

We wish to thank L. Heaton and E. S. Fisher for discussion, H. W. Knott for cutting crystal 2, and R. L. Hitterman for his invaluable assistance in the data collection. We are grateful to C. M. Vincent for help in writing a *SPEAKEASY* program for the least-squares refinement.

### References

- ANDRES, K. (1968). *Phys. Rev.* **170**, 614.  
 BARRETT, C. S., MUELLER, M. H. & HITTERMAN, R. L. (1963). *Phys. Rev.* **129**, 625.  
 COPPENS, P. & HAMILTON, W. C. (1968). American Crystallographic Association Meeting, Buffalo, New York. Abstract G2.  
 FISHER, E. S. & DEVER, D. (1968). *Phys. Rev.* **170**, 607.  
 HEATON, L., MUELLER, M. H. & HITTERMAN, R. L. (1968). American Crystallographic Association Meeting, Buffalo, New York. Abstract H3.  
 HOUGH, A., MARPLES, J. A. C., MORTIMER, M. J. & LEE, J. A. (1968). *Phys. Letters*, **27A**, 222.  
 ROSS, J. W. & LAM, D. J. (1968). *Phys. Rev.* **165**, 617.  
 WILLIS, B. T. M. (1962). *Pile Neutron Research in Physics*, p. 455. Vienna: I.A.E.A.  
 ZACHARIASEN, W. H. (1967). *Acta Cryst.* **23**, 558.  
 ZACHARIASEN, W. H. (1968a). *Acta Cryst.* **A24**, 212.  
 ZACHARIASEN, W. H. (1968b). *Acta Cryst.* **A24**, 324.

*Acta Cryst.* (1970). **B26**, 136

## Representation of Phase Probability Distributions for Simplified Combination of Independent Phase Information

BY WAYNE A. HENDRICKSON\* AND EATON E. LATTMAN

*The Johns Hopkins University, Thomas C. Jenkins Department of Biophysics, Baltimore, Maryland 21218, U.S.A.*

(Received 27 January 1969)

The phase probability distributions associated with several types of phase information of possible use in protein structure analysis have been cast in the simplified representation,

$$P(\alpha) = N \exp(A \cos \alpha + B \sin \alpha + C \cos 2\alpha + D \sin 2\alpha).$$

This formulation permits the combination of independent phase information from different sources by simple addition of the constant coefficients  $A$ ,  $B$ ,  $C$  and  $D$  which encode the phase information. Also, use of this form for the phase probability expedites numerical integration of the centroid phase integrals and makes analytic integration possible. In order to achieve the simplified representation for the isomorphous replacement method it was necessary to reformulate the error analysis for that method. A computational comparison of phase determination by various methods for isomorphous replacement substantiates the validity of the new approach.

### Introduction

Crystal structure analyses usually proceed, by one means or another, to a trial structure which can then be refined to a best match between the calculated and observed intensities. This is not possible in protein structure analysis for the structural complexity and practical resolution limits make a satisfactory trial structure unobtainable in the early stages. Structural information must be obtained by direct interpretation of the electron density map. It is most important, then, that this Fourier synthesis be as accurate as possible.

Blow & Crick (1959) have shown that the Fourier synthesis with minimal mean-square error is achieved by the use of coefficients defined by the centroid of the structure-factor probability distribution. In practice

this has been reduced to the centroid of the phase probability distribution, assuming the amplitude to be fixed. The resulting Fourier synthesis has as coefficients

$$\xi = F_P \frac{\int_0^{2\pi} \exp(i\alpha) P(\alpha) d\alpha}{\int_0^{2\pi} P(\alpha) d\alpha} \quad (1)$$

where  $F_P$  is a structure factor amplitude from the protein crystal and  $P(\alpha) d\alpha$  is the probability that its phase angle  $\alpha$  be between  $\alpha$  and  $\alpha + d\alpha$ . This formula has been implemented for the isomorphous replacement method by determining the phase probability distribution from the lack of closure errors in phase determinations (Blow & Crick, 1959; Dickerson, Kendrew & Strandberg, 1961; Cullis, Muirhead, Perutz, Rossmann & North, 1961). The extension to phase information from anomalous scattering has been made in a similar way by North (1965) and Matthews (1966).

\* Present address: Laboratory for the Structure of Matter, U.S. Naval Research Laboratory, Washington, D.C. 20390, U.S.A.

The phase probability distributions formulated in this way are rather complicated transcendental functions. Storage of phase information other than in the most elementary form is space consuming, combination of phase information from various sources is difficult and evaluation of the integrals in (1) must be done numerically. By using an approximation for the lack of closure error, Rossmann & Blow (1961) were able to represent the phase probability distribution for the isomorphous replacement method in a simplified form. This permitted easy storage of phase information and ready incorporation of information derived from additional isomorphous replacements or from the 'heavy-atom' method. They also proposed an evaluation of the integrals by a three-dimensional table look-up.

The developments below modify and extend the ideas of Rossmann & Blow. By redefinition of the error in the isomorphous replacement method a simplified representation of the phase probability,

$$P(\alpha) = N \exp (A \cos \alpha + B \sin \alpha + C \cos 2\alpha + D \sin 2\alpha),$$

is found without approximation and its validity is established by computational tests.  $A$ ,  $B$ ,  $C$  and  $D$  are constant coefficients encoding the phase information and  $N$  is a non-essential normalization factor. Representations of this form are also derived for phase information from anomalous scattering, direct methods, the partial structure method and non-crystallographic symmetry. Since the total probability is the product of the independent contributing probabilities, phase information from various sources can be combined by simple addition of exponential coefficients. Finally, an analytic evaluation of the phase integrals of (1) is presented.

### Isomorphous replacement phase information

In an isomorphous replacement experiment the structure factor amplitudes,  $F_P$  and  $F_{H_j}$ , of the protein and the  $j$ th heavy-atom derivative are measured. The heavy-atom parameters are found and from them the heavy-atom contribution,  $\mathbf{f}_j = f_j \exp(i\varphi_j) = a_j + i b_j$ , to a structure factor is calculated. One must then consider the probability of a protein phase angle  $\alpha$  being consistent with these observations. A formally correct solution is obtained when  $\mathbf{F}_P + \mathbf{f}_j = \mathbf{F}_{H_j}$ . In practice, of course, there are uncertainties in these values which in general prevent identity even for the correct solution. Consequently an evaluation of errors must be made before a proper assessment of phase probability can be achieved.

The analysis of Blow & Crick treats the lack of closure error in the above equation as residing in the derivative amplitude  $F_{H_j}$ . Thus

$$|\mathbf{F}_P + \mathbf{f}_j| = F_{H_j} + \varepsilon'_j$$

which yields

$$F_P^2 + f_j^2 + 2F_P f_j \cos(\alpha - \varphi_j) = F_{H_j}^2 + \varepsilon_j'^2 + 2F_{H_j} \varepsilon'_j. \quad (2)$$

To obtain a simplified representation of this error,

Rossmann & Blow considered the  $\varepsilon_j'^2$  term in (2) to be negligible and proceeded from the approximation

$$F_P^2 + f_j^2 + 2F_P f_j \cos(\alpha - \varphi_j) = F_{H_j}^2 + 2F_{H_j} \varepsilon'_j. \quad (3)$$

An alternative to the Blow & Crick analysis, and *a priori* a no less legitimate procedure, is to take the error to formally lie in the derivative intensity  $F_{H_j}^2$ . Of course, this is really a lumped error with contributions from many sources, and its relationship to the substantive errors in individual measured and calculated quantities is readily demonstrated. The error equation which obtains with this lumped error is

$$|\mathbf{F}_P + \mathbf{f}_j|^2 = F_{H_j}^2 + \varepsilon_j$$

or

$$F_P^2 + f_j^2 + 2F_P f_j \cos(\alpha - \varphi_j) = F_{H_j}^2 + \varepsilon_j. \quad (4)$$

Expression (4), like (3), is in a form from which a simplified representation can be derived, but unlike (3) it is not an approximation.

Assuming a Gaussian distribution of errors,\* the probability distribution for isomorphous replacement phase information is

$$P_{\text{iso}_j}(\alpha) = N_j \exp(-\varepsilon_j^2(\alpha)/2E_j^2) \quad (5)$$

where  $E_j$  is the standard deviation of the errors and  $N_j$  is a normalizing factor.  $E_j$  can be estimated from the centrosymmetric zones and refined from the set of  $\varepsilon_j$  at preliminary phase angles, just as can the standard deviation  $E'_j$  associated with the distribution of  $\varepsilon'_j$ . However, since  $\varepsilon_j = \varepsilon'_j(\varepsilon'_j + 2F_{H_j})$  and  $E'_j$  has been shown to be only slightly dependent on the measured intensities (Blow & Crick, 1959),  $E_j$  will necessarily be a function of structure factor amplitude.

From the definition of  $\varepsilon_j$  (4), the exponent in the probability distribution (5) may be evaluated. Upon squaring the error,

$$\begin{aligned} \varepsilon_j^2(\alpha) = & (F_P^2 + f_j^2 - F_{H_j}^2)^2 + 4(F_P^2 + f_j^2 - F_{H_j}^2)F_P f_j \cos(\alpha - \varphi_j) \\ & + 4F_P^2 f_j^2 \cos^2(\alpha - \varphi_j). \end{aligned}$$

Then by utilizing the relationships

$$\cos^2(\alpha - \varphi_j) = \frac{1}{2}[1 + \cos 2(\alpha - \varphi_j)],$$

$$f_j \cos(\alpha - \varphi_j) = a_j \cos \alpha + b_j \sin \alpha$$

and

$$f_j^2 \cos 2(\alpha - \varphi_j) = (a_j^2 - b_j^2) \cos 2\alpha + 2a_j b_j \sin 2\alpha$$

this can be written as

\* It is worth noting that while in principle if the distribution of errors in  $F_{H_j}^2$  is Gaussian then that of errors in  $F_{H_j}$  is not, and *vice versa*, in fact they are both nearly Gaussian. The experimentally determined distributions of  $\varepsilon_j$  and  $\varepsilon'_j$  (computed from data of Green, Ingram & Perutz, 1954) both approximately follow Gaussian distributions. Further, if the distribution of errors in  $F_{H_j}^2$  is assumed to be Gaussian and the distribution of errors in  $F_{H_j}$  is derived therefrom, the Taylor series expansions of the two distribution functions have low-order coefficients of about the same relative values.

$$\begin{aligned}\varepsilon_j^2(\alpha) = & (F_P^2 + f_j^2 - F_{H_j}^2)^2 + 2F_P^2 f_j^2 \\ & + 4(F_P^2 + f_j^2 - F_{H_j}^2)F_P a_j \cos \alpha \\ & + 4(F_P^2 + f_j^2 - F_{H_j}^2)F_P b_j \sin \alpha \\ & + 2F_P^2(a_j^2 - b_j^2) \cos 2\alpha \\ & + 4F_P^2 a_j b_j \sin 2\alpha.\end{aligned}$$

Now the exponent of (5) can be put into the simplified form,

$$-\varepsilon_j^2(\alpha)/2E_j^2 = K_{\text{iso}_j} + A_{\text{iso}_j} \cos \alpha + B_{\text{iso}_j} \sin \alpha + C_{\text{iso}_j} \cos 2\alpha + D_{\text{iso}_j} \sin 2\alpha \quad (6)$$

where

$$\begin{aligned}K_{\text{iso}_j} &= -\frac{(F_P^2 + f_j^2 - F_{H_j}^2)^2 + 2F_P^2 f_j^2}{2E_j^2}; \\ A_{\text{iso}_j} &= -\frac{2(F_P^2 + f_j^2 - F_{H_j}^2)F_P}{E_j^2} a_j; \\ B_{\text{iso}_j} &= -\frac{2(F_P^2 + f_j^2 - F_{H_j}^2)F_P}{E_j^2} b_j; \\ C_{\text{iso}_j} &= -\frac{F_P^2}{E_j^2} (a_j^2 - b_j^2) \text{ and } D_{\text{iso}_j} = -\frac{2F_P^2}{E_j^2} (a_j b_j).\end{aligned}$$

This result is similar to that given in equation (4) of Rossmann & Blow (1961). In fact the coefficient  $c_j$  of that equation is related to the coefficients above by

$$A_{\text{iso}_j} \cos \alpha + B_{\text{iso}_j} \sin \alpha = (4F_{H_j}^2 E_j^2 / E_j^2) c_j \cos(\alpha - \varphi_j)$$

and  $d_j$  is likewise related to  $C_{\text{iso}_j}$  and  $D_{\text{iso}_j}$ . The results differ because of the formulations of error as defined in (3) and (4) above and in the choice of representation. The representation of equation (6) separates the phase-information parameters from the functions of phase angle  $\alpha$  to facilitate combinations of phase information and expedite numerical integration of (1). These matters are discussed below.

### Anomalous scattering phase information

When measurements are made on the Friedel mates,  $F_{H_k}^+$  and  $F_{H_k}^-$ , of an anomalously scattering heavy-atom derivative,  $k$ , additional phase information is obtained. North (1965) and Matthews (1966) demonstrated that the lack of closure error in the anomalous scattering experiment is, for an assumed protein phase angle  $\alpha$ ,

$$\varepsilon_k(\alpha) = \Delta H_k + \frac{2F_P \delta_k}{F_c} \sin(\varphi_k - \alpha). \quad (7)$$

The anomalous scattering difference is  $\Delta H_k = (F_{H_k}^+ - F_{H_k}^-)$ . The contribution to the derivative structure factor from the real parts of heavy-atom scattering factors is  $f_k$  with phase angle  $\varphi_k$  and the contribution from the imaginary parts has amplitude  $\delta_k$ .  $F_c$  is the amplitude of the vector sum  $\mathbf{F}_P + \mathbf{f}_k$ . This is replaced by the constant value  $F_{H_k} = \frac{1}{2}(F_{H_k}^+ + F_{H_k}^-)$  as suggested by

North. Matthews found that while the use of  $F_c$  is physically more acceptable, 'calculations using  $F_H$  and  $F_c$  do not lead to widely divergent results.'

The standard deviation  $E_k$  of the distribution of errors can be estimated as discussed by North or found directly from the  $\varepsilon_k$  at preliminary phase angles. Having this information the probability distribution of  $\alpha$  is given by

$$P_{\text{ano}_k}(\alpha) = N_k \exp(-\varepsilon_k^2(\alpha)/2E_k^2). \quad (8)$$

Evaluation of  $\varepsilon_k$  proceeds as in the isomorphous replacement case. First, the squared error is

$$\varepsilon_k^2(\alpha) = \Delta H_k^2 + \frac{4F_P \delta_k \Delta H_k}{F_{H_k}} \sin(\varphi_k - \alpha) + \frac{4F_P^2 \delta_k^2}{F_{H_k}^2} \times \sin^2(\varphi_k - \alpha).$$

Then using the relationships

$$\sin^2(\varphi_k - \alpha) = \frac{1}{2}[1 - \cos 2(\varphi_k - \alpha)],$$

$$\sin(\varphi_k - \alpha) = \frac{1}{f_k} [b_k \cos \alpha - a_k \sin \alpha]$$

and

$$\cos 2(\varphi_k - \alpha) = \frac{1}{f_k^2} [(a_k^2 - b_k^2) \cos 2\alpha + 2a_k b_k \sin 2\alpha],$$

this becomes

$$\begin{aligned}\varepsilon_k^2(\alpha) = & \Delta H_k^2 + \frac{2F_P^2 \delta_k^2}{F_{H_k}^2} + \frac{4F_P \delta_k \Delta H_k}{f_k F_{H_k}} b_k \cos \alpha \\ & - \frac{4F_P \delta_k \Delta H_k}{f_k F_{H_k}} a_k \sin \alpha \\ & - \frac{2F_P^2 \delta_k^2}{f_k^2 F_{H_k}^2} (a_k^2 - b_k^2) \cos 2\alpha \\ & - \frac{4F_P^2 \delta_k^2}{f_k^2 F_{H_k}^2} a_k b_k \sin 2\alpha.\end{aligned}$$

Thus the simplified form of the exponent in (8) is

$$-\varepsilon_k^2(\alpha)/2E_k^2 = K_{\text{ano}_k} + A_{\text{ano}_k} \cos \alpha + B_{\text{ano}_k} \sin \alpha + C_{\text{ano}_k} \cos 2\alpha + D_{\text{ano}_k} \sin 2\alpha \quad (9)$$

where

$$K_{\text{ano}_k} = -\frac{\Delta H_k^2 + (2F_P^2 \delta_k^2 / F_{H_k}^2)}{2E_k^2},$$

$$A_{\text{ano}_k} = -\frac{2F_P \delta_k \Delta H_k}{f_k F_{H_k} E_k^2} b_k,$$

$$B_{\text{ano}_k} = \frac{2F_P \delta_k \Delta H_k}{f_k F_{H_k} E_k^2} a_k,$$

$$C_{\text{ano}_k} = \frac{F_P^2 \delta_k^2}{f_k^2 F_{H_k}^2 E_k^2} (a_k^2 - b_k^2)$$

$$\text{and } D_{\text{ano}_k} = \frac{2F_P^2 \delta_k^2}{f_k^2 F_{H_k}^2 E_k^2} a_k b_k.$$

### Direct methods phase information

The prospect of employing direct phase relations to proteins in conjunction with the isomorphous replacement method (Coulter, 1965; Karle 1966) presents the problem of combining this independent phase information with that from other methods. The most useful relation for non-centrosymmetric structure analysis has been the tangent formula (Karle & Hauptman, 1956),

$$\tan \langle \alpha_h \rangle = \frac{\sum_k |E_k E_{h-k}| \sin (\alpha_k + \alpha_{h-k})}{\sum_k |E_k E_{h-k}| \cos (\alpha_k + \alpha_{h-k})} \quad (10)$$

The expected value of the phase angle for reflection  $h$ ,  $\langle \alpha_h \rangle$ , is determined from those reflections with Miller indices  $k$  and  $h-k$  for which the phase angles  $\alpha_k$  and  $\alpha_{h-k}$  are known. According to a generalization by Karle & Karle (1966) of a probability formula first described by Cochran (1955), the probability distribution of  $\alpha$  is

$$P_{\tan}(\alpha) = N \exp [\kappa \cos (\alpha - \langle \alpha_h \rangle)] \quad (11)$$

where

$$\kappa = 2 \frac{\sigma_3}{\sigma_2^{3/2}} |E_h| \left\{ \left[ \sum_k |E_k E_{h-k}| \cos (\alpha_k + \alpha_{h-k}) \right]^2 + \left[ \sum_k |E_k E_{h-k}| \sin (\alpha_k + \alpha_{h-k}) \right]^2 \right\}^{1/2}.$$

The  $E$ 's are normalized structure factors and  $\sigma_n = \sum_j f_j^n$ , where  $f_j$  is the scattering factor of the  $j$ th atom. This formulation is readily converted to the representation used for other phase information by expansion of the cosline of the difference between two angles. Thus

$$\kappa \cos (\alpha - \langle \alpha_h \rangle) = A_{\tan} \cos \alpha + B_{\tan} \sin \alpha \quad (12)$$

where

$$A_{\tan} = \kappa \cos \langle \alpha_h \rangle \text{ and } B_{\tan} = \kappa \sin \langle \alpha_h \rangle.$$

### Partial structure phase information

The partial structure method of phase determination, wherein the phase of an observed structure factor is taken as that calculated from only some of the atoms of the structure, has also been subjected to probability analysis. For some special proteins this method may be of use in the usual sense of a few heavy atoms comprising the known portion. Moreover, Rossmann & Blow (1961) have suggested that it may also be used to refine and extend the partial structure of a protein determined by other means.

According to Sim (1959) the probability of the phase angle of  $F_P$  being  $\alpha$ , given the partial structure factor contribution,  $F_K = F_K \exp (i\alpha_K)$ , of the atoms of known position is

$$P_{\text{par}}(\alpha) = N \exp [X \cos (\alpha - \alpha_K)] \quad (13)$$

where

$$X = \frac{2F_P F_K}{\Sigma}$$

and  $\Sigma$  represents the contribution of the unknown atoms which is  $\sum_u f_u^2$  by Wilson statistics. Rossmann &

Blow proposed that inaccuracies in the 'known' partial structure be taken into account in  $\Sigma$  to yield

$$\Sigma = \sum_u f_u^2 + \frac{4}{3} \pi S^2 \sum_k \sigma_k^2 f_k^2.$$

Atoms of unknown position have scattering factor  $f_u$  and those of 'known' position have scattering factors  $f_k$  and standard errors in position  $\sigma_k$ . If the isotropic positional errors are in absolute units, the magnitude of the reciprocal-space vector,  $S$ , is  $2 \sin \theta / \lambda$ . The probability formulation from partial structure computations is like that for the tangent formula and can be placed in the standard representation by

$$X \cos (\alpha - \alpha_K) = A_{\text{par}} \cos \alpha + B_{\text{par}} \sin \alpha \quad (14)$$

where

$$A_{\text{par}} = X \cos \alpha_K \text{ and } B_{\text{par}} = X \sin \alpha_K.$$

### Non-crystallographic symmetry phase information

Phase determination by use of non-crystallographic symmetry is another method that promises to be important in protein structure analysis. Rossmann & Blow (1963) have developed equations expressing the phase information due to molecular symmetry within the asymmetric unit. A solution to these equations (Rossmann & Blow, 1964), which has also formed the basis for a solution to the more general equations of Main & Rossmann (1966), uses the phase probability distribution

$$P_{\text{sym}}(\alpha) = N \exp (-\varepsilon^2(\alpha)/2E^2) \quad (15)$$

where

$$\varepsilon(\alpha) = |T_i^c \exp [i(\alpha + \Phi_i^c)] + A_{ii} \exp [i(2\alpha + \varphi_{ii})] - S_i|.$$

The parameters,  $T_i^c$ ,  $\Phi_i^c$ ,  $A_{ii}$ ,  $\varphi_{ii}$  and  $S_i$  (as defined by Rossmann & Blow, 1963; 1964), of the lack of closure error for the  $i$ th reflection can be calculated from the structure amplitudes, rotational and translational disposition of the molecules and a starting set of phases. A procedure has also been given for estimating the mean-square error,  $E^2$ .

By trigonometric expansion of  $\varepsilon^2(\alpha)$  this phase information can also be cast in the standard form. The resulting expansion is

$$-\varepsilon^2(\alpha)/2E^2 = K_{\text{sym}} + A_{\text{sym}} \cos \alpha + B_{\text{sym}} \sin \alpha + C_{\text{sym}} \cos 2\alpha + D_{\text{sym}} \sin 2\alpha \quad (16)$$

where

$$K_{\text{sym}} = -(T_i^{c2} + A_{ii}^2 + S_i^2)/2E^2,$$

$$A_{\text{sym}} = -T_i^c [A_{ii} \cos (\Phi_i^c - \varphi_{ii}) + S_i \cos \Phi_i^c]/E^2$$

$$B_{\text{sym}} = -T_i^c [A_{ii} \sin (\Phi_i^c - \varphi_{ii}) + S_i \sin \Phi_i^c]/E^2$$

$$C_{\text{sym}} = -A_{ii} S_i \cos \varphi_{ii}/E^2$$

and

$$D_{\text{sym}} = A_{ii} S_i \sin \varphi_{ii}/E^2.$$

Similar phase information can be derived from the direct-space use of non-crystallographic symmetry, as by Muirhead, Cox, Mazzarella & Perutz (1967). However, no probability distribution has yet been formulated for this information and the method for combination of this with other phase information remains uncertain.

### Combination of phase information

Phase information from the various methods and from the different derivatives is independent, so the overall phase probability is the product of the individual phase probabilities from (5), (8), (11), (13) and (15):

$$P(\alpha) = \left[ \prod_j P_{\text{iso}_j}(\alpha) \right] \cdot \left[ \prod_k P_{\text{ano}_k}(\alpha) \right] \cdot P_{\text{tan}}(\alpha) \cdot P_{\text{par}}(\alpha) \cdot P_{\text{sym}}(\alpha).$$

Since the probability function for each source of information,  $s$ , has been put into the form

$$P_s(\alpha) = N_s \exp (K_s + A_s \cos \alpha + B_s \sin \alpha + C_s \cos 2\alpha + D_s \sin 2\alpha)$$

the combination of phase information is achieved by simply adding coefficients in the exponent,

$$P(\alpha) = \prod_s P_s(\alpha) = N' \exp \left( \sum_s K_s + \sum_s A_s \cos \alpha + \sum_s B_s \sin \alpha + \sum_s C_s \cos 2\alpha + \sum_s D_s \sin 2\alpha \right).$$

The constant factor  $\exp \left( \sum_s K_s \right)$  may be included in the normalization factor, so

$$P(\alpha) = N \exp (A \cos \alpha + B \sin \alpha + C \cos 2\alpha + D \sin 2\alpha) \quad (17)$$

where the coefficients  $A$ ,  $B$ ,  $C$  and  $D$  are obtained by summation from (6), (9), (12), (14) and (16). The normalization factor  $N$  is of no consequence as it cancels when the phase probability is used in (1).

Two important advantages accrue from this formulation of the phase probabilities. First, the four coefficients  $A$ ,  $B$ ,  $C$  and  $D$  constitute a complete record of the phase information for a reflection. Second, inclusion of newly obtained phase information only requires additions to these coefficients. This obviates the need for trigonometric inversion to extract the angular parameters  $\Phi_1$  and  $\Phi_2$  from the vector sums  $\mathbf{C}_{is}$  and  $\mathbf{D}_{is}$  of Rossmann & Blow (1961, equation 6). Without the simplified representation either the several observed and calculated parameters associated with each derivative or the complete phase probability curves must be stored. Then to include new information either a complete recalculation of phase probability or point-

by-point multiplication of the probability curves is necessary.

It is noteworthy that within the limits of the error models upon which they are based, the individual probability distributions correctly weight the phase information from various sources. Thus a 'bad' isomorphous derivative will have large standard errors,  $E$ , leading to broad phase probability distributions which will be de-emphasized on combination with the sharper distributions from derivatives with smaller  $E$ 's. Similarly, the overall sharpness, and hence the weight, of the distributions from other sources derives from factors which assess the worth of the phase information. For example, the weight of tangent-formula phase information depends on the number of atoms, the magnitude and number of the contributing terms and the consistency among the contributors. It may, of course, happen that other factors than those included in a probability distribution will prove to be important. If an *ad hoc* weighting of a source of phase information is deemed necessary to account for such factors, the weight would best be applied to  $A$ ,  $B$ ,  $C$  and  $D$  so as to effect a change in the variance for that source.

### Evaluation of the phase integrals

Inclusion of the formulation for the total phase probability (17) in the definition of the coefficients for the 'best Fourier' (1) produces the integrals

$$\begin{pmatrix} Q_1 \\ Q_2 \\ Q_3 \end{pmatrix} = \int_0^{2\pi} \begin{pmatrix} 1 \\ \sin \alpha \\ \cos \alpha \end{pmatrix} \exp (A \cos \alpha + B \sin \alpha + C \cos 2\alpha + D \sin 2\alpha) d\alpha \cdot \begin{pmatrix} 18a \\ 18b \\ 18c \end{pmatrix}$$

Numerical integration is facilitated by this form in that the trigonometric functions of the dummy variable  $\alpha$  are separated from the other parameters. Consequently a short, exact table look-up for the sines and cosines is possible.

Analytic integration of (18) is also possible. Letting

$$S = \sqrt{A^2 + B^2}, \quad T = \sqrt{C^2 + D^2},$$

$$\tan \sigma = -B/A \quad \text{and} \quad \tan \tau = -D/C,$$

(18), upon rearrangement, yields

$$\begin{pmatrix} Q_1 \\ Q_2 \\ Q_3 \end{pmatrix} = \int_0^{2\pi} \begin{pmatrix} 1 \\ \sin \alpha \\ \cos \alpha \end{pmatrix} \exp [S \cos (\alpha + \sigma)] \times \exp [T \cos (2\alpha + \tau)] d\alpha \cdot \begin{pmatrix} 19a \\ 19b \\ 19c \end{pmatrix}$$

The exponentials in (19) are of the form  $\exp(t \cos u)$  which may be expanded in an infinite series of Bessel functions of imaginary argument

$$\exp(t \cos u) = I_0(t) + 2 \sum_{n=1}^{\infty} I_n(t) \cos nu.$$

This is but a special case of the expansion of the generating function for Bessel coefficients and so the series is absolutely convergent, in this case for all values of  $t$  and  $u$  (Watson, 1952). Use of this expansion in (19) produces

$$\begin{aligned} \begin{pmatrix} Q_1 \\ Q_2 \\ Q_3 \end{pmatrix} &= \int_0^{2\pi} \begin{pmatrix} 1 \\ \sin \alpha \\ \cos \alpha \end{pmatrix} [I_0(S)I_0(T) \\ &+ 2I_0(S) \sum_{n=1}^{\infty} I_n(T) \cos n(2\alpha + \tau) \begin{pmatrix} 20a \\ 20b \\ 20c \end{pmatrix} \\ &+ 2I_0(T) \sum_{m=1}^{\infty} I_m(S) \cos m(\alpha + \sigma) \\ &+ 4 \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} I_m(S)I_n(T) \cos m(\alpha + \sigma) \cos n(2\alpha + \tau)] d\alpha, \end{aligned}$$

the integrands of which, being the products of absolutely convergent series, are absolutely convergent. The integrands are also uniformly convergent for  $S, T < T'$  where  $T'$  is any positive number, as can be seen by comparison with the series with cosine arguments of 0 and  $S = T = T'$ . This justifies term-by-term integration and assures the uniform convergence of the results (Whittaker & Watson, 1952).

Integration of (20a) is performed first. Since

$$\cos u \cos v = \frac{1}{2} [\cos(u+v) + \cos(u-v)]$$

the general term may be written

$$2I_m(S)I_n(T) \{ \cos [(m+2n)\alpha + m\sigma + n\tau] + \cos [(m-2n)\alpha + m\sigma - n\tau] \}.$$

Now taking note that

$$\int_0^{2\pi} \cos(l\theta + \delta) d\theta = \begin{cases} 0 & l=1, 2, \dots \\ 2\pi \cos \delta & l=0 \end{cases}$$

term-by-term integration of (20a) yields non-vanishing integrals only for the first term and when  $m=2n$  in the last term, and the result is

$$\begin{aligned} Q_1 &= 2\pi I_0(S)I_0(T) \\ &+ 4\pi \sum_{n=1}^{\infty} I_{2n}(S)I_n(T) \cos n(2\sigma - \tau). \end{aligned} \quad (21a)$$

By similar means (20b) and (20c) can be integrated to give

$$\begin{aligned} \begin{pmatrix} Q_2 \\ Q_3 \end{pmatrix} &= 2\pi I_0(T)I_1(S) \frac{\sin}{\cos}(-\sigma) \\ &+ 2\pi \sum_{n=1}^{\infty} I_n(T) \{ I_{2n-1}(S) \frac{\sin}{\cos}[-\sigma + n(2\sigma - \tau)] \\ &+ I_{2n-1}(S) \frac{\sin}{\cos}[-\sigma - n(2\sigma - \tau)] \} \end{aligned}$$

or, upon expansion of the trigonometric terms and use of the identity

$$I_{m-1}(u) - I_{m+1}(u) = \frac{2m}{u} I_m(u),$$

$$\begin{aligned} \begin{pmatrix} Q_2 \\ Q_3 \end{pmatrix} &= 2\pi I_0(T)I_1(S) \frac{\sin}{\cos}(-\sigma) \quad \begin{pmatrix} 21b \\ 21c \end{pmatrix} \\ &+ 2\pi \sum_{n=1}^{\infty} I_n(T) \{ [I_{2n-1}(S) + I_{2n+1}(S)] \cos n(2\sigma - \tau) \\ &\frac{\sin}{\cos}(-\sigma) \begin{pmatrix} + \\ - \end{pmatrix} \frac{4n}{S} I_{2n}(S) \sin n(2\sigma - \tau) \frac{\cos}{\sin}(-\sigma) \} \end{aligned}$$

Unfortunately, the results (21) are not in closed form; however, these convergent series may be evaluated to any desired accuracy by inclusion of sufficient terms. This evaluation is expedited by use of recurrence relations for the Bessel functions as well as for the sines and cosines.

In the centrosymmetric case the phase of a structure factor is constrained to two possible values. This constraint can be introduced into the phase integrals of (1) by modifying the integrands with a factor which is the sum of two delta functions, each centered at a permissible phase angle. Then integration leads directly to the 'best Fourier' coefficient

$$\xi = F_P \tanh A \quad (22a)$$

for pure real structure factors and

$$\xi = iF_P \tanh B \quad (22b)$$

for pure imaginary structure factors. The phases computed from (22) will be the same as those obtained by unconstrained evaluation of (1), but the figures of merit, or weights, will be higher as a result of the centrosymmetric constraint.

### Computational tests

Since the simplified form used here to describe phase information from the method of isomorphous replacement involves a reformulation of the error analysis, the validity of the new analysis was checked by phase computations on a solved problem. The data and heavy-atom parameters from the 5.5 Å resolution structure analysis of *Glycra* hemoglobin (Padlan & Love, 1968) were used in the tests. Dr Hilary Muirhead's program (see Lipscomb, Coppola, Hartsuck, Ludwig, Muirhead, Searl & Steitz, 1966) for phase determination was modified to compute probability distributions and centroid phases by (1) the method of Blow & Crick (1959) as implemented by Dickerson, Kendrew & Strandberg (1961), (2) the method of Cullis, Muirhead, Perutz, Rossmann & North (1961) as programmed by Muirhead, (3) the approximation method of Rossmann & Blow (1961) and (4) the method proposed in this paper. For each method the mean figure of merit,  $\langle m \rangle$ , the mean absolute phase discrepancy,  $\langle |\Delta\alpha| \rangle$ , and root-mean-square relative phase discrepancy,  $\langle (\Delta\alpha/\cos^{-1}m)^2 \rangle^{1/2}$ , were computed for five classes of

structure factor amplitude. The Blow & Crick method was taken as the standard so that  $\Delta\alpha$  and  $\cos^{-1}m$  are taken with respect to that method. Initial estimates of the standard deviation of error,  $E$ , for each method were refined, as a function of  $F_p$ , to the root-mean-square lack of closure error at the centroid phase angles obtained from the estimated  $E$ 's. Two cycles sufficed for convergence.

The results of these tests are shown in Table 1. Two tests were done with essentially the same comparative results: one using all five of the available heavy-atom derivatives of Glycera hemoglobin; the other using only two of the five (HAuCl<sub>4</sub> and HgX). It is clear that the proposed method is entirely adequate for phase determination. Phases determined by this method are on the average, and especially for strong reflections, even nearer the Blow & Crick phases than are the phases determined by the proven method of Cullis *et al.* In both cases the phase discrepancies are well within the standard errors of the phase angles. As shown by the mean figures of merit, phase probability distributions computed by the proposed method do tend to be somewhat broader than those from the methods of Blow & Crick or Cullis *et al.* particularly for weak reflections.

However, inspection of the actual distributions finds them fully reasonable. On the other hand, the disparity of figures of merit for the approximation method of Rossmann & Blow, lower than in the Blow & Crick method for weak reflection and higher for strong reflections, is paralleled by several unreasonably flat or spike-like phase probability distributions.

### Conclusion

The proposed method of representation of phase probability distributions has several advantages over the usual procedures.

(1) Phase information from various sources can all be treated under one unified scheme.

(2) A complete, ready-to-use record of the phase information for a reflection is contained in four parameters.

(3) Newly obtained phase information can be combined with previous information by simple addition of coefficients.

(4) Numerical integration of the phase integrals is facilitated.

(5) The form is amenable to analytic integration.

Table 1. Comparison of methods for isomorphous replacement phase computations on Glycra hemoglobin data

Parameter and method	Values for structure factor amplitude classes					
	Total	1	2	3	4	5
Reflections						
Total	542	72	126	148	120	76
Acentric	421	43	106	127	92	53
Mean structure factor amplitude, $\langle F_p \rangle$	33.7	7.9	17.9	29.8	45.7	73.1
Two-derivative case						
Mean figure of merit, $\langle m \rangle$						
BC, DKS <sup>a</sup>	0.73	0.59	0.74	0.71	0.83	0.73
CMPRN, M <sup>b</sup>	0.75	0.61	0.78	0.74	0.83	0.76
RB <sup>c</sup>	0.66	0.31	0.59	0.64	0.87	0.81
HL <sup>d</sup>	0.72	0.55	0.73	0.70	0.82	0.73
Mean phase discrepancy, $\langle  \Delta\alpha  \rangle$						
CMPRN, M	5.0°	11.8°	4.3°	4.4°	2.9°	5.9°
RB	8.3	16.1	8.1	7.1	6.8	8.0
HL	4.5	11.4	5.0	3.6	2.9	2.6
R.m.s. relative phase discrepancy, $\langle (\Delta\alpha/\cos^{-1}m)^2 \rangle^{1/2}$						
CMPRN, M	0.19	0.37	0.16	0.16	0.13	0.19
RB	0.35	0.58	0.37	0.30	0.29	0.28
HL	0.19	0.31	0.23	0.13	0.15	0.12
Five-derivative case						
Mean figure of merit						
BC, DKS	0.87	0.73	0.88	0.89	0.91	0.86
CMPRN, M	0.86	0.75	0.87	0.88	0.89	0.87
RB	0.84	0.58	0.80	0.87	0.93	0.93
HL	0.84	0.69	0.83	0.86	0.90	0.84
Mean phase discrepancy*						
CMPRN, M	6.1°	10.1°	6.7°	5.3°	5.2°	5.0°
RB	7.7	12.2	8.1	7.8	5.9	5.8
HL	4.7	10.4	6.1	4.2	2.4	2.3
R.m.s. relative phase discrepancy*						
CMPRN, M	0.41	0.72	0.39	0.36	0.37	0.29
RB	0.62	0.81	0.78	0.62	0.35	0.34
HL	0.35	0.44	0.47	0.35	0.19	0.12

(a) Method of Blow & Crick (1959) and Dickerson, Kendrew & Strandberg (1961); (b) method of Cullis *et al.* (1961) and Muirhead (Lipscomb *et al.*, 1966); (c) method of Rossmann & Blow (1961); (d) method proposed here.

\* Computed over acentric reflections only.

*Note added in proof:*— We have recently learned that the distribution  $\exp(\kappa \cos \alpha)$ , which is fundamental to the phase probability distributions above, was first derived by von Mises (1918) from the principle of maximum likelihood of Gauss. Its relation to the Gaussian distribution has prompted statisticians to refer to it as the circular normal distribution. A review of the properties and applications of it and other circular distributions has been written by Batschelet (1965).

We are grateful to Eduardo Padlan for the use of his data from *Glycera* hemoglobin, to Michael Rossmann and David Blow for their helpful comments on the manuscript and to Warner Love for his encouragement and support. We also acknowledge the support of the U.S. Public Health Service through predoctoral fellowship GM-15907 (EEL) and GM30009 (WAH) and grant AM02528 and thank the Johns Hopkins Faculty of Arts and Sciences for computing funds.

### References

- BATSCHULET, E. (1965). *Statistical Methods for the Analysis of Problems in Animal Orientation and Certain Biological Rhythms*. Amer. Inst. of Biol. Sci. Monograph Washington, D.C.
- BLOW, D. M. & CRICK, F. H. C. (1959). *Acta Cryst.* **12**, 794.
- COCHRAN, W. (1955). *Acta Cryst.* **8**, 473.
- COULTER, C. L. (1965). *J. Mol. Biol.* **12**, 292.
- CULLIS, A. F., MUIRHEAD, H., PERUTZ, H. 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.
- GREEN, D. W., INGRAM, V. M. & PERUTZ, M. F. (1954). *Proc. Roy. Soc. A* **225**, 287.
- KARLE, J. (1966). *Acta Cryst.* **21**, 273.
- KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635.
- KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* **21**, 849.
- 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.
- MAIN, P. & ROSSMANN, M. G. (1966). *Acta Cryst.* **21**, 67.
- MATTHEWS, B. W. (1966). *Acta Cryst.* **20**, 82.
- MISES, R. VON (1918). *Physical. Z.* **19**, 490.
- MUIRHEAD, H., COX, J. M., MAZZARELLA, L. & PERUTZ, M. F. (1967). *J. Mol. Biol.* **28**, 117.
- NORTH, A. C. T. (1965). *Acta Cryst.* **18**, 212.
- PADLAN, E. A. & LOVE, W. E. (1968). *Nature, Lond.* **220**, 376.
- ROSSMANN, M. G. & BLOW, D. M. (1961). *Acta Cryst.* **14**, 641.
- ROSSMANN, M. G. & BLOW, D. M. (1963). *Acta Cryst.* **16**, 39.
- ROSSMANN, M. G. & BLOW, D. M. (1964). *Acta Cryst.* **17**, 1474.
- SIM, G. A. (1959). *Acta Cryst.* **12**, 813.
- WATSON, G. N. (1952). *A Treatise on the Theory of Bessel Functions*, Chap. 2 & 3. Cambridge Univ. Press.
- WHITTAKER, E. T. & WATSON, G. N. (1952). *A Course of Modern Analysis*, Chap. 2–4. Cambridge Univ. Press.

*Acta Cryst.* (1970). B26, 143

## The Crystal Structure of *N*-Benzyl-4-methylthiazolium Bromide

BY L. POWER,\* J. PLETCHER AND M. SAX

*The Biocrystallography Laboratory, Veterans Administration Hospital, Pittsburgh, Pa. 15240, U.S.A.*  
*and The Crystallography Laboratory, University of Pittsburgh, Pittsburgh, Pa. 15213, U.S.A.*

(Received 27 February 1969)

The crystal structure of *N*-benzyl-4-methylthiazolium bromide,  $C_{11}NSH_2Br$ , was determined from three-dimensional X-ray diffraction data. The crystals are monoclinic and the space group is  $P2_1/c$  with four molecules per unit cell. The axial dimensions are  $a = 9.162 \pm 0.003$ ,  $b = 11.770 \pm 0.004$  and  $c = 11.070 \pm 0.004$  Å,  $\beta = 81.93 \pm 0.04^\circ$ . The structure was solved by the heavy-atom method and refined by means of full-matrix least squares. The final  $R$  value was 0.056 on 1991 reflections which included 221 unobserved. The bond lengths in the thiazolium ring are compared with those in thiamine and thiamine pyrophosphate. The conformation of the rings in this molecule differs substantially from that observed for the other two molecules. A weak  $CH \cdots Br$  bifurcated hydrogen bond occurs in the structure. Another interaction is observed between the sulfur atom and the bromide ion as the  $Br \cdots S$  interatomic distances average 0.3 Å less than the sum of their van der Waals radii.

### Introduction

Breslow & McNelis (1958) have pointed out that in a number of biochemical reactions involving thiamine

pyrophosphate (TPP) as coenzyme the thiazolium zwitterion is the site of primary reaction. It was shown that the attachment of an aromatic ring on the *N*-methylene group of a thiazolium ring facilitates the formation of the catalytic zwitterion due to the inductive effect of the aromatic ring. The *N*-benzyl thiazolium salt and TPP are both effective catalysts to these

\* Permanent address: The University of Townsville, Pimlico, Townsville, Queensland, Australia.