

## Combining Direct Methods with Isomorphous Replacement or Anomalous Scattering Data. I\*

BY FAN HAI-FU, HAN FU-SON, QIAN JIN-ZI AND YAO JIA-XING

*Institute of Physics, Chinese Academy of Sciences, Beijing, China*

(Received 2 August 1983; accepted 20 March 1984)

### Abstract

A direct-methods procedure has been proposed for separating the phase doublet resulting from the use of either isomorphous replacement or anomalous scattering techniques. The phase doublet is expressed as  $\varphi_{\mathbf{H}} = \varphi'_{\mathbf{H}} \pm |\Delta\varphi_{\mathbf{H}}|$ . Formulae combining the structure-factor relationships with the phase-doublet information are given. Problems concerning the practical applications are also discussed. A test calculation with the error-free data for the protein insulin showed a satisfactory result.

### Introduction

The idea of combining direct methods with isomorphous replacement or anomalous scattering data was first introduced in the 1960s by several authors independently. Coulter (1965) suggested the use of the tangent formula with starting phases extracted from the single isomorphous replacement data. This method does not make full use of the information which could be obtained from a single isomorphous pair. Fan (1965*a*) and Karle (1966) suggested the use of 'component relationships', *i.e.* the relationships among the real and imaginary components of the structure factors. If the arrangement of the heavy atoms is centrosymmetric then this makes the problem of breaking the phase ambiguities just that of finding the signs for the real or imaginary components of the structure factors by direct-method procedures. However, if the arrangement of heavy atoms is non-centrosymmetric then the component relation is not convenient to use. A phase-difference relation is given here together with the associated probability formula. This enables one to treat the problem of phase ambiguities arising from single isomorphous replacement (SIR), as well as that from one-wavelength anomalous scattering (OAS), by a simple and unified manner, no matter what the arrangement of the heavy atoms is. Recently, Hauptman (1982*a, b*) integrated the probabilistic theory of the three-phase structure invariants with the techniques of isomorphous replacement and anomalous scattering leading to a series of complex formulae. Our method differs from

Hauptman's in that, instead of two sets of diffraction data, *e.g.* one for the native protein and one for the derivative in the SIR case, only one set of diffraction data but with the phase-doublet information are introduced into the probability formula.

### Enantiomorphous phase doublets from isomorphous replacement or anomalous scattering

In the case of SIR (see Blundell & Johnson, 1976, for details), for a given reciprocal vector  $\mathbf{H}$ , we have

$$F_{\mathbf{H},P} = F_{\mathbf{H},PQ} - F_{\mathbf{H},Q}, \quad (1)$$

where  $F_{\mathbf{H},P}$  is the structure factor of the native protein,  $F_{\mathbf{H},PQ}$  is that of the heavy-atom derivative and  $F_{\mathbf{H},Q}$  the heavy-atom contribution to  $F_{\mathbf{H},PQ}$ . From experiment the magnitudes of  $F_{\mathbf{H},P}$  and  $F_{\mathbf{H},PQ}$  can be obtained. Accordingly, the parameters of the heavy atoms can be found and  $F_{\mathbf{H},Q}$  be calculated. Hence, we have two ways for drawing the triangle of (1) leading to an enantiomorphous phase doublet for  $F_{\mathbf{H},P}$  or for  $F_{\mathbf{H},PQ}$  in the phase-vector diagram, as shown in Fig. 1. The phase doublets are of the form

$$\varphi_{\mathbf{H}} = \varphi'_{\mathbf{H}} \pm |\Delta\varphi_{\mathbf{H}}|, \quad (2)$$

where  $\varphi_{\mathbf{H}}$  is the phase of the structure factor  $F_{\mathbf{H}}$  for the native protein or for the heavy-atom derivative,

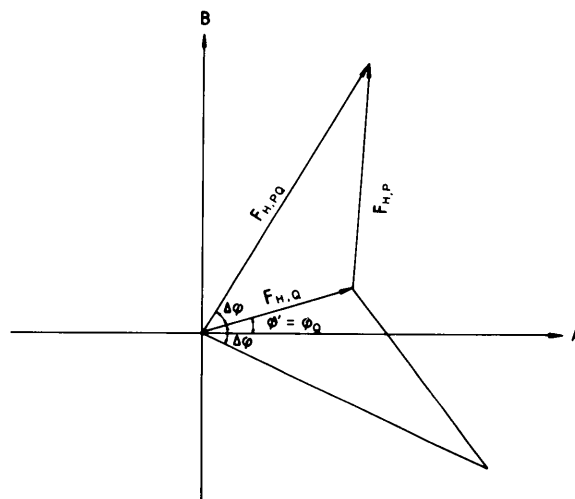


Fig. 1. Phase-vector diagram showing the enantiomorphous phase doublet arising from the SIR method.

\* Part of this paper was presented at the IUCr Summer School on Crystallographic Computing, Kyoto, Japan, 1983.

$\varphi'_H$  equals the phase of  $F_{H,Q}$  and  $\Delta\varphi_H$  is the difference between  $\varphi_H$  and  $\varphi'_H$ .

In the case of OAS (see Blundell & Johnson, 1976), we have

$$F_{H,PQ}^+ = F_{H,PQ} + F_{H,Q}'' \quad (3)$$

and

$$F_{H,PQ}^{-*} = F_{H,PQ} - F_{H,Q}'' \quad (4)$$

Here  $F_{H,PQ}$  is the contribution of both the normal scattering and the real part of the anomalous scattering,  $F_{H,Q}''$  is the contribution of the imaginary part of the anomalous scattering,  $F_{H,PQ}^{-*}$  denotes the conjugate of  $F_{H,PQ}$ . Subtracting (4) and (3), it follows that

$$F_{H,PQ}^+ - F_{H,PQ}^{-*} = 2F_{H,Q}'' \quad (5)$$

The magnitudes of  $F_{H,PQ}^+$  and  $F_{H,PQ}^{-*}$  can be obtained from experiment, while  $F_{H,Q}''$  can then be derived. Hence we also have two ways for drawing the triangle of (5) leading to an enantiomorphous phase doublet for  $F_{H,PQ}$ , as shown in Fig. 2, having the same form of (2), but this time

$$\varphi'_H = \varphi_{H,Q} + \omega,$$

where  $\varphi_{H,Q}$  is the phase of  $F_{H,Q}$  without the imaginary part of anomalous correction,  $\omega$  is the phase difference between  $F_{H,Q}''$  and  $F_{H,Q}$ , which equals  $\pi/2$  if all the anomalous-scattering atoms are of the same kind. In addition to the above two cases, it is interesting to notice that real-space enantiomorphous ambiguities in an electron density map or an  $E$  map, resulting from the determination of various kinds of small structures, also cause the enantiomorphous phase doublets. They all have the same form as (2) but with  $\varphi'_H = 2\pi\mathbf{H} \cdot \mathbf{r}_0$ , where  $\mathbf{r}_0$  is the positional vector of the pseudo inverse centre relating the true structure and its enantiomorph in real space.

To summarize, the problem of splitting any kind of enantiomorphous phase doublets can be converted

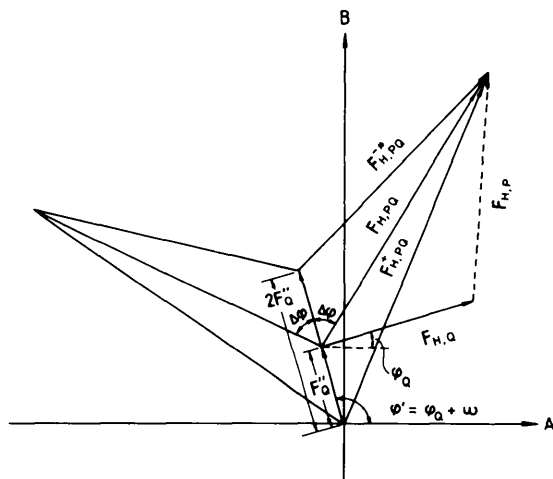


Fig. 2. Phase-vector diagram showing the enantiomorphous phase doublet arising from the OAS method.

into that of finding the signs for  $\Delta\varphi_H$  by making use of the known values of  $\varphi'_H$  and  $|\Delta\varphi_H|$ . Alternatively, defining

$$A'_H = |F_H| \cos \Delta\varphi_H \quad \text{and} \quad B'_H = |F_H| \sin \Delta\varphi_H,$$

the problem becomes that of finding the signs for  $B'_H$  from the known values of  $A'_H$  and  $|B'_H|$ .

### The phase-difference relation

We start from a modified Sayre equation (Fan Hai-fu, 1965*b*, 1975):

$$F_H = \frac{\theta_{H,p}}{V} \sum_{H'} F_{H'} F_{H-H'} - \sum_q \left( \frac{\theta_{H,p}}{\theta_{H,q}} - 1 \right) F_{H,q} \quad (6)$$

where  $\theta$  is an atomic form factor, the subscripts  $p$  and  $q$  denote the light and the heavy atoms, respectively. Replacing  $F_H$  by  $|F_H| \exp(i\varphi_H)$  and replacing  $\varphi_H$  by  $\varphi'_H + \Delta\varphi_H$ , (6) becomes

$$\begin{aligned} |F_H| \exp(i\Delta\varphi_H) &= \frac{\theta_{H,p}}{V} \sum_{H'} |F_{H'} F_{H-H'}| \\ &\times \exp[i(-\varphi'_H + \varphi'_{H'} + \varphi'_{H-H'} + \Delta\varphi_{H'} + \Delta\varphi_{H-H'})] \\ &- \sum_q \left( \frac{\theta_{H,p}}{\theta_{H,q}} - 1 \right) |F_{H,q}| \exp[i(\varphi_{H,q} - \varphi'_H)]. \end{aligned} \quad (7)$$

Taking the imaginary part of (7) and denoting  $-\varphi'_H + \varphi'_{H'} + \varphi'_{H-H'}$  by  $\Phi'_3$ , we have

$$\begin{aligned} |F_H| \sin \Delta\varphi_H &= \frac{\theta_{H,p}}{V} \left[ \sum_{H'} |F_{H'} F_{H-H'}| \sin(\Phi'_3 + \Delta\varphi_{H'} + \Delta\varphi_{H-H'}) \right] \\ &- \sum_q \left( \frac{\theta_{H,p}}{\theta_{H,q}} - 1 \right) |F_{H,q}| \sin(\varphi_{H,q} - \varphi'_H) \end{aligned} \quad (8)$$

or

$$\begin{aligned} B'_H &= \frac{\theta_{H,p}}{V} \left[ \sum_{H'} (A'_{H'} A'_{H-H'} - B'_{H'} B'_{H-H'}) \sin \Phi'_3 \right. \\ &\quad \left. + 2 \sum_{H'} A'_{H'} B'_{H-H'} \cos \Phi'_3 \right] \\ &- \sum_q \left( \frac{\theta_{H,p}}{\theta_{H,q}} - 1 \right) |F_{H,q}| \sin(\varphi_{H,q} - \varphi'_H). \end{aligned} \quad (9)$$

Equation (8) is the so-called phase-difference relation, while (9) may be regarded as a modified component relation. In the case of SIR with a centrosymmetric heavy-atom arrangement, (9) reduces to one of the original component relations (Fan Hai-fu, 1965*a*)

$$B_H = (2\theta_{H,p}/V) \sum_{H'} A_{H'} B_{H-H'} - \sum_q \left( \frac{\theta_{H,p}}{\theta_{H,q}} - 1 \right) B_{H,q} \quad (10)$$

while in the case of OAS with a centrosymmetric heavy-atom arrangement, (9) reduces to the other

original component relation

$$A_{\mathbf{H}} = (\theta_{\mathbf{H},p}/V) \sum_{\mathbf{H}'} (A_{\mathbf{H}'} A_{\mathbf{H}-\mathbf{H}'} - B_{\mathbf{H}'} B_{\mathbf{H}-\mathbf{H}'}) - \sum_q \left( \frac{\theta_{\mathbf{H},p}}{\theta_{\mathbf{H},q}} - 1 \right) A_{\mathbf{H},q}. \quad (11)$$

Here  $A_{\mathbf{H}}$  and  $B_{\mathbf{H}}$  are the real and imaginary components of the structure factor, respectively.

With (8) or (9) it is possible to derive the signs of  $\Delta\varphi_{\mathbf{H}}$  or  $B'_{\mathbf{H}}$  by a routine direct-method procedure, provided  $|\Delta\varphi_{\mathbf{H}}|$  or  $A'_{\mathbf{H}}$  and  $|B'_{\mathbf{H}}|$  are known *a priori*. In other words, (8) or (9) may in principle be used to break any kind of enantiomorphous phase doublets occurring in single-crystal structure analysis. Moreover, (8) may also be used to refine the values of  $|\Delta\varphi_{\mathbf{H}}|$  once a trial set of their signs, mostly correct, is known. This is important since the values of  $|\Delta\varphi_{\mathbf{H}}|$  from either isomorphous replacement or anomalous scattering include large errors.

If the normalized structure factors  $E_{\mathbf{H}}$  are to be used instead of  $F_{\mathbf{H}}$ , we have instead of (8) and (9) the following equations:

$$|E_{\mathbf{H}}| \sin \Delta\varphi_{\mathbf{H}} = \frac{(\varepsilon\sigma_2)^{1/2}}{Z_p V} \left[ \sum_{\mathbf{H}'} |E_{\mathbf{H}'} E_{\mathbf{H}-\mathbf{H}'}| \times \sin (\Phi'_3 + \Delta\varphi_{\mathbf{H}'} + \Delta\varphi_{\mathbf{H}-\mathbf{H}'}) \right] - \sum_q \left( \frac{Z_q}{Z_p} - 1 \right) |E_{\mathbf{H},q}| \sin (\varphi_{\mathbf{H},q} - \varphi'_{\mathbf{H}}) \quad (12)$$

$$\beta_{\mathbf{H}} = \frac{(\varepsilon\sigma_2)^{1/2}}{Z_p V} \left[ \sum_{\mathbf{H}'} (\alpha_{\mathbf{H}'} \alpha_{\mathbf{H}-\mathbf{H}'} - \beta_{\mathbf{H}'} \beta_{\mathbf{H}-\mathbf{H}'}) \sin \Phi'_3 + 2 \sum_{\mathbf{H}'} \alpha_{\mathbf{H}'} \beta_{\mathbf{H}-\mathbf{H}'} \cos \Phi'_3 \right] - \sum_q \left( \frac{Z_q}{Z_p} - 1 \right) |E_{\mathbf{H},q}| \sin (\varphi_{\mathbf{H},q} - \varphi'_{\mathbf{H}}). \quad (13)$$

Here

$$\alpha_{\mathbf{H}} = |E_{\mathbf{H}}| \cos \Delta\varphi_{\mathbf{H}}, \quad \beta_{\mathbf{H}} = |E_{\mathbf{H}}| \sin \Delta\varphi_{\mathbf{H}}.$$

### Incorporating phase-doublet information into the probability formula of triplet structure invariants

According to Cochran (1955), for a triplet structure invariant,  $\varphi_{\mathbf{H}} - \varphi_{\mathbf{H}'} - \varphi_{\mathbf{H}-\mathbf{H}'}$ , the conditional probability distribution of  $\varphi_{\mathbf{H}}$  given  $|E_{\mathbf{H}}|$ ,  $|E_{\mathbf{H}'}|$ ,  $|E_{\mathbf{H}-\mathbf{H}'}|$ ,  $\varphi_{\mathbf{H}'}$  and  $\varphi_{\mathbf{H}-\mathbf{H}'}$  is as follows:

$$P(\varphi_{\mathbf{H}}) = [2\pi I_0(K_{\mathbf{H},\mathbf{H}'})]^{-1} \times \exp [K_{\mathbf{H},\mathbf{H}'} \cos (\varphi_{\mathbf{H}} - \varphi_{\mathbf{H}'} - \varphi_{\mathbf{H}-\mathbf{H}'})], \quad (14)$$

where  $K_{\mathbf{H},\mathbf{H}'} = 2\sigma_3\sigma_2^{-3/2}|E_{\mathbf{H}}E_{\mathbf{H}'}E_{\mathbf{H}-\mathbf{H}'}|$ ,  $\sigma_n = \sum_{j=1}^N Z_j^n$ ,  $N$  is the number of atoms in the unit cell,  $I_0(K_{\mathbf{H},\mathbf{H}'})$

is the modified Bessel function of the first kind with  $K_{\mathbf{H},\mathbf{H}'}$  as argument. If for a given  $\mathbf{H}$  there are simultaneously  $t$  phase indications, the distribution becomes

$$P(\varphi_{\mathbf{H}}) = M \exp \left[ \sum_1^t K_{\mathbf{H},\mathbf{H}'} \cos (\varphi_{\mathbf{H}} - \varphi_{\mathbf{H}'} - \varphi_{\mathbf{H}-\mathbf{H}'}) \right], \quad (15)$$

where  $M$  is a normalizing factor. Replacing  $\varphi$  by  $\varphi' + \Delta\varphi$  in (15), we can obtain the conditional probability for  $\Delta\varphi_{\mathbf{H}}$  to have a positive or negative sign, when  $|\Delta\varphi_{\mathbf{H}}|$ ,  $\Delta\varphi_{\mathbf{H}'}$ ,  $\Delta\varphi_{\mathbf{H}-\mathbf{H}'}$ ,  $\varphi'_{\mathbf{H}}$ ,  $\varphi'_{\mathbf{H}'}$  and  $\varphi'_{\mathbf{H}-\mathbf{H}'}$  are fixed. The result is\*

$$P_{\pm}(\Delta\varphi_{\mathbf{H}}) = \frac{1}{2} \pm \frac{1}{2} \tanh \left[ 2\sigma_3\sigma_2^{-3/2}|E_{\mathbf{H}}| \sin |\Delta\varphi_{\mathbf{H}}| \times \sum_1^t |E_{\mathbf{H}'} E_{\mathbf{H}-\mathbf{H}'}| \sin (\Phi'_3 + \Delta\varphi_{\mathbf{H}'} + \Delta\varphi_{\mathbf{H}-\mathbf{H}'}) \right] \quad (16)$$

or identically

$$P_{\pm}(\beta_{\mathbf{H}}) = \frac{1}{2} \pm \frac{1}{2} \tanh \left\{ 2\sigma_3\sigma_2^{-3/2}|\beta_{\mathbf{H}}| \times \sum_1^t [(\alpha_{\mathbf{H}'} \alpha_{\mathbf{H}-\mathbf{H}'} - \beta_{\mathbf{H}'} \beta_{\mathbf{H}-\mathbf{H}'}) \sin \Phi'_3 + (\alpha_{\mathbf{H}'} \beta_{\mathbf{H}-\mathbf{H}'} + \beta_{\mathbf{H}'} \alpha_{\mathbf{H}-\mathbf{H}'}) \cos \Phi'_3] \right\}. \quad (17)$$

When dealing with structures having centrosymmetric heavy-atom arrangements, (17) reduces to either of the two probability formulae given by Karle (1966) according to whether SIR or OAS data are used. On the other hand, while the Karle formulae are not applicable to structures having non-centrosymmetric heavy-atom arrangements, (16) or (17) is very efficient for resolving phase ambiguities in this case.

### Practical application

#### 1. Derivation of initial sign of $\Delta\varphi_{\mathbf{H}}$

According to (16), signs of  $\Delta\varphi_{\mathbf{H}}$  can be determined through the signs of  $\Delta\varphi_{\mathbf{H}'}$  and  $\Delta\varphi_{\mathbf{H}-\mathbf{H}'}$ . In dealing with macromolecular structures, a large starting set is essential. This may be achieved by a random approach such as YZARC (Declercq, Germain & Woolfson, 1979) or RANTAN (Yao Jia-xing, 1981). However, it would be preferable to have some way of deriving more reliable initial signs for  $\Delta\varphi_{\mathbf{H}}$  without knowing the signs of  $\Delta\varphi_{\mathbf{H}'}$  and  $\Delta\varphi_{\mathbf{H}-\mathbf{H}'}$ . Consider the averaged value of  $\sin (\Phi'_3 + \Delta\varphi_{\mathbf{H}'} + \Delta\varphi_{\mathbf{H}-\mathbf{H}'})$  over the

\* See Appendix.

four possible sign combinations of  $\Delta\varphi_{\mathbf{H}'}$  and  $\Delta\varphi_{\mathbf{H}-\mathbf{H}'}$ ,

$$\begin{aligned} & \langle \sin(\Phi'_3 + \Delta\varphi_{\mathbf{H}'} + \Delta\varphi_{\mathbf{H}-\mathbf{H}'}) \rangle_{S_1, S_2} \\ &= \frac{1}{4} \sum_{S_1=-1}^{+1} \sum_{S_2=-1}^{+1} \sin(\Phi'_3 + \Delta\varphi_{\mathbf{H}'} + \Delta\varphi_{\mathbf{H}-\mathbf{H}'}) \\ &= \sin \Phi'_3 \cos \Delta\varphi_{\mathbf{H}'} \cos \Delta\varphi_{\mathbf{H}-\mathbf{H}'}, \end{aligned}$$

where  $S_1$  and  $S_2$  are the signs of  $\Delta\varphi_{\mathbf{H}'}$  and  $\Delta\varphi_{\mathbf{H}-\mathbf{H}'}$  respectively. If the averaged value,  $\sin \Phi'_3 \cos \Delta\varphi_{\mathbf{H}'} \cos \Delta\varphi_{\mathbf{H}-\mathbf{H}'}$ , is positive, the actual sign of  $\sin(\Phi'_3 + \Delta\varphi_{\mathbf{H}'} + \Delta\varphi_{\mathbf{H}-\mathbf{H}'})$  is more likely to be +1 than -1. The converse is also true. Hence, it is reasonable to expect that  $\sin(\Phi'_3 + \Delta\varphi_{\mathbf{H}'} + \Delta\varphi_{\mathbf{H}-\mathbf{H}'})$  will have the same sign as  $\sin \Phi'_3 \cos \Delta\varphi_{\mathbf{H}'} \cos \Delta\varphi_{\mathbf{H}-\mathbf{H}'}$ , at least when the absolute value of  $\sin \Phi'_3 \cos \Delta\varphi_{\mathbf{H}'} \cos \Delta\varphi_{\mathbf{H}-\mathbf{H}'}$  is large. So we can replace  $\sin(\Phi'_3 + \Delta\varphi_{\mathbf{H}'} + \Delta\varphi_{\mathbf{H}-\mathbf{H}'})$  by

$$Q \sin \Phi'_3 \cos \Delta\varphi_{\mathbf{H}'} \cos \Delta\varphi_{\mathbf{H}-\mathbf{H}'}$$

Here  $Q$  is a scaling factor. Equation (16) then becomes

$$\begin{aligned} P_{\pm}(\Delta\varphi_{\mathbf{H}}) &\sim \frac{1}{2} \pm \frac{1}{2} \tanh [2Q\sigma_3\sigma_2^{-3/2}|E_{\mathbf{H}}| \sin |\Delta\varphi_{\mathbf{H}}| \\ &\quad \times \sum_{\mathbf{H}'} |E_{\mathbf{H}'} E_{\mathbf{H}-\mathbf{H}'}| \sin \Phi'_3 \cos \Delta\varphi_{\mathbf{H}'} \cos \Delta\varphi_{\mathbf{H}-\mathbf{H}'}]. \quad (18) \end{aligned}$$

Equation (18) with  $Q=1$  can be derived more rigorously by making use of the concept of 'best phase relationship' (Fan Hai-fu, Han Fu-son & Qian Jin-zi, 1984).

The application of (18) to the SIR case may be considered as a reciprocal-space equivalent to the method of Ramachandran & Raman (1959) or Blow & Rossmann (1961). They used a special Fourier synthesis with SIR data to reveal the true structure. Wang (1981) showed that electron density modification by iterative Fourier and inverse Fourier calculations using SIR data can break the phase ambiguities and perform the phase refinement. This can be simulated in reciprocal space by using (12) with starting signs derived from (18). On the other hand, the application of (18) to the OAS case is similar to the 'resolved anomalous scattering method' of Hendrickson & Teeter (1981). However, this latter method separates a given phase doublet by the heavy-atom phase corresponding to only one reflection with the same index, while (18) uses a large number of heavy-atom phases with indices ranging over the whole reciprocal space within the resolution limit. It should be noticed that, in the case of SIR with a centrosymmetric heavy-atom arrangement, (18) cannot be used since  $\sin \Phi'_3$  will then always be zero. However, in this case (16) can give a large number of sign relationships between two  $\Delta\varphi_{\mathbf{H}}$ 's. This will be discussed in detail later.

## 2. Convergence mapping

In the usual convergence process (Germain *et al.*, 1970), the controlling factor is  $K_{\mathbf{H},\mathbf{H}'} =$

$2\sigma_3\sigma_2^{-3/2}|E_{\mathbf{H}}E_{\mathbf{H}'}E_{\mathbf{H}-\mathbf{H}'}|$ , which is a measure of strength of a single phase indication. In the absence of phase information, the corresponding measure of strength for a multiple phase indication is  $\langle \alpha_{\mathbf{H}}^2 \rangle^{1/2}$ , where

$$\begin{aligned} \langle \alpha_{\mathbf{H}}^2 \rangle &= \sum_{\mathbf{H}'} K_{\mathbf{H},\mathbf{H}'}^2 \\ &\quad + 2 \sum_{\substack{\mathbf{H}' \\ \mathbf{H}'' \\ \mathbf{H}' \neq \mathbf{H}''}} K_{\mathbf{H},\mathbf{H}'} K_{\mathbf{H},\mathbf{H}''} \frac{I_1(K_{\mathbf{H},\mathbf{H}'})}{I_0(K_{\mathbf{H},\mathbf{H}'})} \frac{I_1(K_{\mathbf{H},\mathbf{H}''})}{I_0(K_{\mathbf{H},\mathbf{H}''})}. \end{aligned}$$

When a direct method is used to resolve enantiomorphous phase ambiguities, partial phase information is available. In this case only the strong 'phase-difference relationships' are of importance. However, a strong  $\sum_2$  relationship is not necessarily a strong phase-difference relationship. Hence, we need a new controlling factor for the convergence process. In view of (16) the measure of strength may be

$$K'_{\mathbf{H},\mathbf{H}'} = K_{\mathbf{H},\mathbf{H}'} \sin |\Delta\varphi_{\mathbf{H}}| \sin (\Phi'_3 + \Delta\varphi_{\mathbf{H}'} + \Delta\varphi_{\mathbf{H}-\mathbf{H}'}) \quad (19)$$

for a single indication and

$$\alpha'_{\mathbf{H}} = \left| \sum_{\mathbf{H}'} K_{\mathbf{H},\mathbf{H}'} \sin (|\Delta\varphi_{\mathbf{H}}|) \sin (\Phi'_3 + \Delta\varphi_{\mathbf{H}'} + \Delta\varphi_{\mathbf{H}-\mathbf{H}'}) \right| \quad (20)$$

for a multiple indication.  $K'_{\mathbf{H},\mathbf{H}'}$  and  $\alpha'_{\mathbf{H}}$  are proportional to the probability that the indication will be true. In the absence of the information about the signs of  $\Delta\varphi_{\mathbf{H}'}$  and  $\Delta\varphi_{\mathbf{H}-\mathbf{H}'}$ , we can replace  $|\sin(\Phi'_3 + \Delta\varphi_{\mathbf{H}'} + \Delta\varphi_{\mathbf{H}-\mathbf{H}'})|$  by  $[\langle \sin^2(\Phi'_3 + \Delta\varphi_{\mathbf{H}'} + \Delta\varphi_{\mathbf{H}-\mathbf{H}'}) \rangle_{S_1, S_2}]^{1/2}$  where  $\langle \rangle_{S_1, S_2}$  is the average over the four possible sign combinations of  $\Delta\varphi_{\mathbf{H}'}$  and  $\Delta\varphi_{\mathbf{H}-\mathbf{H}'}$ . This is equal to

$$(\frac{1}{2} - \frac{1}{2} \cos 2\Phi'_3 \cos 2\Delta\varphi_{\mathbf{H}'} \cos 2\Delta\varphi_{\mathbf{H}-\mathbf{H}'})^{1/2}.$$

So we have

$$\begin{aligned} (K'_{\mathbf{H}-\mathbf{H}'})_{\text{est}} &= K_{\mathbf{H},\mathbf{H}'} \sin (|\Delta\varphi_{\mathbf{H}}|) (\frac{1}{2} - \frac{1}{2} \cos 2\Phi'_3 \\ &\quad \times \cos 2\Delta\varphi_{\mathbf{H}'} \cos 2\Delta\varphi_{\mathbf{H}-\mathbf{H}'})^{1/2} \quad (21) \end{aligned}$$

for a single indication and

$$(\alpha'_{\mathbf{H}})_{\text{est}} = \sum_{\mathbf{H}'} (K'_{\mathbf{H},\mathbf{H}'})_{\text{est}} \quad (22)$$

for a multiple indication. These should be used instead of  $K_{\mathbf{H},\mathbf{H}'}$  and  $(\alpha_{\mathbf{H}})_{\text{est}}$  in the convergence mapping.

## 3. Two-sign relationships

As mentioned above, in the case of SIR with a centrosymmetric heavy-atom arrangement, (18) cannot help in deriving the initial signs of  $\Delta\varphi_{\mathbf{H}}$ 's. In this

case, since  $\sin \Phi'_3 = 0$ , (16) becomes

$$P_{\pm}(\Delta\varphi_{\mathbf{H}}) = \frac{1}{2} \pm \frac{1}{2} \tanh \left[ 2\sigma_3 \sigma_2^{-3/2} |E_{\mathbf{H}}| \sin |\Delta\varphi_{\mathbf{H}}| \right. \\ \left. \times \sum_{\mathbf{H}'} |E_{\mathbf{H}'} E_{\mathbf{H}-\mathbf{H}'}| \cos \Phi'_3 \right. \\ \left. \times \sin (\Delta\varphi_{\mathbf{H}'} + \Delta\varphi_{\mathbf{H}-\mathbf{H}'}) \right]. \quad (23)$$

If, in addition, either  $\Delta\varphi_{\mathbf{H}'}$  is close to  $\pm\pi/2$  or  $\Delta\varphi_{\mathbf{H}-\mathbf{H}'}$  is close to 0 or  $\pi$  then (23) may reduce further to give

$$P_{\pm}(\Delta\varphi_{\mathbf{H}}) = \frac{1}{2} \pm \frac{1}{2} \tanh \left[ 2\sigma_3 \sigma_2^{-3/2} |E_{\mathbf{H}}| \sin |\Delta\varphi_{\mathbf{H}}| \right. \\ \left. \times \sum_{\mathbf{H}'} |E_{\mathbf{H}'} E_{\mathbf{H}-\mathbf{H}'}| \cos \Phi'_3 \right. \\ \left. \times \cos \Delta\varphi_{\mathbf{H}-\mathbf{H}'} \sin \Delta\varphi_{\mathbf{H}'} \right]. \quad (24)$$

If the summation over  $\mathbf{H}'$  in the above equation includes only those terms which are symmetrically related to  $\mathbf{H}'$ , then (24) gives the probability that  $\Delta\varphi_{\mathbf{H}}$  has sign the same as or opposite to  $\Delta\varphi_{\mathbf{H}'}$ , without the necessity of knowing any of the signs of  $\Delta\varphi_{\mathbf{H}-\mathbf{H}'}$ . A modified convergence procedure controlled by

$$|K''_{\mathbf{H},\mathbf{H}'}| = |K_{\mathbf{H},\mathbf{H}'}| \sin |\Delta\varphi_{\mathbf{H}}| \cos \Phi'_3 \\ \times \cos \Delta\varphi_{\mathbf{H}-\mathbf{H}'} \sin |\Delta\varphi_{\mathbf{H}'}|$$

and

$$\alpha''_{\mathbf{H}} = \left| \sum_{\mathbf{H}'} K''_{\mathbf{H},\mathbf{H}'} \right|$$

can be used to find out systematically the strongest two-sign relationships. After that, by giving a positive or negative sign to a  $\Delta\varphi_{\mathbf{H}}$  at the bottom of the convergence map, which is equivalent to fixing the enantiomorph, the phase ambiguities can be resolved easily. If, however, the indications at the bottom of the convergence map are not strong enough or whenever a weak link exists, then a multiple starting set of signs must be used and the phase ambiguities resolved by a *MULTAN*-like procedure.

The procedure suggested in this section may also be used in resolving enantiomorphous ambiguities arising from the determination of various kinds of small structures.

#### 4. Weighting schemes and figures of merit

All the weighting functions and figures of merit used in *MULTAN80* (Main *et al.*, 1980) can also work here with the corresponding  $\alpha_{\mathbf{H}}$ ,  $(\alpha_{\mathbf{H}})_{\text{est}}$  and  $(\alpha_{\mathbf{H}})_{\text{rand}}$  given as follows:

$$\alpha_{\mathbf{H}} = \alpha'_{\mathbf{H}} \\ (\alpha_{\mathbf{H}})_{\text{est}} = (\alpha'_{\mathbf{H}})_{\text{est}} \\ (\alpha_{\mathbf{H}})_{\text{rand}} = 0,$$

Table 1. *The correlation between the calculated probability and the percentage of reflections with the signs of  $\Delta\varphi_{\mathbf{H}}$  correctly determined*

$P$ :  $|P_+ - \frac{1}{2}| + \frac{1}{2}$ ; Ngr: Number of reflections in the group; %: Percentage of reflections with signs correctly determined.

$P \times 100$	Results from (16)		Results from (18)	
	Ngr	%	Ngr	%
98-100	426	98.8	57	100.0
92-98	128	94.6	83	94.0
85-95	151	91.4	138	89.9
70-90	192	81.2	297	78.8
60-80	176	65.3	356	65.5
50-70	253	61.7	533	52.2

where  $\alpha'_{\mathbf{H}}$  and  $(\alpha'_{\mathbf{H}})_{\text{est}}$  are given by (20) and (22), respectively. In addition, it would be better to replace the  $E_{\mathbf{H}}$  values by  $\beta_{\mathbf{H}} = |E_{\mathbf{H}}| \sin \Delta\varphi_{\mathbf{H}}$  in the  $\Psi_0$  and  $R$ (Karle) figures of merit.

#### Test of the probability formulae

Formulae (16) and (18) have been tested with the error-free SIR data from the protein insulin, molecular weight  $\sim 12\,000$ , and its Pb derivative. Crystals of insulin belong to space group *R3* with unit-cell parameters  $a = 82.5$ ,  $c = 34.0$ ,  $\gamma = 120^\circ$  and  $Z = 9$ . The data were calculated from the known atomic parameters. There are 6371 independent reflections within resolution limit of  $1.9 \text{ \AA}$ . In the test calculation, 1000 largest  $E$ 's of the native protein together with the corresponding  $\Delta\varphi_{\mathbf{H}}$  were selected and 60 000 of the total 75 568  $\sum_2$  relationships were involved. The results are summarized in Tables 1 and 2. In Table 1 the reflections are grouped according to the calculated probability. As can be seen the calculated probabilities,  $P$ , from either (16) or (18) are in good agreement with the percentages of reflections having signs of  $\Delta\varphi_{\mathbf{H}}$  correctly determined by the corresponding probability formula. This means that the probabilities calculated from either (16) or (18) are reliable quantitatively. In Table 2 the reflections are cumulated in 20 groups. It shows that, with (18), it is possible to obtain a very large starting set of good quality without any preliminary knowledge of the sign of  $\Delta\varphi_{\mathbf{H}}$ . Furthermore, such a starting set can in turn be improved considerably by making use of (16).

Tests with experimental data of known protein structures are now in progress. The results will be published in due course.

The authors are grateful to Drs G. Dodson and E. Dodson for the use of their insulin data and to Professor M. M. Woolfson for his helpful comments and for kindly correcting the manuscript.

Table 2. Errors in the estimation of signs of  $\Delta\varphi_{\mathbf{H}}$  with respect to the cumulative probabilities

I: Results from (18) using  $E_{\mathbf{H}}$  and  $|\Delta\varphi_{\mathbf{H}}|$  as input; II: Results from (16) using the results from (18) as input; \* III: Results from (16) calculated with the signs of  $\Delta\varphi_{\mathbf{H}}$  derived from the known structure; Gnr: Group number; Ngr: Number of reflections in the group; P: Minimum value of  $|P_+ - \frac{1}{2}| + \frac{1}{2}$  in the group; Nwr: Number of reflections with wrong signs; ER: Averaged error of phases in degrees.

Gnr	Ngr	I			II			III		
		$P \times 100$	Nwr	ER(°)	$P \times 100$	Nwr	ER(°)	$P \times 100$	Nwr	ER(°)
1	50	98.3	0	0	100.0	0	0	100.0	0	0
2	100	95.2	2	2	99.7	2	2	100.0	0	0
3	150	91.4	5	4	99.2	3	2	100.0	0	0
4	200	88.2	11	7	98.3	5	2	99.9	0	0
5	250	84.4	17	7	97.0	10	4	99.8	1	0
6	300	80.9	32	12	95.5	12	4	99.6	1	0
7	350	77.4	44	14	92.7	19	5	99.1	2	0
8	400	74.6	55	15	89.7	27	7	98.2	5	1
9	450	71.0	65	16	85.9	35	8	96.7	7	1
10	500	67.6	80	17	82.5	49	10	94.0	10	2
11	550	65.3	92	18	78.5	63	11	90.6	15	2
12	600	63.2	116	20	74.8	84	13	87.2	19	3
13	650	60.7	146	22	70.0	102	15	82.0	24	3
14	700	58.4	174	23	66.2	123	16	75.6	37	4
15	750	56.5	197	23	62.8	149	16	69.3	54	5
16	800	55.3	220	24	59.4	173	17	65.2	71	6
17	850	53.7	247	24	57.3	192	17	59.7	89	6
18	900	52.2	268	24	54.5	216	18	55.4	106	7
19	950	51.0	294	24	52.0	243	18	52.5	126	7
20	998	50.0	324	25	50.0	262	18	50.0	150	7

\* See Fan Hai-fu *et al.* (1984) for further details.

## APPENDIX

While (16) comes from Cochran's distribution (14), the *a priori* knowledge of  $E_{\mathbf{H},Q}$  can lead to a distribution different from (14). Taking this into account we can use instead of (14) a combined distribution which is the product of (14) and the distribution of Sim (1959). This results in

$$P(\Delta\varphi_{\mathbf{H}}) = [2\pi I_0(\alpha)]^{-1} \exp[\alpha \cos(\Delta\varphi_{\mathbf{H}} - \beta)] \quad (14')$$

with

$$\alpha^2 = \left[ \sum_{\mathbf{H}'} K_{\mathbf{H}\mathbf{H}'} \cos(\Phi_3' + \Delta\varphi_{\mathbf{H}'} + \Delta\varphi_{\mathbf{H}-\mathbf{H}'}) + x \cos \omega \right]^2 + \left[ \sum_{\mathbf{H}'} K_{\mathbf{H}\mathbf{H}'} \sin(\Phi_3' + \Delta\varphi_{\mathbf{H}'} + \Delta\varphi_{\mathbf{H}-\mathbf{H}'}) - x \sin \omega \right]^2$$

and

$$\tan \beta = \frac{\sum_{\mathbf{H}'} K_{\mathbf{H}\mathbf{H}'} \sin(\Phi_3' + \Delta\varphi_{\mathbf{H}'} + \Delta\varphi_{\mathbf{H}-\mathbf{H}'}) - x \sin \omega}{\sum_{\mathbf{H}'} K_{\mathbf{H}\mathbf{H}'} \cos(\Phi_3' + \Delta\varphi_{\mathbf{H}'} + \Delta\varphi_{\mathbf{H}-\mathbf{H}'}) + x \cos \omega},$$

where

$$x = 2|E_{\mathbf{H}} E_{\mathbf{H},Q}| / \sigma_p^2, \\ \sigma_p^2 = \sum_{p=1}^{N_p} Z_p^2 / \sigma_2,$$

$p$  denotes the atoms of unknown positions,

$$\omega = 0,$$

except for the OAS case. By maximizing (14') we have

$$\tan(\Delta\varphi_{\mathbf{H}}) = \left[ \sum_{\mathbf{H}'} |E_{\mathbf{H}'} E_{\mathbf{H}-\mathbf{H}'}| \sin(\Phi_3' + \Delta\varphi_{\mathbf{H}'} + \Delta\varphi_{\mathbf{H}-\mathbf{H}'}) - \sigma_3^{-1} \sigma_2^{3/2} \sigma_p^{-2} |E_{\mathbf{H},Q}| \sin \omega \right] \times \left[ \sum_{\mathbf{H}'} |E_{\mathbf{H}'} E_{\mathbf{H}-\mathbf{H}'}| \cos(\Phi_3' + \Delta\varphi_{\mathbf{H}'} + \Delta\varphi_{\mathbf{H}-\mathbf{H}'}) + \sigma_3^{-1} \sigma_2^{3/2} \sigma_p^{-2} |E_{\mathbf{H},Q}| \cos \omega \right]^{-1}. \quad (15')$$

We can also obtain from (14') that

$$P_+(\Delta\varphi_{\mathbf{H}}) = \frac{1}{2} + \frac{1}{2} \tanh \left\{ 2\sigma_3 \sigma_2^{-3/2} |E_{\mathbf{H}} \sin \Delta\varphi_{\mathbf{H}}| \times \sum_{\mathbf{H}'} |E_{\mathbf{H}'} E_{\mathbf{H}-\mathbf{H}'}| \sin(\Phi_3' + \Delta\varphi_{\mathbf{H}'} + \Delta\varphi_{\mathbf{H}-\mathbf{H}'}) - \sigma_p^{-2} |E_{\mathbf{H},Q}| \sin \omega \right\}. \quad (16')$$

In the SIR case, since  $\omega = 0$ , (16') is the same as (16). In the OAS case, as far as macromolecular structures are concerned, the term  $\sigma_p^{-2} |E_{\mathbf{H},Q}| \sin \omega$  will be comparatively small and there will be little difference between (16) and (16'). On the other hand, when we are trying to derive a complete structure from its known part, (15') and (16') will then be very helpful. This will be discussed in detail in a forthcoming paper.

## References

- BLOW, D. M. & ROSSMANN, M. G. (1961). *Acta Cryst.* **14**, 1195–1202.
- BLUNDELL, T. L. & JOHNSON, L. N. (1976). *Protein Crystallography*, pp. 153–158, 177–180. London: Academic Press.
- COCHRAN, W. (1955). *Acta Cryst.* **8**, 473–478.
- COULTER, C. L. (1965). *J. Mol. Biol.* **12**, 292–295.
- DECLERCQ, J.-P., GERMAIN, G. & WOOLFSON, M. M. (1979). *Acta Cryst.* **A35**, 622–626.
- FAN HAI-FU (1965a). *Acta Phys. Sin.* **21**, 1114–1118.
- FAN HAI-FU (1965b). *Acta Phys. Sin.* **21**, 1105–1113.
- FAN HAI-FU (1975). *Acta Phys. Sin.* **24**, 57–60.
- FAN HAI-FU, HAN FU-SON & QIAN JIN-ZI (1984). *Acta Cryst.* **A40**, 495–498.
- GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1970). *Acta Cryst.* **B26**, 274–285.
- HAUPTMAN, H. (1982a). *Acta Cryst.* **A38**, 289–294.
- HAUPTMAN, H. (1982b). *Acta Cryst.* **A38**, 632–641.
- HENDRICKSON, W. A. & TEETER, M. M. (1981). *Nature (London)*, **290**, 107–113.
- KARLE, J. (1966). *Acta Cryst.* **21**, 273–276.
- MAIN, P., FISKE, S. J., HULL, S. E., LESSINGER, L., GERMAIN, G., DECLERCQ, J.-P. & WOOLFSON, M. M. (1980). *A System of Computer Programs for the Automatic Solution of Crystal Structures from X-ray Diffraction Data*. Universities of York, England, and Louvain, Belgium.
- RAMACHANDRAN, G. N. & RAMAN, S. (1959). *Acta Cryst.* **12**, 957–964.
- SIM, G. A. (1959). *Acta Cryst.* **12**, 813–815.
- WANG, B. C. (1981). *Acta Cryst.* **A37**, C11.
- YAO JIA-XING (1981). *Acta Cryst.* **A37**, 642–644.

*Acta Cryst.* (1984). **A40**, 495–498

## Combining Direct Methods with Isomorphous Replacement or Anomalous Scattering Data. II. The Treatment of Errors

BY FAN HAI-FU, HAN FU-SON AND QIAN JIN-ZI

*Institute of Physics, Chinese Academy of Sciences, Beijing, China*

(Received 13 October 1983; accepted 20 March 1984)

### Abstract

The error treatment given by Blow & Crick [*Acta Cryst.* (1959). **12**, 794–802] in the isomorphous replacement method can be readily introduced into direct methods when they are used to handle the single isomorphous replacement (SIR) or the one-wavelength anomalous scattering (OAS) data. The 'best phase relationship' is defined similarly to the 'best Fourier'. Expressions of the 'best phase' and the 'figure of merit' for individual reflection have been derived for the SIR or OAS case. These enable initial signs to be obtained for a set of  $\Delta\varphi_{\mathbf{H}}$  without knowing the sign of any one  $\Delta\varphi_{\mathbf{H}}$  in advance. Finally, a weighting scheme is proposed for the refinement of signs and magnitudes of  $\Delta\varphi_{\mathbf{H}}$ .

### The best phase relationship

Following Blow & Crick (1959), we consider the error in a single triplet relationship arising from the error of only one reflection (say  $\mathbf{E}_{\mathbf{H}}$ ). We write

$$\Delta\mathbf{E}_{\mathbf{H}} = K'(\mathbf{E}_{\mathbf{H}}^e - \mathbf{E}_{\mathbf{H}}^t)\mathbf{E}_{\mathbf{H}-\mathbf{H}}, \quad (1)$$

where  $K'$  is a constant,  $\mathbf{E}_{\mathbf{H}}^e$  is the value of  $\mathbf{E}_{\mathbf{H}}$  employed in the calculation and  $\mathbf{E}_{\mathbf{H}}^t$  is the true value of  $\mathbf{E}_{\mathbf{H}}$ . From (1),

$$\Delta\mathbf{E}_{\mathbf{H}}^2 = K'^2|\mathbf{E}_{\mathbf{H}-\mathbf{H}}|^2|\mathbf{E}_{\mathbf{H}}^e - \mathbf{E}_{\mathbf{H}}^t|^2 = K|\mathbf{E}_{\mathbf{H}}^e - \mathbf{E}_{\mathbf{H}}^t|^2, \quad (2)$$

where  $K$  is also a constant. Now a best phase relationship is defined as that which leads to a minimum value of  $\Delta\mathbf{E}_{\mathbf{H}}^2$ . In practice,  $\mathbf{E}_{\mathbf{H}}^e$  can only be expressed in the form of a probability distribution. Accordingly,

$$\Delta\mathbf{E}_{\mathbf{H}}^2 = K \int |\mathbf{E}_{\mathbf{H}}^e - \mathbf{E}_{\mathbf{H}}|^2 P(\mathbf{E}_{\mathbf{H}}) d\mathbf{E}_{\mathbf{H}}. \quad (3)$$

Let  $\partial(\Delta\mathbf{E}_{\mathbf{H}}^2)/\partial\mathbf{E}_{\mathbf{H}}^e = 0$ , then

$$\mathbf{E}_{\mathbf{H}}^e = \int \mathbf{E}_{\mathbf{H}} P(\mathbf{E}_{\mathbf{H}}) d\mathbf{E}_{\mathbf{H}} = \mathbf{E}_{\mathbf{H}}^{\text{best}}. \quad (4)$$

$\mathbf{E}_{\mathbf{H}}$  can be expressed by  $E_{\mathbf{H}} \exp(i\alpha_{\mathbf{H}})$ , where  $E_{\mathbf{H}}$  can be derived from the experimental data. Since the error to be considered in  $\mathbf{E}_{\mathbf{H}}$  is the phase error, (4) can be written as

$$\mathbf{E}_{\mathbf{H}}^{\text{best}} = E_{\mathbf{H}} \int \exp(i\alpha_{\mathbf{H}}) P(\alpha_{\mathbf{H}}) d\alpha_{\mathbf{H}}. \quad (5)$$

Let

$$\mathbf{m}_{\mathbf{H}} = m_{\mathbf{H}} \exp(i\alpha_{\mathbf{H}}^{\text{best}}) = \int \exp(i\alpha_{\mathbf{H}}) P(\alpha_{\mathbf{H}}) d\alpha_{\mathbf{H}}, \quad (6)$$

(5) becomes

$$\mathbf{E}_{\mathbf{H}}^{\text{best}} = m_{\mathbf{H}} E_{\mathbf{H}} \exp(i\alpha_{\mathbf{H}}^{\text{best}}). \quad (7)$$

Here  $\alpha_{\mathbf{H}}^{\text{best}}$  and  $m_{\mathbf{H}}$  are known as the best phase and the figure of merit in protein crystallography. They