derived from the mean of F_H^+ and F_H^- for compound 2. The chain line is the anomalous scattering probability curve. Effectively equal E values were used for the two curves. The solid line is the joint probability.

The new expression has been used to recalculate the phase probability curves for the reflexion discussed above. Firstly (Fig. 6) to simulate the Blow & Rossmann method, equal values of E were used for the heavy atom differences and anomalous scattering difference curves. (In practice, it was necessary to divide the E for the first of these by $\sqrt{2}$, to allow for the fact that each F_c was the mean of two observations). The weighting effect of anomalous scattering on the two peaks can now clearly be seen. The joint probability curve is not significantly different from the original one.

For Fig. 7, the E' for the anomalous scattering probability curve was reduced to a third, a more appropriate value. The anomalous scattering curve now shows greater variation, so that one of the maxima of the joint probability curve is reduced to about a quarter of the other. The change in the effective weight of the anomalous scattering data has little effect on the probability curve for compound 1, for which the anomalous scattering maximum lies roughly in the middle of the broad heavy atom difference maximum. The change in the curve for compound 2 is therefore predominant, causing the position of the centroid of the combined distribution to move from C_1 to C_2 (Fig. 4).

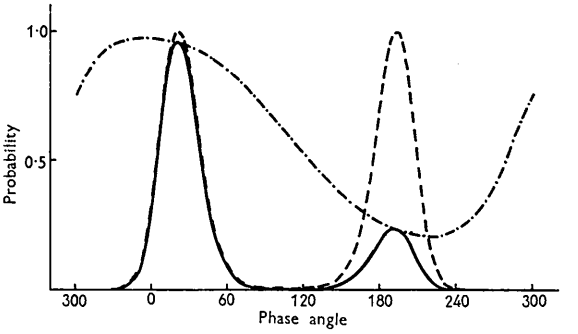

Fig. 7. As Fig. 6 but with a lower E value for the anomalous scattering probability curve.

It is clear therefore that, given appropriate weighting, anomalous scattering differences can be a useful supplement to multiple isomorphous replacement in phase determination.

References

Bijvoet, J. M. (1954). *Nature, Lond.* **173**, 888.

Blake, C. C. F., Fenn, R. H., North, A. C. T., Phillips, D. C. & Poljak, R. J. (1962). *Nature, Lond.* **196**, 1173.

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

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

Blow, D. M. & Rossmann, M. G. (1961). *Acta Cryst.* **14**, 1195.

Cullis, A. F., Muirhead, H., Perutz, M. F., Rossmann, M. G. & North, A. C. T. (1961). *Proc. Roy. Soc.* A, **265**, 15.

Cullis, A. F., Muirhead, H., Perutz, M. F., Rossmann, M. G. & North, A. C. T. (1962). *Proc. Roy. Soc.* A, **265**, 161.

Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C. & Shore, V. C. (1960). *Nature, Lond.* **181**, 662.

Kraut, J., Sieker, L. C., High, D. F. & Freer, S. T. (1962). *Proc. Nat. Acad. Sci., Wash.* **48**, 1417.