On the Application of Phase Relationships to Complex Structures. XXII. Techniques for Random Phase Refinement

By T. Debaerdemaeker

Sektion für Röntgen- und Electronenbeugung, Universität Ulm, Oberer Eselsberg, D-7900 Ulm, Federal Republic of Germany

and M. M. Woolfson

Physics Department, University of York, Heslington, York YO1 5DD, England

Abstract

In recent years there have been a number of developments in direct methods involving refinement processes applied to initially random sets of phases. Procedures which have been used for refinement include least-squares and gradient methods applied to triple-phase relationships expressed as linear equations and also the tangent formula. In the present investigation seven functions are investigated for the refinement of random phases; because of the awkward form of these functions the refinement process used is based on a parameter-shift algorithm. Some of the functions appear to be more effective than others but the most effective one was discovered through making a mistake with one of the others and no rational explanation for its efficacy can be given. Trials have been made with known structures and with three unknown structures which were originally solved by the processes described in the paper.

Introduction

In recent years a great deal of attention has been paid to multisolution direct methods which refine, or refine and extend, initially random sets of phases. Using random phases, with phase relationships expressed as linear equations, Baggio, Woolfson, Declercq & Germain (1978) showed how true phases came from a refinement process in a limited number of trials. This is the basis for the very successful direct-methods procedure *YZARC* (Declercq, Germain, Woolfson & Wright, 1981). Furusaki (1979) has described a Monte Carlo method where a large *MULTAN* starting set is given random phases which are then extended and refined by the tangent formula. In the *RANTAN* procedure (Yao Jia-xing, 1981) a very large set of reflexions, sometimes of sufficient size to define the structure, has allocated random phases and initial

weights and these phases are then refined by a controlled use of a weighted tangent formula.

Here there will be examined various refinement procedures, based on different functions, using as a refinement procedure the parameter-shift method described by Bhuiya & Stanley (1963). This has two advantages – firstly that it offers the possibility of jumping over, or escaping from, subsidiary maxima or minima, and secondly that it can be applied to awkward functions of the variables being refined.

The refinement functions

(I) *The first group of functions*

Here there will be considered refinement functions based on Sayre's equation (Sayre, 1952) which may be written in the form:

$$|E_h|^2 = K_h \sum_{h'} E_h E_{-h'} E_{h'-h}. \tag{1}$$

Taking the real and imaginary sides of the equation separately this gives

$$|E_h|^2 = K_h \sum_{h'} |E_h E_{h'} E_{h-h'}| \cos(\varphi_h - \varphi_{h'} - \varphi_{h-h'}) \tag{2}$$

and

$$0 = K_h \sum_{h'} |E_h E_{h'} E_{h-h'}| \sin(\varphi_h - \varphi_{h'} - \varphi_{h-h'}). \tag{3}$$

Because these equations are going to be applied to structures with non-equal atoms and with only a limited number of terms in the summations it is necessary to find a scale factor, S, such that the equations

$$|E_h|^2 = S \sum_{h'} |E_h E_{h'} E_{h-h'}| \cos(\varphi_h - \varphi_{h'} - \varphi_{h-h'}) \tag{4}$$

apply as closely as possible. One way of doing this is

through the result given by Germain, Main & Woolfson (1970):

$$\langle \cos(\varphi_h - \varphi_{h'} - \varphi_{h-h'})\rangle = \frac{I_1(\kappa_{h,h'})}{I_0(\kappa_{h,h'})} = \xi_{h,h'}, \qquad (5)$$

where $I_1(x)$ and $I_0(x)$ are modified Bessel functions and

$$\kappa_{h,h'} = 2N^{-1/2}|E_h E_{h'} E_{h-h'}| \qquad (6)$$

for N equal atoms in the unit cell.

A single overall scale factor for all the equations can be found as

$$S = \frac{\sum_h |E_h|^2}{\sum_h \sum_{h'} |E_h E_{h'} E_{h-h'}| \xi_{h,h'}}. \qquad (7)$$

This gives rise to our first refinement function

$$\psi_A = \sum_h \{(SX_h - |E_h|^2)^2 + (SY_h)^2\}, \qquad (8)$$

where

$$X_h = \sum_{h'} |E_h E_{h'} E_{h-h'}| \cos(\varphi_h - \varphi_{h'} - \varphi_{h-h'}) \qquad (9)$$

and

$$Y_h = \sum_{h'} |E_h E_{h'} E_{h-h'}| \sin(\varphi_h - \varphi_{h'} - \varphi_{h-h'}). \qquad (10)$$

The function ψ_A would be expected to be a minimum for a correct set of phases.

Experience with individual Sayre-type equations revealed that, with an overall scale factor, the agreement between left-hand and right-hand sides strongly depended on the number of terms in the summation. For this reason there was tried an individual scale factor for each equation, given by

$$S_h = |E_h|^2 / \sum_{h'} |E_h E_{h'} E_{h-h'}| \xi_{h,h'}. \qquad (11)$$

This gives a second refinement function

$$\psi_B = \sum_h \{(S_h X_h - |E_h|^2)^2 - (S_h Y_h)^2\}, \qquad (12)$$

which again should be a minimum.

(II) The second group of functions

The usual criterion for a good set of phases, which is implicit in the MULTAN figure of merit ABSFOM, is having a high value for the sum of all the quantities X_h in (9). For the next refinement function this condition is combined with having the quantities Y_h as small as possible by requiring

$$\psi_C = \sum_h \{X_h - |Y_h|\} \qquad (13)$$

to be large.

During tests with ψ_C it was found by accident that another function to be maximized

$$\psi_D = \sum_h \{X_h - Y_h\} \qquad (14)$$

also works well although the reason for this is somewhat obscure. In fact from (9) and (10) one finds

$$\psi_D = \sqrt{2} \sum_h \sum_{h'} |E_h E_{h'} E_{h-h'}|$$
$$\times \cos(\varphi_h - \varphi_{h'} - \varphi_{h-h'} + \pi/4), \qquad (15)$$

which seems to drive the values of triple-phase relationships towards $-\pi/4$. It would be just as reasonable to replace the difference of X_h and Y_h in (14) by the sum, in which case it would appear that the relationship was being driven towards $+\pi/4$. Despite the fact that a rational basis for function ψ_D cannot be found we have investigated its use.

(III) The third group of functions

The result given in (5) for the expectation value of the cosine of a triple-phase invariant leads to the refinement function

$$\psi_E = \sum_h \sum_{h'} |E_h E_{h'} E_{h-h'}|$$
$$\times \{\cos(\varphi_h - \varphi_{h'} - \varphi_{h-h'}) - \xi_{h,h'}\}^2, \qquad (16)$$

which should be minimized.

In ψ_E only the expectation value of the cosine invariant is included. The fact that the sine invariant is expected to be close to zero may be incorporated in the form

$$\psi_F = \sum_h \sum_{h'} |E_h E_{h'} E_{h-h'}| \{[\cos(\varphi_h - \varphi_{h'} - \varphi_{h-h'}) - \xi_{h,h'}]^2$$
$$+ \sin^2(\varphi_h - \varphi_{h'} - \varphi_{h-h'})\}, \qquad (17)$$

which should be minimized. After simplification this leads to the refinement function

$$\psi_F = \sum_h \sum_{h'} |E_h E_{h'} E_{h-h'}| \xi_{h,h'} \cos(\varphi_h - \varphi_{h'} - \varphi_{h-h'}), \qquad (18)$$

which should be maximized. A similar refinement function has been used by Hull, Viterbo, Woolfson & Zhang Shao-hui (1981).

Another variant of this general type of function involves using the expectation value of $\sin^2(\varphi_h - \varphi_{h'} - \varphi_{h-h'})$. From the basic probability distribution for the triple-phase invariants (Cochran, 1955) it is readily found that

$$\langle \cos^2(\varphi_h - \varphi_{h'} - \varphi_{h-h'})\rangle = \frac{1}{2} + \frac{1}{2}\frac{I_2(\kappa_{h,h'})}{I_0(\kappa_{h,h'})} \qquad (19a)$$

and

Table 1. *Tests of refinement functions*

The bracketed pair of numbers following the name of the structure gives the number of phases and number of relationships in the system. NT means 'number of trials' and MPE means 'mean phase error in degrees'. All results with MPE ≤ 60 are given.

Function	ψ_A		ψ_B		ψ_C		ψ_D		ψ_E		ψ_F		ψ_G	
Structure	NT	MPE	NT	MPE	NT	MPE	NT	MPE	NT	MPE	NT	MPE	NT	MPE
Cinobufagin (200/5769)	52	60	55		59	23 41	49	18 20	73		60	27 39	68	
Ergocalciferol (200/5157)	128		62		49	49 59	49	47	90		84		84	
Factor S (100/1584)	199	53 57	199	58	54	55 57	154	56	113	53	177	37 56	199	55 59
Munich (100/1506)	95		138	58	60	41 45 47 52 52 60	66	15	190	53 53 54 56 58	190	29 31 37 39 42 45 45 46 46 46 48 48 49 51 52 52 52 53 54 55 55 58		

$$\langle \sin^2(\varphi_h - \varphi_{h'} - \varphi_{h-h'}) \rangle = \frac{1}{2} - \frac{1}{2}\frac{I_2(\kappa_{h,h'})}{I_0(\kappa_{h,h'})}. \quad (19b)$$

This leads to the final refinement function

$$\psi_G = \sum_h \sum_{h'} |E_h E_{h'} E_{h-h'}| \left\{ [\cos(\varphi_h - \varphi_{h'} - \varphi_{h-h'}) - \xi_{h,h'}]^2 \right.$$
$$\left. + \left| \sin^2(\varphi_h - \varphi_{h'} - \varphi_{h-h'}) - \frac{1}{2} + \frac{1}{2}\frac{I_2(\kappa_{h,h'})}{I_0(\kappa_{h,h'})} \right| \right\}.$$

Some experiments will now be described using the seven refinement functions ψ_A to ψ_G.

Results

For four structures trials have been carried out with all seven of the refinement functions. These are: (i) cinobufagin: $C_{26}H_{27}O_6$, $P2_12_12_1$, $Z = 4$ (Declercq, Germain & King, 1977); (ii) ergocalciferol (vitamin D_2): $C_{28}H_{44}O$, $P2_12_12_1$, $Z = 8$ (Hull, Leban, Main, White & Woolfson, 1976); (iii) factor S: $C_{43}H_{49}N_7O_{10}$, $P2_12_12_1$, $Z = 4$ (Declercq, Germain, Van Meerssche, Hull & Irwin, 1978); (iv) munich: $C_{39}H_{16}$, $C2$, $Z = 8$ (Szeimies-Seebach, Harnisch, Szeimies, Van Meerssche, Germain & Declercq, 1978).

Of these structures cinobufagin is fairly straightforward. At the time it was solved by one of the techniques described here munich had resisted the solution by other direct methods. The other two structures are both very challenging and usually can only be solved by developing initial fragments representing a small fraction of the total structure.

The results are shown in Table 1. The interesting observation is that function ψ_D, the one for which there is no rational explanation, seems to perform better than the others and other tests, not reported here, support

Table 2. *Solutions of three unknown structures with refinement function ψ_D*

The notation is as for Table 1.

Structure	NT	MPE
WDS 117[a] (300/6000)	44	37
DEBH 22[b]	36	23 23 24 37 37 39 47 48
CH3PC2H5[c] (250/6750)	22	14 14 15 19 19 20 20 20 20 20 20 55
(350/10050)	15	18 23 23 23 24

References: (a) Debaerdemaeker (1981): $C_{28}H_{18}Cl_4N_2O_3$, $P2_1/n$, $Z = 4$; (b) Friedrichsen, Schröer & Debaerdemaeker (1980): $C_{14}H_{22}N_2O_3S$, $P2_12_12_1$, $Z = 8$; (c) Debaerdemaeker (1982): $C_{25}H_{23}OP.Cr(CO)_3$, $P2_12_12_1$, $Z = 4$.

this view. Indeed, this function has been used to solve three unknown structures and the test results for these are shown in Table 2.

A characteristic of several of the refinement functions described here is that they do not tend to drive the triple-phase invariants towards zero (modulo 2π). Methods which do drive invariants towards zero have the characteristic that the errors in phases are heavily correlated *via* the phase relationships. This has an adverse effect on the quality of the resultant E maps (Silva & Viterbo, 1980). If the errors in phases are not so heavily correlated, as will be so for many of the functions here described, then even relatively large r.m.s. errors can give reasonable E maps with considerable fragments visible. In tests we have found that, with a sufficiently large number of reflexions, phase sets with r.m.s. errors up to $60°$ can give developable fragments which is why we have chosen this limit in Tables 1 and 2.

We conclude that, notwithstanding the theoretical uncertainties, the refinement function ψ_D can be a useful aid to solving crystal structures from initially

random phases and we intend to distribute a computer package, based on the *MULTAN* system, to implement this and perhaps one or two of the other refinement functions as well. The computer time necessary for these parameter-shift refinement programs is not excessive and, as it turns out, for the second group of functions which includes ψ_D the timings are least. As an example, for the structure CH3PC2H5 (Table 2) with 350 reflexions and 10 050 relationships, the refinement time was about 4 min per trial on a Telefunken TR440 and would take about one half of that time or less on a DEC System 10 computer. Perhaps it should be said that while there is no strong indication that these refinement procedures are markedly better than those already in use they are different, are comparable in power and offer an alternative approach for solving difficult structures.

We wish to express our gratitude to the Computing Centre of the University of Ulm for the use of its facilities and to the Deutches Forschungsgemeinschaft and North Atlantic Treaty Organisation for support of our collaborative activities.

References

BAGGIO, R., WOOLFSON, M. M., DECLERCQ, J. P. & GERMAIN, G. (1978). *Acta Cryst.* A34, 883–892.
BHUIYA, A. K. & STANLEY, E. (1963). *Acta Cryst.* 16, 981–984.
COCHRAN, W. (1955). *Acta Cryst.* 8, 473–478.
DEBAERDEMAEKER, T. (1981). Unpublished.
DEBAERDEMAEKER, T. (1982). To be published in *Cryst. Struct. Commun.*
DECLERCQ, J. P., GERMAIN, G. & KING, G. S. D. (1977). Fourth European Crystallographic Meeting, Oxford. Abstracts A, pp. 279–280.
DECLERCQ, J. P., GERMAIN, G., VAN MEERSSCHE, M., HULL, S. E. & IRWIN, M. J. (1978). *Acta Cryst.* B34, 3644–3648.
DECLERCQ, J. P., GERMAIN, G., WOOLFSON, M. M. & WRIGHT, H. (1981). *Acta Cryst.* A37, C237.
FRIEDRICHSEN, W., SCHRÖER, W. D. & DEBAERDEMAEKER, T. (1980). *Justus Liebigs Ann. Chem.*, pp. 1850–1858.
FURASAKI, A. (1979). *Acta Cryst.* A35, 220–224.
GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1970). *Acta Cryst.* B26, 274–285.
HULL, S. E., LEBAN, I., MAIN, P., WHITE, P. S. & WOOLFSON, M. M. (1976). *Acta Cryst.* B32, 2374–2381.
HULL, S. E., VITERBO, D. WOOLFSON, M. M. & ZHANG SHAO-HUI (1981). *Acta Cryst.* A37, 566–572.
SAYRE, D. (1952). *Acta Cryst.* 5, 60–65.
SILVA, A. M. & VITERBO, D. (1980). *Acta Cryst.* A36, 1065–1070.
SZEIMIES-SEEBACH, U., HARNISCH, J., SZEIMIES, G., VAN MEERSSCHE, M., GERMAIN, G. & DECLERCQ, J. P. (1978). *Agnew. Chem. Int. Ed. Engl.* 17, 848–850.
YAO JIA-XING (1981). *Acta Cryst.* A37, 642–644.

A Strategy for Combining Restrained Least Squares with Computer Graphics in the Refinement of Protein Structures

BY A. J. MORFFEW

IBM UK Scientific Centre, Athelstan House, St Clement Street, Winchester, Hampshire SO23 9DR, England

AND D. S. MOSS

Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, England

Abstract

A method of combining interactive molecular graphics with restrained least-squares refinement is described. This enables a researcher using the interactive computer graphics terminal to weight the shifts applied to the individual atoms in following cycles of refinement according to the confidence in their positions. The weights are incorporated into the least-squares procedure using the Marquardt factor which was originally introduced to handle the ill-conditioned case [Moss & Morffew (1982). *Comput. Chem.* 6, 1–3; Marquardt

(1963). *SIAM J. Appl. Math.* 11, 431–441]. The described refinement strategy has been imlemented at Birkbeck College, London, and, using avian pancreatic polypeptide (APP) as test data, some refinements have been carried out. In order to compare the results of these refinements, histograms have been drawn that show the distribution of distances between the corresponding atoms in the different models. These histograms show that the model is improved by including the 'confidence' weighting. An improvement of 7·5% was achieved in the mean distance between the corresponding atoms in the fully refined model and a