Acta Cryst. (1976). A32, 473

A New Synthesis for the Refinement of Heavy-Atom Parameters in Protein Crystallography

BY GREGORY A. PETSKO*

Laboratory of Molecular Biophysics, Department of Zoology, South Parks Road, Oxford, England

(*Received* 23 *January* 1975; *accepted* 14 *October* 1975)

Both of the conventional 'double-difference' or 'residual' syntheses used in protein crystallography to detect errors in refined heavy-atom parameters are approximations to the desired vector difference synthesis between the observed and calculated heavy-atom structure factors. By using the measured anomalous differences between Friedel-related pairs of reflexions from the parent-plus-heavy-atom crystal, it is possible to calculate the coefficients for the vector difference synthesis. Tests indicate that this synthesis can provide a better signal-to-noise ratio than the conventional syntheses. In the Appendix it is shown that $\bar{F}_{PH}$, parent-plus-heavy-atom structure amplitude in the absence of anomalous scattering, is properly calculated as the mean of the structure amplitudes of the Friedel-related reflexions, not the root mean square.

Notation

Structure factors are indicated as vectors $\mathbf{F}_P$, or as a magnitude and phase, $F_P \exp i\alpha_P$. The structure amplitude is denoted by F_P, or by the absolute value of the vector, $|F_P \exp i\alpha_P|$. Subscript P denotes the parent crystal; PH the parent-plus-heavy-atom crystal, and H the contribution from the heavy atom alone. Subscript c refers to calculated quantities; all others are taken as observed.

W is a weighting factor, usually $=1$. k is the ratio of the real to the imaginary part of the heavy-atom scattering factor. α, β, γ and μ are phase angles; δ is the imaginary part of the heavy-atom contribution.

F^+_{PH} and F^-_{PH} refer to the structure amplitudes from the Friedel-related reflexions $F_{PH}(hkl)$ and $F_{PH}(\bar{h}\bar{k}\bar{l})$, respectively.

Introduction

The solution of protein crystal structures by the method of isomorphous replacement (Green, Ingram & Perutz, 1954) requires accurate determination of the position, occupancy, and thermal parameters of the added heavy atoms. Estimates of these parameters are usually arrived at by difference Patterson synthesis or direct methods, and the trial values refined by least-squares. Two types of 'residual' or 'double-difference' Fourier syntheses are used to reveal errors in the refined heavy-atom parameters; the form of the synthesis depends on the type of refinement. When heavy-atom parameter refinement has been carried out by minimizing

$$\sum_{hkl} W(F_H - F_{Hc})^2, \tag{1}$$

where F_H is calculated for centrosymmetric reflexions by

$$F_H = |F_{PH} \pm F_P| \tag{2}$$

(Green *et al.*, 1954) and for non-centrosymmetric reflexions by

$$F_H = (F_P^2 + F_{PH}^2 - 2F_P F_{PH}\{1 - [k(F^+_{PH} - F^-_{PH})/2F_P]^2\}^{1/2})^{1/2} \tag{3}$$

(Dodson & Vijayan, 1971), then a 'residual' synthesis of the form

$$R\Delta F \exp i\alpha_{Hc}\,, \tag{4}$$

where

$$R\Delta F = (F_H - F_{Hc}) \tag{5}$$

and α_{Hc} is the phase angle for the calculated heavy-atom structure, will often reveal errors in the refined heavy-atom parameters (Adams, 1968; Evans, 1973; Petsko, 1973). However, protein crystallographers most often refine the heavy-atom parameters of isomorphous derivatives by the process commonly known as 'phase refinement' (Dickerson, Kendrew & Strandberg, 1961). This method involves the minimization of

$$\sum_{hkl} W(F_{PH} - F_{PHc})^2, \tag{6}$$

where F_{PH} is usually calculated from measurements of F^+_{PH} and F^-_{PH} (Appendix I) and F_{PHc} is given by

$$F_{PHc} = |F_P \exp i\alpha_P + F_{Hc} \exp i\alpha_{Hc}|\,. \tag{7}$$

The corresponding 'double difference' synthesis (Blake *et al.*, 1963) is calculated with coefficients

$$m\Delta\Delta F \exp i\alpha_{PHc} \tag{8}$$

where m is the 'figure of merit' for the corresponding native phase angle α_P, α_{PHc} is the phase angle calculated for the protein-plus-heavy-atom structure, and $\Delta\Delta F$ is given by

$$\Delta\Delta F = (F_{PH} - F_{PHc})\,. \tag{9}$$

This synthesis will reveal errors in the refined heavy-atom parameters; the effect of specific errors on features in it has been discussed by Bloomer (1972).

* Present address: Department of Biochemistry, Wayne State University School of Medicine, 540 E. Canfield, Detroit, Michigan 48201, U.S.A.

The purpose of this paper is to point out that both of these syntheses are approximations to what is really desired, and to suggest a more accurate synthesis.

The conventional double difference syntheses

Fig. 1 shows the relations among the various quantities involved in protein phase-angle determination for a reflexion with a large difference between the observed and calculated heavy-atom structure factor. To detect errors in heavy-atom parameters, one would wish to use a synthesis whose coefficients represent the vector difference between the observed and calculated F_H's. The quantity desired is

$$(\mathbf{F}_H - \mathbf{F}_{Hc}) , \tag{10}$$

which is denoted in Fig. 1 by

$$V\Delta F \exp i\beta , \tag{11}$$

where β is actual phase angle of $(\mathbf{F}_H - \mathbf{F}_{Hc})$ and

$$V\Delta F = |F_H \exp i\alpha_H - F_{Hc} \exp i\alpha_{Hc}| . \tag{12}$$

Neither of the conventional double difference syntheses provides these coefficients. Both are 'residual syntheses' in the sense that they use as the magnitudes of their coefficients the residuals of the corresponding least-squares refinement. In each case the phase angles of the coefficients are those of the calculated F's in the refinement process.

Least-squares refinement in three dimensions by equation (1) depends on the measured Bijvoet differences between F_{PH}^+ and F_{PH}^- to provide an accurate estimate of F_H for non-centrosymmetric reflexions [by equation (3)]. The $R\Delta F$ synthesis used after this type of refinement is related to the conventional difference Fourier synthesis used in the refinement of small-molecule crystal structures (Cochran, 1951). This type of synthesis has been analysed by Dodson & Vijayan (1971) and Henderson & Moffat (1971); they note that it is only an approximation to the vector difference between F observed and F calculated. From Fig. 1 it can be seen that the $R\Delta F$ synthesis represents the vector component of $(\mathbf{F}_H - \mathbf{F}_{Hc})$ in the direction of α_{Hc}, but includes no component perpendicular to this direction. This is the source of the intrinsic error in this synthesis.

The same considerations apply to the $\Delta\Delta F$ synthesis used after phase refinement. As Fig. 1 shows, the $\Delta\Delta F$ synthesis represents the difference between $\mathbf{F}_{PH}$ and $\mathbf{F}_{PHc}$ in the direction of α_{PHc} but includes no contribution in the perpendicular direction. It therefore suffers from the same type of intrinsic error as the $R\Delta F$ synthesis. Although the $\Delta\Delta F$ synthesis has coefficients equal to the so-called 'lack-of-closure' vector in phase refinement (Blow & Matthews, 1973), the difference that is calculated actually represents a component of $(\mathbf{F}_H - \mathbf{F}_{Hc})$, as can be seen in Fig. 1.

Thus both the $R\Delta F$ and $\Delta\Delta F$ syntheses are approximations to the desired quantity. Both contain intrinsic error. It should be noted that the two syntheses do not provide identical coefficients (Fig. 1).

The new vector difference synthesis

The intrinsic error in the $R\Delta F$ and $\Delta\Delta F$ syntheses can be avoided by calculation of the true vector difference synthesis, equation (11). If the Bijvoet differences $(F_{PH}^+ - F_{PH}^-)$ have been measured for the reflexions from the protein-plus-heavy-atom crystal, it is possible to calculate the coefficients of the $V\Delta F$ synthesis. This is because the Bijvoet differences can be used to obtain information about the phase angle β of $(\mathbf{F}_H - \mathbf{F}_{Hc})$. Indeed, it is possible to derive an exact expression for β from the relationships illustrated in Fig. 1, but this is not essential (see below).

A number of workers have already noted that the Bijvoet differences enable an accurate estimate of F_H to be made (Kartha & Parthasarathy, 1965; Matthews, 1966; Singh & Ramaseshan, 1966; Dodson & Vijayan, 1971). F_H is calculable from the isomorphous and Bijvoet differences by equation (3), provided reflexions are rejected where $|\alpha_P - \alpha_{PH}|$ is likely to be $>90°$ (Dodson & Vijayan, 1971). k, the ratio of the real to the anomalous scattering of the heavy-atom group, may be determined empirically (Matthews, 1966). It is further possible to use the Bijvoet differences to derive an accurate estimate of α_H. Matthews (1966) has shown that this is given by

$$\alpha_H = \alpha_P + \mu - \pi , \tag{13}$$

where

$$\sin\mu = \frac{F_{PH}\sin(\alpha_P - \alpha_{PH})}{F_H} ; \quad \cos\mu = \frac{(F_P^2 + F_H^2 - F_{PH}^2)}{2F_P F_H} \tag{14}$$

and, if the heavy atoms are 'of the same type',

$$\sin(\alpha_P - \alpha_{PH}) \simeq k(F_{PH}^+ - F_{PH}^-)/2F_P . \tag{15}$$

Since F_H and α_H can be calculated, the calculation of the vector difference synthesis coefficients is straight-

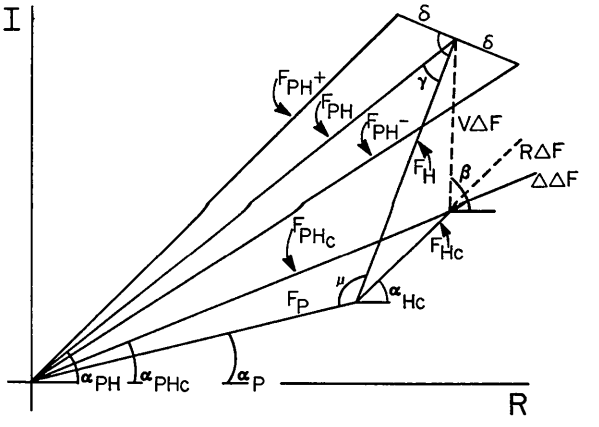

Fig. 1. Vectors on the complex plane showing the relations needed to derive the new vector difference synthesis. The (*hkl*) diagram has been reflected through the real axis onto the (*hkl*) diagram.

forward. Equations (3), (13), (14) and (15) are used to calculate $\mathbf{F}_H$, and $\mathbf{F}_{Hc}$ is subtracted from it:

$$V\Delta F \exp i\beta = (F_H \exp i\alpha_H - F_{Hc} \exp i\alpha_{Hc}). \quad (16)$$

Since the angle β depends on α_P [*via* α_H and equation (13)], it seems justifiable to weight each reflexion in the synthesis with the figure of merit of the corresponding protein phase. To avoid any bias or feedback into the synthesis it is desirable to omit the derivative being refined from the calculation of α_P.

Test

The value of this synthesis has been tested by the use of data taken from studies of triose phosphate isomerase at 6 Å resolution (Banner, Bloomer, Petsko, Phillips & Pogson, 1971). A computer program was written in Fortran to calculate the $V\Delta F$ coefficients from the observed isomorphous and Bijvoet differences, the protein phase angle, and the calculated heavy-atom parameters. Equation (16) was used with figure-of-merit weighting. In using equations (3) and (15) to estimate $\mathbf{F}_H$, empirical rather than theoretical values of k were used. Protein phase angles α_P were calculated from all derivatives except ethyl mercuric phosphate (EMP), which was chosen as the subject of the refinement. In the calculation of $\mathbf{F}_{Hc}$ three of the four EMP sites were included, two correctly and one shifted 2·5 Å from its true position. The program also calculated coefficients for the conventional $\Delta\Delta F$ synthesis [equation (8)] for comparison. After calculation of the two syntheses, the r.m.s. error was calculated as the root-mean-square density in featureless regions.

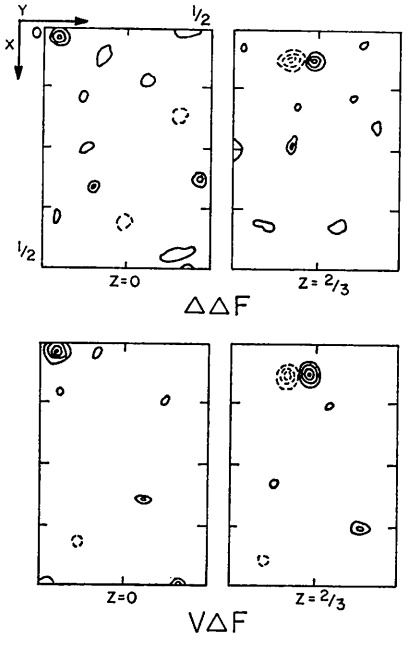

Fig. 2. Sections of the full three-dimensional double difference maps for the ethyl mercuric phosphate derivative of chicken triose phosphate isomerase. The left sections at $Z=0$ show the site omitted from the refinement. The right sections at $Z=\frac{2}{3}$ illustrate the appearance of a site incorrectly positioned. The top sections are the conventional double difference synthesis; the bottom sections are from the new vector difference synthesis. Contours at equal and arbitrary intervals; zero contour omitted, negative contours dashed.

The resulting map sections around the site omitted and the site incorrectly located are shown in Fig. 2. For both syntheses the correct positions of the sites are clearly indicated, but in the $V\Delta F$ map the peaks are higher and appear against a lower background. In both syntheses the areas around the correct sites are featureless. The r.m.s. errors in the two maps are:

$$\text{r.m.s.}\ (\sigma\Delta\Delta F) = 0{\cdot}076\ \text{e/Å}^3,$$
$$\text{r.m.s.}\ (\sigma V\Delta F) = 0{\cdot}052\ \text{e/Å}^3.$$

The heights of the peak omitted entirely are 10 and 5·7 times the r.m.s. errors for the $V\Delta F$ and $\Delta\Delta F$ maps respectively.

It is simple to modify any least-squares refinement program to calculate the $V\Delta F$ synthesis, no matter which refinement method is used. Despite the requirement in this synthesis for accurate measurement of Bijvoet differences it would seem to have some utility, as it avoids the intrinsic error present in the conventional residual syntheses.

This synthesis may be compared with that of Matthews (1966), from which equations (13) and (14) have been taken. Both syntheses endeavour to arrive at an accurate determination of the parameters of heavy-atom derivatives and both use information available in measurements of Bijvoet differences and preliminary protein phase angles. Obviously, the $V\Delta F$ synthesis cannot be used until some preliminary information about the derivative in question is available to allow initiation of refinement. Such information may logically be obtained by Matthews's synthesis; however, since it makes no use of F_{Hc} this latter synthesis is not as easy to use in revealing errors in *refined* heavy-atom parameters. Also, as the synthesis of Matthews (1966) is an F_{obs} synthesis, it is more sensitive to series-termination errors than a difference synthesis such as the $V\Delta F$.

APPENDIX

The calculation of F_{PH} from F^+_{PH} and F^-_{PH}

Ordinarily, in protein crystallography, the reflexions from each derivative are measured in 'Friedel-related' pairs, in order to make use of anomalous scattering effects in heavy-atom location and phase determination. However, these methods also require F_{PH}, the derivative structure amplitude in the absence of anomalous scattering. Although F_{PH} can of course be calculated exactly from F^+_{PH} and F^-_{PH} once the protein phases are known, in their absence it is necessary to approximate it by $\bar{F}_{PH}$, which has customarily been calculated by one of two different expressions:

$$\bar{F}^2_{PH} = [(F^+_{PH} + F^-_{PH})/2]^2 \quad (A1)$$
$$\bar{F}^2_{PH} = [(F^+_{PH})^2 + (F^-_{PH})^2]/2\,. \quad (A2)$$

Equation (*A*1) has been used extensively, but not exclusively; for example, in their derivation for F_H, Singh & Ramaseshan (1966) require $\bar{F}^2_{PH}$ as estimated by equation (*A*2). There appears to have been no previous analysis of which approximation leads to the least error. In principle (*A*2) might be preferred as it uses the squares of the structure amplitudes, which are more directly related to the experimentally observed intensities, but the question can be resolved by reference to Fig. 1 and a little algebra. If $\mathbf{F}_{PH}$, the resultant vector of $\mathbf{F}_P$ and $\mathbf{F}_H$ in the absence of anomalous scattering, makes an angle γ with $\mathbf{F}_H$, then (*cf.* Phillips, 1966):

$$\left.\begin{aligned}(F^+_{PH})^2 &= F^2_{PH}+\delta^2+2\delta F_{PH}\sin\gamma \\ \text{and}\qquad (F^-_{PH})^2 &= F^2_{PH}+\delta^2-2\delta F_{PH}\sin\gamma\end{aligned}\right\} \qquad (A3)$$

where δ is the imaginary part of the heavy-atom contribution. It is necessary to assume that δ makes an angle of 90° to F_H; *i.e.*, all the anomalous scatterers are the 'same type'. Equations (*A*3) are the only *exact* equations relating F_{PH} to F^+_{PH} and F^-_{PH}. Adding them together gives

$$\frac{(F^+_{PH})^2+(F^-_{PH})^2}{2} = F^2_{PH}+\delta^2. \qquad (A4)$$

Equation (*A*4) is also *exact*. There is no error if the measurements are accurate. For the moment it will be assumed that they are.

Comparing (*A*4) with (*A*2) shows that using (2) leads to an error of δ^2. By choosing $\bar{F}_{PH}$ as the r.m.s. of F^+_{PH} and F^-_{PH} one in fact overestimates it by δ^2. Equation (*A*1) is in error from (*A*4) by

$$\frac{-[(F^+_{PH})^2+(F^-_{PH})^2]}{4}+\frac{(F^+_{PH})\,(F^-_{PH})}{2}+\delta^2$$

which can be reduced to

$$-\tfrac{1}{4}[(F^+_{PH})-(F^-_{PH})]^2+\delta^2.$$

This is clearly always $\leq\delta^2$. It is equal to δ^2 when the anomalous difference is 0, in which case (*A*1) and (*A*2) are equivalent anyway. Whenever the anomalous difference is non-zero, the mean provides a better estimate of $\bar{F}_{PH}$ than the root mean square. The error involved in using either expression increases as the size of the anomalous difference increases.

The best expression to use when there is no error is (*A*1); if there is only one measurement each for F^+_{PH} and F^-_{PH} and if each measurement has associated with it an estimated standard deviation, the most precise estimate of $\bar{F}_{PH}$ will be given by the weighted mean. However, the most precise value is not statistically the value free of the anomalous scattering effects.

If the heavy atoms are all of the same type, statistical considerations require that one give equal weights to (F^+_{PH}) and (F^-_{PH}) in the determination of the weighted mean. The individual estimated standard deviations may be used to calculate the standard deviation of $\bar{F}_{PH}$. This quantity is useful in refinement and phase determination and can also be used to determine whether anomalous differences are statistically significant.

If there is more than one measurement each for F^+_{PH} and F^-_{PH}, circumstances of data collection influence how (or if) they should be combined before refinement. Tukey (1974) has given a lucid discussion of the considerations involved.

I am grateful to Professor David C. Phillips, F.R.S., and Professor John W. Tukey for advice and stimulating discussions. I thank the Rhodes Trust and the Danforth Foundation for fellowships. The referee made some very useful comments.

References

Adams, M. J. (1968). D. Phil. Thesis, Univ. of Oxford.

Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C. & Pogson, C. I. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 151–155.

Blake, C. C. F., Fenn, R. H., Koenig, D. F., North, A. C. T., Phillips, D. C. & Poljak, R. J. (1963). *Acta Cryst.* **16**, A77.

Bloomer, A. C. (1972). *Crystallographic Studies of Oligomeric Enzymes*. Thesis, Univ. of Oxford.

Blow, D. M. & Matthews, B. W. (1973). *Acta Cryst.* A**29**, 56–62.

Cochran, W. (1951). *Acta Cryst.* **4**, 408–411.

Dickerson, R. E., Kendrew, J. C. & Strandberg, B. E. (1961). In *Computing Methods and the Phase Problem in X-ray Crystal Analysis*, pp. 236–251. Oxford: Pergamon Press.

Dodson, E. & Vijayan, M. (1971). *Acta Cryst.* B**27**, 2402–2411.

Evans, P. R. (1973). *X-ray Structural Studies of Enzymes*. Thesis, Univ. of Oxford.

Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). *Proc. Roy. Soc.* A**225**, 287–307.

Henderson, R. & Moffat, J. K. (1971). *Acta Cryst.* B**27**, 1414–1420.

Kartha, G. & Parthasarathy, R. (1965). *Acta Cryst.* **18**, 745–749.

Matthews, B. W. (1966). *Acta Cryst.* **20**, 230–239.

Petsko, G. A. (1973). *Crystallographic Studies of Triose Phosphate Isomerase*. Thesis, Univ. of Oxford.

Phillips, D. C. (1966). *Advanc. Struct. Res. Diffr. Meth.* **2**, 75–140.

Singh, A. K. & Ramaseshan, S. (1966). *Acta Cryst.* **21**, 279.

Tukey, J. W. (1974). In *Critical Evaluation of Chemical and Physical Structural Information*, edited by D. R. Lide, Jr. & M. A. Paul. pp. 3–41. Washington D. C.: National Academy of Sciences.