Acta Cryst. (1968). B**24**, 997

A Least-Squares Refinement Method for Isomorphous Replacement*

By Richard E. Dickerson and Jon E. Weinzierl
Gates and Crellin Laboratories of Chemistry, California Institute of Technology, Pasadena, California, U.S.A.

and Rex A. Palmer
Department of Crystallography, Birkbeck College, London, England

(*Received* 3 *January* 1967 *and in revised form* 3 *October* 1967)

A least-squares method is described for refining the heavy atom parameters of an isomorphous replacement phase analysis, based upon minimizing the weighted sum of squares of the lack-of-closure errors for each derivative in turn, the phase angles having been determined earlier by the use of all derivatives at once. Conclusions based on trials with cytochrome C are described. A comparison is made between the mean figure of merit and the Kraut and least-squares R factors (R_K and R_L) as working criteria of phase analysis and refinement. It is observed that heavy atoms will refine in position correctly if their initial displacement is as much as half the nominal resolution. The behavior of substitution number, A, and radial falloff factor, B, under varying conditions is studied.

Introduction

After rough heavy atom parameters have been found in an isomorphous replacement phase analysis, they must be refined prior to the final phase determination. Various schemes have been used previously with proteins, including a correlation function least-squares method (Rossmann, 1960), and a trial-and-error least-squares method coupled with a method of steepest descents, applicable only to centric data (Hart, 1961). The Hart method has worked quite well, and has continued to be used since its first application to myoglobin in 1959. Its handicap is its inability to handle acentric reflections and hence to be applicable to proteins in three dimensions.

The purpose of this paper is to describe a method of least-squares refinement, to present some experiences with the method, and to compare several criteria of refinement and of phase determination. The method was first used to refine the relative y coordinates of the heavy atoms in sperm whale myoglobin (Dickerson, Kendrew & Strandberg, 1961) after the other parameters had been refined by the Hart method. It has since been used independently by Kraut, Sieker, High & Freer (1962) with chymotrypsinogen, Muirhead (1966) for haemoglobin, and Lipscomb, Coppola, Hartsuck, Ludwig, Muirhead, Searl & Steitz (1966) for carboxypeptidase.

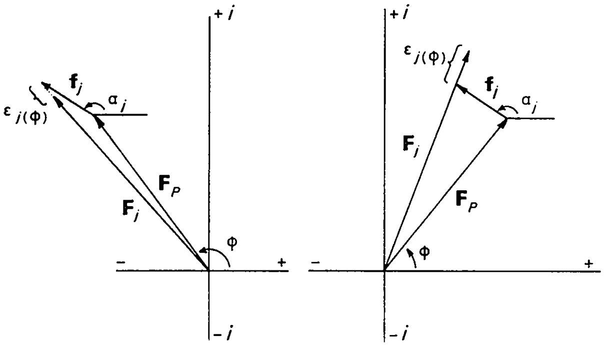

Fig. 1. Phase triangles on the complex plane, solutions of the equation: $\mathbf{F}_j=\mathbf{F}_P+\mathbf{f}_j$. Illustrating lack-of-closure errors, ε_j, from incorrect choices of phases, φ.

Principles of refinement

The quantities known at the start of an isomorphous phase analysis are the magnitudes of the structure factors of the parent protein, F_P, and the jth heavy atom derivative, F_j, and the magnitude, f_j, and phase, α_j, of the total scattering contribution of the heavy atom group in the jth derivative. In terms of the vector triangle formed by equation: $\mathbf{F}_j=\mathbf{F}_P+\mathbf{f}_j$ on the complex plane, the quantities known are the lengths of all three sides and the orientation of one, $\mathbf{f}_j$. The quantity wanted is φ, the orientation of side $\mathbf{F}_P$.

For each derivative there will be two choices of φ for which the phase triangle closes and the vector equation is satisfied. For all other φ's, there will be a lack-of-closure error $\varepsilon_{j(\varphi)}$. (See Fig. 1.) This lack-of-closure error is defined as:

$$\varepsilon_{j(\varphi)}\equiv\left|\,|F_j|-|D_{j(\varphi)}|\,\right| \tag{1}$$

where $D_{j(\varphi)}$is the length of the third side of the phase triangle defined by F_P, φ, f_j and α_j, and is given by:

$$D^2_{j(\varphi)}=F^2_P+f^2_j+2|F_P|\,|f_j|\cos(\varphi-\alpha_j)\,. \tag{2}$$

In a sense, D_j bears the same relationship to the observed heavy atom derivative structure factor, F_j, as the calculated structure factor does to the observed structure factor in a conventional small-molecule analysis.

The refinement program is based upon the theory of phase analysis derived by Blow & Crick (1959). In

---

* Contribution No. 3465 from the Gates and Crellin Laboratories of Chemistry.

their treatment, the unnormalized probability of a given phase, φ, being correct for a particular reflection is given by:

$$P_{(\varphi)} = \exp - \sum_j \left(\frac{\mathscr{E}_j^2(\varphi)}{2E_j^2} \right), \tag{3}$$

where E_j is the root mean square error in derivative j, arising either from data errors or from misinterpretation of the heavy atom sites [equations (15) and (16)]. If, for each reflection, φ_M is the phase angle for which the probability is at a maximum, then the most probable Fourier map will be that which uses the Fourier coefficients:

$$\{|F_P| \exp i\varphi_M\}. \tag{4}$$

The second part of the Blow–Crick theory demonstrates that a somewhat different set of coefficients will produce the map with the least overall root mean square error in electron density. This is the map which uses as phase vectors, not the unit vector in the direction of the most probable phase angle φ_M, but rather the vector to the center of gravity of the probability function $P_{(\varphi)}$ plotted around a unit radius circle. If the polar coordinates of the center of gravity of this probability distribution on the complex plane are (m,φ_B), then the proper Fourier coefficients for this least error map are:

$$\{m|F_P| \exp i\varphi_B\}. \tag{5}$$

The radial component, m, acts to weight down the contribution of a poorly determined reflection. It has been called the 'figure of merit' and used as a criterion of the quality of a phase determination. But it is important to realize that it is not a measure of the absolute probability of the phase determination – only of its sharpness or unambiguity. The value of $P_{(\varphi)}$ itself must be taken into account as well. The individual figure of merit measures the precision and not the accuracy of a phase determination.*

The refinement program to be discussed in this paper first determines phases by the Blow–Crick method using the initial unrefined heavy atom parameters, and then carries out one cycle of full-matrix least-squares refinement of the heavy atom parameters of each derivative with phase angles held fixed. It repeats this alternation of phase determination and refinement as often as desired and then finishes with a final phase cycle and a punched phase deck suitable for Fourier program input. The heavy atom parameters refined are x_k, y_k, z_k, A_k and B_k in the expression:

$$\mathbf{f}_j = \sum_k A_k \exp(-B_k S^2) \exp 2\pi i(hx_k + ky_k + lz_k) \tag{6}$$

or its appropriate space group reduction. (The summation, k, is over all the heavy atom sites in the jth derivative, and $S = 2 \sin \theta$.) The coefficients K_j and β_j of the derivative scale factor, $K_j' = K_j \exp(\beta_j S^2)$, are refined as well, the factor K_j' being defined as the number by which the derivative structure factors F_j must be *divided* in order to put them on the same relative scale as F_P. At present the program to be described works only for space groups $P1$ and $P4_1$, but the adaptation to other space groups is straightforward.

In phase determination the following quantity is minimized to find φ_M:

$$\mathscr{E}_h = \sum_j w_j(|F_{h,j}| - |D_{h,j(\varphi)}|)^2 \tag{7}$$

for each reflection, h, where $w_j = 1/2E_j^2$. (Throughout this paper, the symbol h as a subscript by itself will denote the triplet of Miller indices (h,k,l). The subscript j will denote a particular heavy atom derivative, and subscript k will denote a particular heavy atom site in the jth derivative. Subscripts m and n, when used with ψ, will denote individual members in a sequential list of all the parameters which are refined by least squares.) In the refinement cycle, an analogous sum over all reflections is minimized:

$$\mathscr{E}_j = \sum_h w_h(|F_{h,j}| - |D_{h,j(\varphi)}|)^2 \tag{8}$$

for each derivative, j. The desired shifts in the parameters ψ_m (where ψ_m can be atom coordinates, effective atomic number A_k, temperature factor B_k or scale factor components K_j,β_j) are found by solving the set of normal equations:

$$\sum_m a_{mn} \Delta\psi_m = b_n$$

or:

$$[a_{mn}][\Delta\psi_m] = [b_n], \tag{9}$$

where the subscripts m and n denote the parameters being refined. The matrix coefficients a_{mn} and b_n are

---

* If, for example, a set of phase circles fails to intersect at all, but instead has a point of closest approach, or if an otherwise well-determined phase is spoiled by an aberrant derivative, then there can result a cleanly determined but improbable phase, with a high figure of merit but a low maximum probability, $P_{(\varphi_M)}$. Conversely, if the heavy atom contributions are small and if the derivative circles lie very close to the parent circle at all values of φ, then one can have an ambiguous but intrinsically likely phase determination, with low m and high $P_{(\varphi_M)}$. The greater the lengths of the heavy atom vectors relative to the phase circle radii, the sharper will be the intersection points of the circles and the higher the figure of merit. The greater the errors in phase circle radii on the other hand (or errors in intensity data), the greater the measure by which circles which would otherwise approach tangentially will fail to do so, and the lower will be the $P_{(\varphi)}$ at all values of φ. Since the heavy atom contribution in proteins is usually small on the average relative to $\langle F_P \rangle$, most phase circle sets approach the tangential rather than the perpendicular intersection extreme. Under these conditions, therefore, the figure of merit tends to reflect the quality of the heavy atom substitutions, and the intrinsic probability $P_{(\varphi)}$ (or its negative logarithm, the Relative Error), the quality of the data.

given by:

$$\left\{\begin{aligned} a_{mn} &= \sum_h w_h \left(\frac{\partial D_{h,j}}{\partial \psi_m}\right)\left(\frac{\partial D_{h,j}}{\partial \psi_n}\right) \\ b_n &= \sum_h w_h (F_{h,j} - D_{h,j(\varphi)})\left(\frac{\partial D_{h,j}}{\partial \psi_n}\right) \end{aligned}\right\}. \qquad (10)$$

The phase angles are the products of the previous step of phase determination, and can be either φ_M or φ_B.

No provision was made in this particular program for the incorporation of anomalous scattering data into the analysis, although this could certainly be done using the phase probability method of North (North, 1965; Matthews, 1966). Since the program was designed for protein structure analysis, where centric reflections occur only in a few zones, it was decided to treat all reflections as acentric rather than to use a tanh formula (Blow & Crick, 1959) for centric ones.

Operation of the program

Phase analysis

The program as written can accept up to six sites in the asymmetric unit in each of a maximum of eight heavy atom derivatives. There is the choice of refining all parameters, of suppressing temperature factor (B_k) refinement, or of refining positional coordinates only. Fractional shifts can be applied. The options exist of printing out individual probabilities, $P_{(\varphi)}$, or punching probability profiles across the 72 columns of an IBM card. Either φ_M or φ_B may be used, although it has been found that φ_B is preferable in general and of course is mandatory for centric reflections, where the center of gravity of the two phase solutions symmetric about the real number axis must be used.

The phase determining subprogram calculates $P_{(\phi)}$ at 5° intervals around the phase circle, and finds the phase angle, φ_M, for which the probability is at a maximum. It calculates m and φ_B from:

$$m \cos \varphi_B = \frac{\sum_\varphi P_{(\varphi)} \cos \varphi}{\sum_\varphi P_{(\phi)}}$$

$$m \sin \varphi_B = \frac{\sum_\varphi P_{(\varphi)} \sin \varphi}{\sum_\varphi P_{(\varphi)}}. \qquad (11)$$

It then finds the negative of the natural logarithm of the probability, $P_{(\varphi)}$, for φ_M and φ_B, which will be called *EMAX* and the 'residual error', abbreviated 'R.E.':

$$EMAX \equiv \sum_j \frac{\varepsilon^2_{j(\varphi_M)}}{2E_j^2}$$

$$\text{R.E.} \equiv \sum_j \frac{\varepsilon^2_{j(\varphi_B)}}{2E_j^2}. \qquad (12)$$

As it proceeds, the phase subprogram contributes to the buildup of the mean figure of merit, $\langle m\rangle$ and to the mean residual error, M.R.E. It also contributes to the buildup of the Kraut R factor, R_K (Kraut *et al.*, 1962), defined by:

$$R_{K_j} \equiv \frac{\sum_h ||F_{h,j}|_{\text{obs}} - |F_{h,j}|_{\text{calc}}|}{\sum_h |F_{h,j}|_{\text{obs}}} \times 100\%$$

$$= \frac{\sum_h \varepsilon_{h,j(\varphi_B)}}{\sum_h |F_{h,j}|} \times 100\% \qquad (13)$$

and to a least-squares R factor, R_L, defined by:

$$R_{L_j} \equiv \left\{\frac{\sum_h w_h \varepsilon^2_{hj(\varphi_B)}}{\sum_h w_h |F_{h,j}|^2}\right\}^{1/2} \times 100\% \qquad (14)$$

for each derivative. Since the numerator of R_L^2 is the sum of squares of the residuals and is the quantity minimized by the least-squares subprogram, R_L is the true criterion of refinement. The taking of the square root in the definition of R_L is perhaps unnecessary; yet it makes R_L of the first power in ε, and brings the R_L values into a range comparable with R_K and conventional R factors.* The program calculates the mean discrepancy between φ_M and φ_B, and the mean change in φ_B between the present phase cycle and the previous one, $\langle \Delta\varphi_B \rangle$.

The r.m.s. estimate of error, E_j, for each derivative is initially fed into the phase program as a constant. It is usually found initially from:

$$E_j^2 = \langle (|\Delta F_j| - |f_j|)^2 \rangle_h \qquad (15)$$

where $\Delta F_j = |F_j| - |F_P|$, and the average is over centric reflections only. E_j^2 as defined above is formally equivalent to $\langle \varepsilon^2_{j(\varphi_B)} \rangle$ for centric reflections if crossover terms (in which F_P and F_j have opposite signs) are neglected.

---

* The Kraut R_K factor is closely related to another phase criterion, the centric R_C factor of Cullis, Muirhead, Perutz, Rossmann & North (1961). For centric reflections only:

$$\varepsilon_{h,j} = |\mathbf{F}_{h,j} - \mathbf{D}_{h,j}| = |\mathbf{F}_{h,j} - (\mathbf{F}_{h,p} + \mathbf{f}_{h,j})| = |\Delta\mathbf{F}_{h,j} - \mathbf{f}_{h,j}|$$
$$= [|\Delta F_j| - |f_j|_h]$$

and:

$$R_C = \frac{\sum_h \varepsilon_{h,j}}{\sum_h |\Delta F_{h,j}|} = R_K \cdot \frac{\sum_h |F_{h,j}|}{\sum_h |\Delta F_{h,j}|}.$$

The ratio of mean heavy atom change to mean structure factor is useful in any event, and if two of the three quantities R_K, R_C and $\sum_h |\Delta F_j| / \sum_h |F_j|$ are quoted for *centric* reflections then the third can be calculated. For good derivatives in structures which have been published to date, $\sum_h |\Delta F_j| / \sum_h |F_j|$ has tended to be around 0·25, R_C around 40% and R_K around 10%.

As the phase analysis proceeds, new weighting factors are built up as:

$$E_j^2 = \langle \varepsilon_{j(\varphi_B)}^2 \rangle_h \,, \tag{16}$$

again with the average taken *only over centric reflections*. These new E_j's are better than the originals because of proper treatment of crossover terms. As will be seen below, the mean figure of merit is rather sensitive to the value of E_j used. An underestimation of the r.m.s. error in the data will lead to a spuriously high $\langle m \rangle$.

In summary, the refinement criteria computed by the phase subprogram are: $\langle m \rangle$, M.R.E., $\langle \Delta\varphi_B \rangle$, and R_K, R_L and E_j for each derivative. Their use is described below.

Refinement

Each derivative, j, is refined independently for one cycle, minimizing the $\mathscr{E}_j$ of equation (8). The reflection weighting factor, w_h, comes from the just-completed phase cycle. In the earlier relative y application to myoglobin, an empirical weighting factor equal to the square of the figure of merit of the reflection was used. In the current program, the weight is taken as being the inverse of the r.m.s. lack-of-closure error over all derivatives for the reflection in question:

$$w_h = 1/E_h \,, \text{ where } E_h^2 = \langle \varepsilon_{j(\varphi_B)}^2 \rangle_j \,. \tag{17}$$

The least-squares matrices $[a_{mn}]$ and $[b_n]$ are built up in the normal way. A precise definition of D_j, including the scale factor K'_j is:

$$D_j = K'_j \{ F_P^2 + f_j^2 + 2|F_P|\,|f_j| \cos(\varphi - \alpha_j) \}^{1/2}. \tag{18}$$

The partial derivatives of D_j with respect to the scale factor constants are trivial. The derivatives with respect to all the other parameters of refinement can conveniently be written:

$$\frac{\partial D_j}{\partial \psi_n} = \frac{K_j'^2}{D_j}$$

$$\times \left\{ (F_P \cos\varphi + a_j) \frac{\partial a_j}{\partial \psi_n} + (F_P \sin\varphi + b_j) \frac{\partial b_j}{\partial \psi_n} \right\}, \tag{19}$$

where a_j and b_j are the real and imaginary parts of the heavy atom group structure factor, with their appropriate space group reductions. If a given parameter is not to be refined, then its derivatives are not calculated. The program then checks to make sure that the row and column in $[a_{mn}]$ corresponding to this parameter are all zeros and puts a value of 1·0 in the diagonal box common to the row and column. The program solves the matrix equations as given in equation (9), and applies the shifts, fully or partially. It repeats the process for the successive derivatives, then returns control to the phase determining subprogram. On the last cycle, the phase subprogram exits to the final phase output routine rather than to the refinement subprogram.

Experimental

Refinement tests were carried out in two dimensions (*hk*0 zone, space group $P4_1$) and in three dimensions, for Pt ($PtCl_4^{2-}$) and Hg (mersalyl) derivatives of horse heart cytochrome C, at 4 Å resolution. Details have been published by Dickerson, Kopka, Varnum & Weinzierl (1967*a*) and Dickerson, Kopka, Borders, Varnum, Weinzierl & Margoliash (1967*b*). Out of these trials have come some general observations which may be useful to those trying such a refinement method.

Single-derivative refinement in centrosymmetric projection was never very satisfactory. It is as if the heavy atom, having been solely responsible for the sign set against which refinement was to take place, then refused to be influenced by it. In an analogous manner, refinement in three dimensions with the phase-determining minimum of two derivatives is inconclusive. The phases, having been determined so as to be most nearly compatible with the starting heavy atom parameters, then do not exert a strong corrective pressure on these parameters. For this reason most of the conclusions of this paper have been based on centric data refinement.

Best results were obtained with double shifts in coordinates, half to full shifts in A, half shifts in K and no refinement of B. Derivative scale factors, when set to 1·00, refined rapidly to values near the Kraut scale factor of:

$$K = \frac{\sum\limits_h F_P F_j}{\sum\limits_h F_P^2} \,. \tag{20}$$

Five main points emerged from the refinement trials, and can be illustrated by *hk*0 refinement runs using Pt and Hg:

1. The refinement program will pull heavy atoms back to their proper positions from displacements as much as $d_{\min}/2$ Å away, where $d_{\min}$ is the resolution of the data.

Fig. 2 shows the shifts in Pt and Hg positions during trials using these two derivatives alone and various initial displacements. Atoms displaced by 2 Å from their proper positions refined back without difficulty at 4 Å resolution (*S*26), whereas with 3 Å displacements they failed to do so (*S*27). The platinum atom oscillated in place while the mercury atom moved in the wrong direction and then began a long looping spiral which might eventually have converged. This failure to refine was reflected in lower mean figure of merit, higher R_K and R_L and substantially larger r.m.s. errors, E_j. During refinement the substitution numbers A_j fell steadily as if the program, being unable to refine the atoms properly, was bent upon wiping them out. This behavior is quite typical with incorrect atoms.

With 2·5 Å displacement, Pt refined but Hg did not (*S*28). Only when the Hg atom was moved back to 2 Å from its correct position did both atoms refine properly (*S*29). It appears that with 4 Å data there is something like a crest in the refinement plane at 2 Å displacement, beyond which the atoms slip over into false minima. This 2 Å limit is just the point at which the atoms move into the wrong half-cycle of a Fourier ripple from a reflection at the edge of the data sphere, and is the point at which derivatives of equation (19) for these reflections begin to have the wrong sign. With a few wrong-sign terms in the summations of equation (10), the heavy atom rides around the contour without falling into the proper minimum (*S*28), and with a considerable number of wrong signs, the atom wanders off in the wrong direction (*S*27).

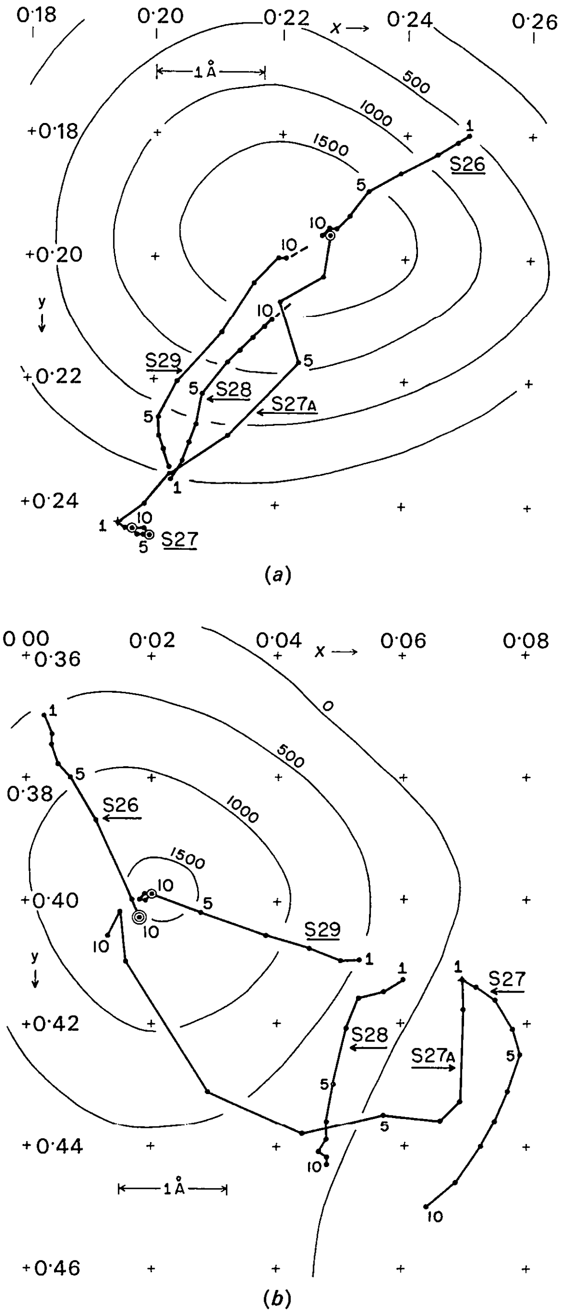

Fig. 2. Refinement of x and y coordinates of (*a*) Pt and (*b*) Hg for runs *S*26, *S*27, *S*27*A*, *S*28 and *S*29. Numbers 1, 5 and 10 alongside trajectories are cycle numbers. Crosses mark intersections of grid lines in 50ths of a cell edge. Standard 1 Å length is marked. Contour lines are from final ΔF difference Fourier maps, and have nothing to do with the refinement other than illustrating the goal.

If this half-cycle displacement argument is valid, then it should be possible to make *S*27 refine by cutting the data off at 6 Å. (Weighting down the outer reflections in early stages of least-squares refinement is, of course, nothing new in standard small-molecule refinement.) The results are shown as run *S*27*A*. Pt refined at once, and Hg followed after an initial swerve along the contour.

2. *A* and *B* will not refine well together at 4 Å resolution, but the Wilson-plot values of *B* are sufficient for low resolution phasing.

A and *B* for each derivative [equation (6)] were obtained initially from Wilson plots like Fig. 3, using initial values of $A = 1{\cdot}0$ and $B = 0{\cdot}0$. (The first power of S was used rather than S^2 in the exponent of equation (6) for reasons described by Dickerson *et al.* (1967*b*), but this does not affect the conclusions about *A*,*B* behavior.) If *B* was varied within a considerable range, *A* was found to refine to a value which best fitted a straight line through the center of gravity of the experimental points on the Wilson plot. In Fig. 3, line *S*26 represents the refined Pt (*A*,*B*) values of (8·00, 7·46), starting with Wilson-plot values of (8·15, 7·46) and keeping *B* fixed. In run *S*26*A*, *B* was increased to 12·00 and held fixed while *A* was set initially to 5·00, represented by the lowest dashed line, which does not come anywhere near the plotted points. Ten cycles of refinement raised *A* to 11·02 and brought the Wilson-plot line up to where it again passed through the cloud of data points (*S*26*A*-final). Hg simultaneously behaved in much the same way. At the same time, the atoms were also displaced in x and y by 2 Å as well, but they refined back immediately, and the added perturbation of changes in *A* and *B* seemed to have no effect upon coordinate refinement.

When run *S*26*A* was restarted, but with *B* allowed to refine (*S*26*B*), it was expected that the optimum values of *S*26 would result. This was not the case; *A* and *B* for both Pt and Hg refined to higher numerical values than ever before (12·75 and 14·91 for Pt), and the Wilson-plot line, although pivoting about the experimental points, changed slope in the wrong sense (*S*26*B*-final).

Detailed checks of sign reversals in determinations with these final values of *A* and *B* showed that the signs are much less sensitive to the particular value of *B* than might have been expected, *as long as A* is refined to its corresponding value. Runs *S*26*A* and *B* in their final stages had only eleven and sixteen sign reversals out of 180 reflections when compared with run *S*26, and in each case only one involved a total change in signed figure of merit of greater than 0·50. The majority of changes occurred in weakly determined reflections which would have little effect on a weighted, 'best' Fourier map.

The simultaneous refinement of A and B at 4 Å resolution was no more successful in three dimensions than in two. The most likely explanation is that the limited range of S for which data are present at 4 Å does not force a proper least-squares convergence on the parameters, and that refinement would be much better at higher resolution.

3. The mean figure of merit is a valid criterion of phasing quality *only* if the r.m.s. errors, E_j, have been chosen properly. If this is not so, then almost any $\langle m\rangle$ can be obtained at will. As a check, when E_j's have been chosen properly, then the mean relative error (M.R.E.) will be equal to 0·5 times the number of derivatives.

In a typical two-derivative $hk0$ refinement with E_j's chosen as in equation (16), $\langle m\rangle$ was 0·46. Doubling the E_j's dropped $\langle m\rangle$ to 0·26 and halving them raised $\langle m\rangle$ to 0·59. That these values were in error was indicated by the fact that the mean residual error per derivative, M.R.E./n, dropped from 0·56 to 0·14 in the first case and rose to 2·18 in the second. Blow (1966) has commented that for one derivative, the mean residual error should be 0·5 when the E_j's are properly chosen:

$$\text{M.R.E.} = \frac{1}{N_h} \sum_h \frac{\varepsilon_h^2}{2E^2} = \frac{\langle\varepsilon^2\rangle}{2E^2} = \frac{1}{2} . \qquad (21)$$

The principle has been observed to be generally valid for two, three, four or five derivatives as well. With E_j's chosen as in equation (16), the mean residual error levels off quickly at 0·5 times the number of derivatives.

As a general rule, whenever mean figures of merit are quoted, M.R.E./n should be given as well. If it is not close to 0·5, then the error estimates are out of line and the mean figure of merit has little or no meaning.

4. R factors such as R_K, R_L or the centric R_C, are superior to $\langle m\rangle$ in judging the quality of a phase determination.

These R factors are not influenced by improper error estimates in the way in which $\langle m\rangle$ is. Moreover, the mean figure of merit improves as more degrees of freedom are added in the form of secondary sites, no matter whether these sites are right or wrong, although improvement is less for wrong sites than for right ones. The R factors, in contrast, are improved by true secondary sites and worsened by false ones. Examples of this behavior are to be found in Dickerson *et al.* (1967*b*).

5. If B's are held constant, then the substitution numbers, A_j, tend to refine towards zero for false heavy atom sites.

An example involving badly displaced atoms has been mentioned earlier. Another, involving a wrong secondary site appears in Dickerson *et al.* (1967*b*), and one involving an entirely erroneous derivative is discussed in Dickerson *et al.* (1967*a*).

Operating time

In its space group $P1$ version, with 250 reflections, and with three derivatives having a total of 17 sites, the program took roughly five minutes per phase/refinement cycle on the IBM 7090. In space group $P4_1$, with 180 centric reflections, one derivative with one site took a little over one minute per cycle on the IBM 7094, and five derivatives with a total of nine sites took two minutes per cycle. Again in $P4_1$, with 1416 predominantly acentric reflections, two derivatives with one site each took approximately seven minutes per cycle.

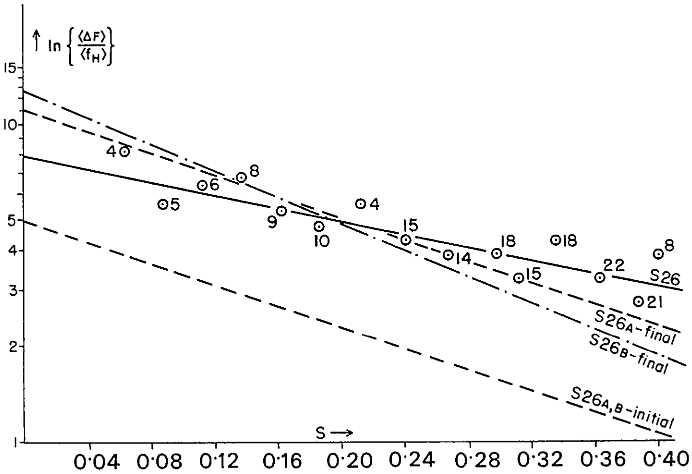

Fig. 3. Wilson plot of $\ln\{\langle \Delta F\rangle/\langle f_j\rangle\}$ *versus* S for the Pt derivative. The number beside each point is the number of reflections contributing to the average value. The solid line represents the best straight line through the points, and is the result of refinement run *S*26. The broken lines represent initial and final A and B values from runs *S*26*A* and *S*26*B*.

Conclusions

The myoglobin relative y refinement, the previously cited work by others, and the cytochrome C trials described here demonstrate that the phase refinement method works well and converges rapidly to the true minimum point. In its adjustment of extent of substitution, A, to compensate for a particular choice of radial falloff factor, B, its tendency to refine the A value of an incorrect atom towards zero, its refinement of positional parameters to the correct location provided that the atom was initially within half the nominal resolution of its proper value, and in the seeming independence of coordinate and A,B refinement, the program appears to be a 'fail-safe' device – that is, one which will correct for reasonable errors or inaccuracies in the initial parameters.

The mean figure of merit seems to be a good relative measure of quality of phase determination if, and only if, the r.m.s. errors have been chosen correctly. It is also unduly affected by the total number of degrees of freedom available or parameters to be refined, whether in the form of several derivatives or many sites per derivative. The two R factors, R_K and R_L, are less sensitive to such factors and are better measures of the phase analysis. Both R factors distinguish clearly between true and false minor sites, whereas the figure of merit behavior is less straightforward. The least-squares R factor, R_L, being directly related to the quantity minimized, is theoretically the better of the two criteria to watch, yet in the limited experience of the cytochrome work, it and the Kraut R factor behave in an identical manner aside from the tendency of R_L to be quite small for single derivatives. R_L is consistently two to five percentage points lower than R_K, so it is important to distinguish clearly between them. Until their relationships are more clearly understood, it would be well to compute and report both. Two of the three quantities, R_K, R_C and $\Sigma\,|\Delta F|/\,\Sigma\,|F|$, should be reported for the centric reflections.

From a practical standpoint, the time to stop refining is the moment at which the change in phase angles from one cycle to the next ceases to be significant. The phase-changing effect of small positional errors is much greater than that of what would appear at first glance to be comparable errors in substitution number or radial falloff factor. This is because a major error (say 25%) in A or B will displace the end of the heavy atom vector only a short distance on the phase plane, while a similar error in x, y or z will cause the heavy atom vector to swing wildly around the origin for all but the lowest order reflections. The consequence of this situation is that there is considerable phase change as positional coordinates refine, but that the fine adjustments of scale factor and substitution number make little difference to the phase analysis. At the end of refinement of cytochrome C as judged by the leveling off of parameter shifts and changes in figure of merit and the R factors, the mean change in phase angle from one cycle to the next had dropped to a degree or less.

This work was carried out with the support of United States Public Health Grants GM 10862 (formerly A-3185) and GM 12121, whose help is gratefully acknowledged. One of us (R.A.P.) was also supported as a post-doctoral fellow by GM 10862, and another (J.E.W.) has been the recipient of a National Institutes of Health Predoctoral Traineeship (GM 1262).

References

BLOW, D. M. (1966). Private communication.
BLOW, D. M. & CRICK, F. H. C. (1959). *Acta Cryst.* **12**, 794.
CULLIS, A. F., MUIRHEAD, H., PERUTZ, M. F., ROSSMANN, M. G., & NORTH, A. C. T. (1961). *Proc. Roy. Soc.* A**265**, 15.
DICKERSON, R. E., KENDREW, J. C. & STRANDBERG, B. E. (1961). *Acta Cryst.* **14**, 1188.
DICKERSON, R. E., KOPKA, M. L., VARNUM, J. C. & WEINZIERL, J. E. (1967*a*). *Acta Cryst.* **23**, 511.
DICKERSON, R. E., KOPKA, M. L., BORDERS, C. L., VARNUM, J. C., WEINZIERL, J. E. & MARGOLIASH, E. (1967*b*). *J. Mol. Biol.* **29**, 77.
HART, R. G. (1961). *Acta Cryst.* **14**, 1194.
KRAUT, J. (1961). Private communication.
KRAUT, J., SIEKER, L. C., HIGH, D. F. & FREER, S. T. (1962). *Proc. Nat. Acad. Sci. Wash.* **48**, 1417.
LIPSCOMB, W. N., COPPOLA, J. C., HARTSUCK, J. A., LUDWIG, M. L., MUIRHEAD, H., SEARL, J. & STEITZ, T. A. (1966). *J. Mol. Biol.* **19**, 423.
MATTHEWS, B. W. (1966). *Acta Cryst.* **20**, 82.
MUIRHEAD, H. (1966). Private communication.
NORTH, A. C. T. (1965). *Acta Cryst.* **18**, 212.
ROSSMANN, M. G. (1960). *Acta Cryst.* **13**, 221.