Acta Cryst. (1965). **19**, 883

Combination of Multiple Isomorphous Replacement and Anomalous Dispersion Data for Protein Structure Determination. III. Refinement of Heavy Atom Positions by the Least-Squares Method

BY GOPINATH KARTHA

Center for Crystallographie Research, Roswell Park Memorial Institute, Buffalo, New York, U.S.A.

(*Received* 27 *November* 1964 *and in revised form* 2 *February* 1965)

In two earlier communications under the above title from this laboratory, it was shown how the combination of isomorphous replacement and anomalous dispersion data can be used for establishing and correlating the positions of the substituted heavy atoms in the derivative crystals. By suitably combining the difference in amplitudes between the free protein and its heavy atom derivative with the difference between the Friedel pairs for the heavy atom derivative, it is possible to obtain a quantity representing the length of the heavy atom vector. This quantity, computed purely from experimental data, could be used in applying the usual least-squares techniques for refining the positional and thermal parameters of heavy atoms in the derivatives.

Introduction

For the satisfactory determination of phase angles in proteins by the multiple isomorphous series method (Harker, 1956; Dickerson, Kendrew & Strandberg, 1961) it is essential to have as the starting point as accurate an estimate as possible of the substituted heavy atom parameters. In two recent communications from this laboratory (Kartha & Parthasarathy, 1964*a*, *b*; hereinafter referred to as parts I and II) it was shown that the combination of information obtained from isomorphous derivative and parent protein crystals with the anomalous scattering information from the heavy atom derivative leads to easier location and correlation of the heavy atom positions. In this paper it is shown that a similar combination leads to better refinement of the heavy atom parameters and, consequently, a more reliable starting point for the protein phase angle determination. In general, wherever possible, the notation used here will follow that of parts I and II.

Heavy atom parameters

After finding the heavy atoms in the various derivatives it becomes necessary to refine, as best as one can, their various parameters. Assuming that the free protein as well as the heavy atom derivative data have been placed on the same scale, the main parameters to be refined are (*a*) the coordinates of the heavy atom (*b*) their thermal parameters, and (*c*) the occupancy factor.

As pointed out by Dickerson *et al.* (1961), use of the complete set of observed reflections for doing the usual least-squares refinement of atomic parameters against observed data became difficult, because one does not have information about the phase angles to start with, and without this one cannot find for the heavy atom scattering a measured quantity $|F_H|_{obs}$ against which the calculated $|F_H|_{calc}$ could be compared. Hence, in some of the earlier attempts at refinement (Hart, 1961) only a small part of the data was used, *viz.* the centrosymmetric reflections only. In such a case, lack of phase angle information reduces to a sign ambiguity and one can get an experimental value of $|F_H|_{obs}$ by assigning that sign which gives the smaller difference between $|F_H|_{obs}$ and $|F_H|_{calc}$. In addition to the fact that it uses only a very small part of the data, this method has the disadvantage of not being able to refine some of the parameters like the y parameters in the case of space group $P2_1$. Rossmann's (1960) method, in which he refines the peaks in his correlation maps, did, in fact, enable him to use the data from the acentric reflections also. However, his method of weighting reduces the effect of a large number of reflections, and also makes uncertain the parameters of any atoms that have the same, or nearly the same, positions in the two derivatives. This is obvious if one remembers that in his correlation maps the self-Patterson peaks are negative and, hence, any common site will have no maxima in the correlation maps, the positive and negative peaks cancelling each other. This makes the refinement of parameters of common sites uncertain. Unfortunately, this possibility occurs more often than one expects from pure chance, and, hence, it would be advantageous to have a method where one refines the atomic parameters in one heavy atom derivative by comparing it only with the free protein data. In the method indicated below, the coordinates in each derivative are refined independently of other derivatives.

Length of heavy atom vector $|\mathbf{F_H}|_{obs}$

By combining the available experimental information from the anomalous scattering of the heavy atom in the heavy atom derivative, in addition to the usual isomorphous derivative data, it is, however, possible to supplement, in part, the unknown phase angle in-

formation and evaluate a quantity $|F_H|_{obs}$ which can be compared with the length of the calculated heavy atom vector $|F_H|$. It is then possible to apply ordinary least-squares refinement techniques to improve the heavy atom parameters in the usual way.

Consider Fig. 1, which presents the vector diagram showing the relationships between the magnitudes and phase angles of the three vectors $\mathbf{F}_P$, $\mathbf{F}_H$ and $\mathbf{F}_{PH}$. From this diagram it is seen that

$$|F_H|^2=|F_{PH}|^2+|F_P|^2-2|F_{PH}||F_P|\cos(\alpha_{PH}-\alpha_P)$$
$$=(|F_{PH}|-|F_P|)^2+2|F_{PH}||F_P|\{1-\cos(\alpha_{PH}-\alpha_P)\} \quad (1)$$

Now the first term on the right hand side of equation (1) is the square of the difference in magnitude of the structure amplitudes $|F_{PH}|$ and $|F_P|$ and, hence, is an observed quantity. This difference has been designated $\Delta|F|_{iso}$ in part I (equation 3). The earlier refinements were based essentially on this part only. It is the second part that involves phase angles and, hence, is indeterminate.

We can, however, get an estimate of the second part, based on the anomalous scattering information available from the heavy atom derivative. From Fig. 1, consideration of triangle ABC gives

$$|F_H|/\sin(\alpha_{PH}-\alpha_P)=|F_P|/\sin(\alpha_H-\alpha_{PH})$$

so that

$$\sin(\alpha_{PH}-\alpha_P)=(|F_H|/|F_P|)\sin(\alpha_H-\alpha_{PH}) . \quad (2)$$

In part I (equation 26), we have defined $|F_H|\sin(\alpha_{PH}-\alpha_H)$ as $\Delta|F|_{ano}$ and this quantity is obtained from the difference in the measured structure amplitude and its Friedel conjugate, and rescaling this difference by multiplying by the factor $f''_H/2f''_H$ where f' and f'' are the real and imaginary components of the anomalously scattering heavy atom. In practice, however, this scaling factor is obtained by comparing the measured isomorphous derivative and anomalous scattering differences. Making this substitution, we get the expression

$$\sin(\alpha_{PH}-\alpha_P)=-\Delta|F|_{ano}/|F_P| . \quad (3)$$

From equation (3), it is seen that $\sin(\alpha_{PH}-\alpha_P)$ can be deduced from measured quantities based on anomalous scattering information and from this one can easily compute the quantity $\cos(\alpha_{PH}-\alpha_P)$, except for a sign ambiguity, occurring in the right hand side of equation (1). If we make the very reasonable assumption that the difference between the phase angles of the free proteins and the heavy atom derivative, *i.e.* $(\alpha_{PH}-\alpha_P)$ rarely exceeds $\pi/2$, we can rule out the negative value of $\cos(\alpha_{PH}-\alpha_P)$. This leads to the expression

$$|F_H|^2=(\Delta|F|_{iso})^2+2|F_{PH}||F_P|$$
$$\times[1-\{1-(\Delta|F|_{ano}/|F_P|)^2\}^{\frac{1}{2}}] \quad (4)$$

where the positive square root is always implied in equation (4). In most cases it may further be that $|F_H|$ is small compared with $|F_P|$ and $|F_{PH}|$ and for all practical purposes one could replace $|F_P||F_{PH}|$ in equation (4) by $|F_P|^2$, thus reducing this equation to equation (24) of part I. In an actual case studied, it was found that this assumption caused no additional errors and led to same final result as when the more accurate equation (4) was used. All terms on the right hand side of equation (4) being known, the actual length of the vector $\mathbf{F}_H$ can be computed from measured quantities alone and this length can be denoted by $|F_H|_{obs}$. These values can be used for refinement of the heavy atom parameters by the least-squares method by comparing them with the quantities $|F_H|_{calc}$, the length of the heavy atom vector computed on the basis of the assumed parameters. It may be pointed out that even the assumption that $(\alpha_{PH}-\alpha_P)$ is always less than $\pi/2$ and, hence, only the positive square root need be taken

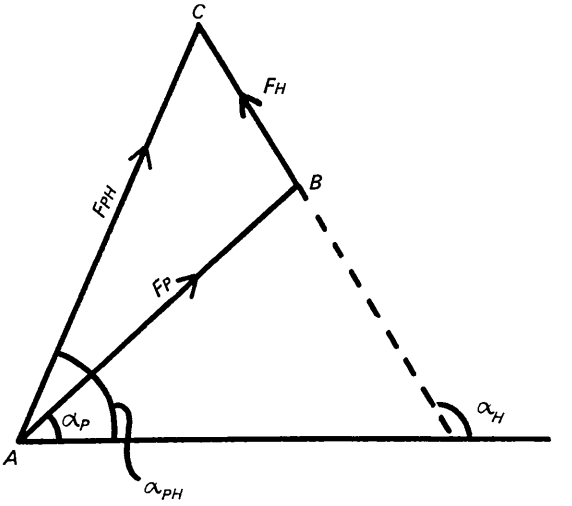

Fig. 1. Vector diagram showing relationship between magnitudes and phase angles of F_P, F_{PH} and F_H.

Table 1. *Discrepancy residuals for four cycles of least-squares refinement*

$|F|_{obs1}$ is full $|F_H|$ amplitude and $|F|_{obs2}$ is the amplitude designated in the text as $|F|_{iso}$. *R* values are given in percentages. All five derivatives contained two main heavy atom sites.

	R with $\|F\|_{obs1}$				*R* with $\|F\|_{obs2}$			
Derivative	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 1	Cycle 2	Cycle 3	Cycle 4
1	56·9	41·7	36·4	36·3	63·2	59·6	53·1	52·9
2	56·4	49·8	43·0	42·8	61·2	60·5	53·3	53·3
3	49·6	47·7	39·4	39·4	57·6	57·4	52·7	52·8
4	63·2	49·7	48·0	47·3	67·6	63·6	60·1	59·6
5	59·8	45·6	42·0	41·8	63·5	59·9	53·2	53·0

in equation (4), can be avoided by adopting a least-squares procedure similar to that of Hart (1961) by comparing $|F_H|_{calc}$ with the value of $|F_H|_{obs}$, using both possible signs of the square root. However, it is very doubtful whether this extra complication is worth while in most practical cases of protein heavy atom derivatives.

Practical application to ribonuclease

The heavy atom parameters used in determining phase angles of reflections from ribonuclease were refined by the least-squares method. The six anisotropic thermal parameters as well as three positional coordinates of the heavy atoms were varied, the occupancy being fixed by the peak heights in the heavy atom difference Fourier synthesis. Refinements were carried out by comparing $|F_H|_{calc}$ against (i) $|F|_{obs1}$ obtained from equation (4) in its simplified form and (ii) $|F|_{obs2} = \Delta|F|_{iso}$. In Table 1 are given the results of least-squares runs for heavy atom refinements for five heavy atom derivatives. In this table all derivatives contained two heavy atoms per asymmetric unit and about 1000 reflections were used in performing the refinement.

It is seen that even though the discrepancies are very large compared with usual standards of single-crystal X-ray analysis, for all the derivatives for which refinement computations were run, the discrepancy index R was 10 to 16% lower when refinements were carried out with $|F|_{obs1}$. However, in the cases studied of ribonuclease derivatives, where the number of heavy atoms refined was only two, refinements with $|F|_{obs1}$ and $|F|_{obs2}$ both gave practically the same positional parameters when more than 1000 reflections within 4 Å sphere were used in the least-squares cycles. This was repeatedly tested by deliberately shifting the input coordinates from their correct positions and then subjecting them to least-squares refinements, first with amplitudes $|F|_{obs1}$ and then $|F|_{obs2}$. It was noticed that, as long as the shifts were less than about 1·5 Å, both refinement runs converged at about the same rate to the correct heavy atom positions. It may be that, in spite of the much larger disagreement between $|F|_{obs}$ and $|F|_{calc}$ in the case of $|F|_{obs2}$ refinement, the ratios of the number of reflections to number of parameters in these cases are so large that the coordinate refinements were equally fast in both cases. However, this was not the case for the thermal parameters; those obtained by refinement involving $|F|_{obs1}$ converge faster and look more reasonable than those when $|F|_{obs2}$ are used. There is reason to hope that, even in the case of positional parameters, the refinement rates would be better for the $|F|_{obs1}$ amplitudes for problems where the ratio of number of reflections to number of coordinates is not as large as in this case.

The author is grateful to Professor David Harker and Dr R. Parthasarathy for many interesting discussions he had with them. Thanks are also due to Drs D. R. Harris, H. H. Mills and Miss Elaine DeJarnette for help in collection and processing of the data used in some of these computations. The investigation was supported by the Public Health Service Research Career Development Program Award GM-K3-16737 from the National Institutes of Health and research grants NIH-A-3942, NSF-GB-429, and by the Dean Langmuir Foundation.

References

DICKERSON, R. E., KENDREW, J. C. & STRANDBERG, B. E. 1961). *Acta Cryst.* **14**, 1188.
HARKER, D. (1956). *Acta Cryst.* **9**, 1.
HART, R. G. (1961). *Acta Cryst.* **14**, 1194.
KARTHA, G. & PARTHASARATHY, R. (1965*a*). *Acta Cryst.* **18**, 745.
KARTHA, G. & PARTHASARATHY, R. (1965*b*). *Acta Cryst.* **18**, 749.
ROSSMANN, M. G. (1960). *Acta Cryst.* **13**, 221.