

where

$$A_{jk} = F^*(\mathbf{h}[C_j]^{-1})F(\mathbf{h}[C_k]^{-1}) \times \exp[2\pi i \mathbf{h} \cdot (\mathbf{u}_j - \mathbf{u}_k)]. \quad (A2)$$

Equation (A1) can be transformed either with respect to the reciprocal-space vectors  $\mathbf{h}$ , resulting in a transform with vectors  $([C_j] - [C_k])\mathbf{t}$ , or with respect to the reciprocal-space cross vectors  $\mathbf{h}([C_j] - [C_k])$ , giving rise to a transform with vectors  $\mathbf{t}$ . The main advantage of the latter lies in the ease of combination of different symmetry elements. A Fourier synthesis with modified coefficients of this sort has been described previously by Rius & Miravittles (1986).

The translation and packing functions are calculated as follows.

(i) Structure factors are calculated for the correctly oriented molecule, placed in a triclinic unit cell of identical lattice constants as the target cell; this avoids interpolation in the ensuing steps.

(ii) The structure factors are sorted according to resolution such that symmetry-related reflections are grouped together.

(iii) For a given vector  $\mathbf{h}$  in the asymmetric unit, the structure factors are found for all symmetry-related vectors  $\mathbf{h}[C_j]^{-1}$ . The complex coefficients (A2) are evaluated for each symmetry element [using (3) and (8) for the packing and translation functions respectively] and assigned to cross vectors  $\mathbf{h}([C_j] - [C_k])$ .

(iv) The resulting difference vector coefficients are added together (from all symmetry elements for the translation function, pairwise for the overlap function) and Fourier transformed. The packing-function coefficients must be weighted by an artificial  $B$  factor to dampen spurious ripples, resulting in a smooth

function (see Fig. 2); this is necessary for the combination with the translation function as outlined in (6).

(v) The packing function is evaluated according to (4) and (5), using an appropriate value for  $\kappa$ .

(vi) Modified translation functions are calculated according to (6).

The current translation-function package requires ~4 min CPU on a VAX 8550 from atomic coordinates to translation function for the example given here (29 823 structure factors between 8 and 3.5 Å, 6 symmetry operations); the packing function currently takes <20 min CPU for the same problem (15 cross vectors), although the use of external programs, with consequent multiple file conversions, represents a considerable fraction of retrievable computation time.

Programs for the packing and translation functions described here are available on request.

#### References

- BOTT, R. & SARMA, R. (1976). *J. Mol. Biol.* **106**, 1037-1046.  
 CROWTHER, R. A. & BLOW, D. M. (1967). *Acta Cryst.* **23**, 544-548.  
 FITZGERALD, P. (1990). Conf. Proc. of Crystallographic Computing School, Bischofsberg, Germany, pp. 287-300.  
 FUJINAGA, M. & READ, R. J. (1987). *J. Appl. Cryst.* **20**, 517-521.  
 HARADA, Y., LIFCHITZ, A., BERTHOU, J. & JOLLES, P. (1981). *Acta Cryst.* **A37**, 398-406.  
 HENDRICKSON, W. A. & WARD, K. B. (1976). *Acta Cryst.* **A32**, 778-780.  
 HOPPE, W. (1957a). *Acta Cryst.* **10**, 750-751.  
 HOPPE, W. (1957b). *Z. Elektrochem.* **61**, 1076.  
 HUBER, R. (1965). *Acta Cryst.* **19**, 353-356.  
 LATTMAN, E. E. & LOVE, W. E. (1970). *Acta Cryst.* **B26**, 1854-1857.  
 NORDMAN, C. E. & NAKATSU, K. (1963). *J. Am. Chem. Soc.* **85**, 353.  
 RIUS, J. & MIRAVITLLES, C. (1986). *Acta Cryst.* **A42**, 402-404.  
 ROSSMANN, M. G. (1972). Editor. *The Molecular Replacement Method*. New York: Gordon and Breach.  
 ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24-31.  
 STUBBS, M. T., LABER, B., BODE, W., HUBER, R., JERALA, R., LENARČIĆ, B. & TURK, V. (1990). *EMBO J.* **9**, 1939-1947.

*Acta Cryst.* (1991). **A47**, 526-533

## On the Application of Direct Methods to Complex Structures. XXXI. Properties and Limitations of Sayre's Equation

BY M. SHIONO AND M. M. WOOLFSON

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

(Received 13 December 1990; accepted 20 March 1991)

### Abstract

Since Sayre's equation is the basis of some direct-methods procedures, the applicability of Sayre's equation has been tested in various circumstances. When a structure contains a heavy atom, it is found that Sayre's equation does not hold well, which is

expected since the condition of equal resolved atoms does not apply. However, what is not expected is that with a heavy atom present the equation actually holds better at low resolution than at high resolution. The cause of this apparent anomaly is discussed and it is shown that there exists a modified Sayre's equation which holds far better in the presence of one kind of

heavy atom in a structure. While this modified equation is difficult to use, it is possible that it may find application if the full power of modern computer technology is utilized.

### Introduction

Sayre (1952) derived an equation which gives relationships between structure factors for structures containing equal resolved atoms in the form

$$F(\mathbf{h}) = [f(\mathbf{h})/g(\mathbf{h})]G(\mathbf{h}) \\ = [f(\mathbf{h})/g(\mathbf{h})V] \sum_{\mathbf{k}} F(\mathbf{k})F(\mathbf{h}-\mathbf{k}), \quad (1)$$

where  $f(\mathbf{h})$  and  $g(\mathbf{h})$  are the scattering factors of the true and squared-electron-density atoms, respectively,  $G(\mathbf{h})$  is the structure factor for the squared structure and  $V$  is the volume of the unit cell. The equation is based on the following idea: if the peaks in the electron-density map are equal and well resolved then the squared electron density is very similar to the original density. The only difference is in the shape of the peaks which is compensated by the factor  $f(\mathbf{h})/g(\mathbf{h})$ .

Recently introduced direct phase-determining methods are based on Sayre's equation. One such method involves the Sayre-equation tangent formula (SETF) (Debaerdemaeker, Tate & Woolfson, 1985, 1988) which attempts to minimize the least-squares residual for the Sayre equation. Sayre's equation for normalized structure factors is written as

$$E(\mathbf{h}) = [f/g(\mathbf{h})V] \sum_{\mathbf{k}} E(\mathbf{k})E(\mathbf{h}-\mathbf{k}). \quad (2)$$

For normalized structure factors,  $f$  is independent of  $\mathbf{h}$  and equals  $N^{-1/2}$  for an equal-atom structure where  $N$  is the number of atoms in the unit cell. The scattering factor for squared atoms,  $g(\mathbf{h})$ , is equivalent to the self-convolution of the scattering factor for the true atoms, thus

$$g(\mathbf{h}) = V^{-1} \sum_{\mathbf{k}} f(\mathbf{k})f(\mathbf{h}-\mathbf{k}). \quad (3)$$

Since  $f(\mathbf{h})$  is constant ( $= N^{-1/2}$ ),  $g(\mathbf{h})$  equals  $(VN)^{-1}$  times the number of the terms in the summation of (3), which is given by the function

$$(\pi/12)V\{4s_{\max} + s(\mathbf{h})\}\{2s_{\max} - s(\mathbf{h})\}^2, \quad (4)$$

where  $s(\mathbf{h})$  is the length of the reciprocal-lattice vector,  $\mathbf{h}$ , and  $s_{\max}$  is the maximum value of  $s(\mathbf{h})$  over all the  $\mathbf{h}$ 's. The formula (4), except for the factor  $V$ , represents the overlapped volume in reciprocal space of the two spheres with radius  $s_{\max}$ , separated by  $s(\mathbf{h})$ . If we define the structure factor for the squared structure multiplied by  $V/g(\mathbf{h})$  by

$$G'(\mathbf{h}) = [g(\mathbf{h})]^{-1} \sum_{\mathbf{k}} E(\mathbf{k})E(\mathbf{h}-\mathbf{k}), \quad (5)$$

then the least-squares residual for the Sayre equation is given by

$$R = \sum_{\mathbf{h}} |E(\mathbf{h}) - KG'(\mathbf{h})|^2 / \sum_{\mathbf{h}} |E(\mathbf{h})|^2, \quad (6)$$

where  $K$  is a scale factor. If all possible terms were included in the summation of the Sayre equation then  $K$  would equal  $f/V$ . However, in the usual practical applications of the equation only a subset of larger  $E$ 's is included. In this case  $K$  must also compensate for the loss of terms in the summation and it can be shown that the value of  $K$  giving the lowest value of  $R$  is

$$K = \sum_{\mathbf{h}} E(\mathbf{h})^* G'(\mathbf{h}) / \sum_{\mathbf{h}} |G'(\mathbf{h})|^2. \quad (7)$$

A plausible phase set is expected to make (6) a minimum. From this point of view, for all  $\mathbf{l}$ ,

$$\partial R / \partial \varphi(\mathbf{l}) = 0.$$

The application of this condition to (6) gives

$$\tan \varphi(\mathbf{l}) = \frac{\text{Im}\{t(\mathbf{l})\} - (2T/3Q)\text{Im}\{q(\mathbf{l})\}}{\text{Re}\{t(\mathbf{l})\} - (2T/3Q)\text{Re}\{q(\mathbf{l})\}}, \quad (8)$$

where  $\text{Im}$  and  $\text{Re}$  indicate imaginary and real parts of complex quantities,

$$t(\mathbf{l}) = \sum_{\mathbf{h}} [1/g(\mathbf{l}) + 1/g(\mathbf{h}) + 1/g(\mathbf{l}-\mathbf{h})] E(\mathbf{h})E(\mathbf{l}-\mathbf{h}),$$

$$q(\mathbf{l}) = \sum_{\mathbf{h}} E(\mathbf{l}-\mathbf{h}) [1/g(\mathbf{h})^2] \sum_{\mathbf{k}} E(\mathbf{k})E(\mathbf{h}-\mathbf{k}),$$

$$T = \sum_{\mathbf{l}} E(\mathbf{l})^* t(\mathbf{l})$$

and

$$Q = \sum_{\mathbf{l}} E(\mathbf{l})^* q(\mathbf{l}) = \sum_{\mathbf{h}} |G'(\mathbf{h})|^2.$$

Formula (8) is the previously mentioned Sayre-equation tangent formula (SETF); Debaerdemaeker, Tate & Woolfson (1985, 1988) have shown that it is much more effective than the conventional tangent formula in direct-methods applications. For example, it was demonstrated by Woolfson & Yao (1990) that it would be effective in solving a small protein, avian pancreatic polypeptide (details given later), with more than 300 atoms in the asymmetric unit - although it should be said that this structure had data of unusually good quality and resolution.

Since the SETF attempts to develop phases so as to make the residual for a system of Sayre equations a minimum, it is interesting to examine the accuracy of the Sayre equation under various conditions. If in some circumstances the equation does not hold well then it may indicate that this method may not be applicable in such a case. What we do here is to report on tests of the Sayre equation at low resolution and also for structures containing heavy atoms.

### An apparently anomalous result

The behaviour of Sayre's equation has been tested on the structure of the 36 amino acid avian pancreatic polypeptide (App) (Glover, Haneef, Pitts, Wood, Moss, Tickle & Blundell, 1983). The molecule forms symmetrical dimers linked through zinc atoms in the crystal lattice. The crystal has space group  $C2$  with dimensions  $a = 34.18$ ,  $b = 32.92$ ,  $c = 28.45$  Å and  $\beta = 105.26^\circ$ . The asymmetric unit contains 301 non-hydrogen atoms belonging to the polypeptide, a zinc atom and 80  $H_2O$  molecules.

The procedure of the test is as follows:

- (1) Calculate the normalized structure magnitudes with the program *MULTAN* from the observed  $F$ 's.
- (2) Calculate the  $E$  map. True phases calculated from known atomic coordinates are used.
- (3) Square the  $E$  map.
- (4) Derive the structure factors by Fourier transformation of the squared  $E$  map and then divide them by  $g(\mathbf{h})$  in order to get structure factors proportional to  $G'(\mathbf{h})$  in (5).
- (5) Compare  $E(\mathbf{h})$ 's and  $G'(\mathbf{h})$ 's. The absolute values of  $E(\mathbf{h})$  and  $G'(\mathbf{h})$  can be compared by means of the residual,

$$R = \sum_{\mathbf{h}} \{ |E(\mathbf{h})| - K' |G'(\mathbf{h})| \}^2 / \sum_{\mathbf{h}} |E(\mathbf{h})|^2. \quad (9)$$

For phases the weighted mean error is calculated as

$$\langle \Delta\varphi \rangle_{EG'} = \frac{\sum_{\mathbf{h}} |E(\mathbf{h})| |G'(\mathbf{h})| |\varphi_E(\mathbf{h}) - \varphi_{G'}(\mathbf{h})|}{\sum_{\mathbf{h}} |E(\mathbf{h})| |G'(\mathbf{h})|}, \quad (10)$$

where  $\varphi_E(\mathbf{h})$ ,  $\varphi_{G'}(\mathbf{h})$  are phases of  $E(\mathbf{h})$  and  $G'(\mathbf{h})$ , respectively, and  $\langle \Delta\varphi \rangle_{EG'}$  is the phase error weighted with the product of the absolute values of  $E(\mathbf{h})$  and  $G'(\mathbf{h})$ . A suitable value of  $K'$ , in (9), is obtained by

$$K' = \sum_{\mathbf{h}} |E(\mathbf{h})| |G'(\mathbf{h})| / \sum_{\mathbf{h}} |G'(\mathbf{h})|^2. \quad (11)$$

The test has been done at various resolutions. The result of this analysis is plotted in Fig. 1 with circles.

A surprising result was obtained. The cumulative  $R$  factor and the weighted mean phase error are smallest at the lowest resolution and their values increase with increasing resolution. We shall discuss later the reason for this result and its implications for the use of Sayre's equation.

### The effect of heavy atoms

In order to confirm that the apparently anomalous effect was not due to experimental error in measuring structure factors, Sayre's equation was tested on the structure factors calculated from the known atomic coordinates of App ignoring both water molecules

and hydrogen atoms. However, the anomaly remained; the  $R$  factor and weighted mean phase error were smallest at lowest resolution although a difference from the result with the observed magnitudes is that the  $R$  factor and phase error improved beyond 1.29 Å resolution.

Eventually it was found that the cause of the anomaly was the presence of the zinc atoms. When Sayre's equation was tested on calculated structure factors for App with zinc atoms replaced by carbon atoms, the anomaly disappeared. A comparison of the results with and without zinc is shown in Fig. 2.

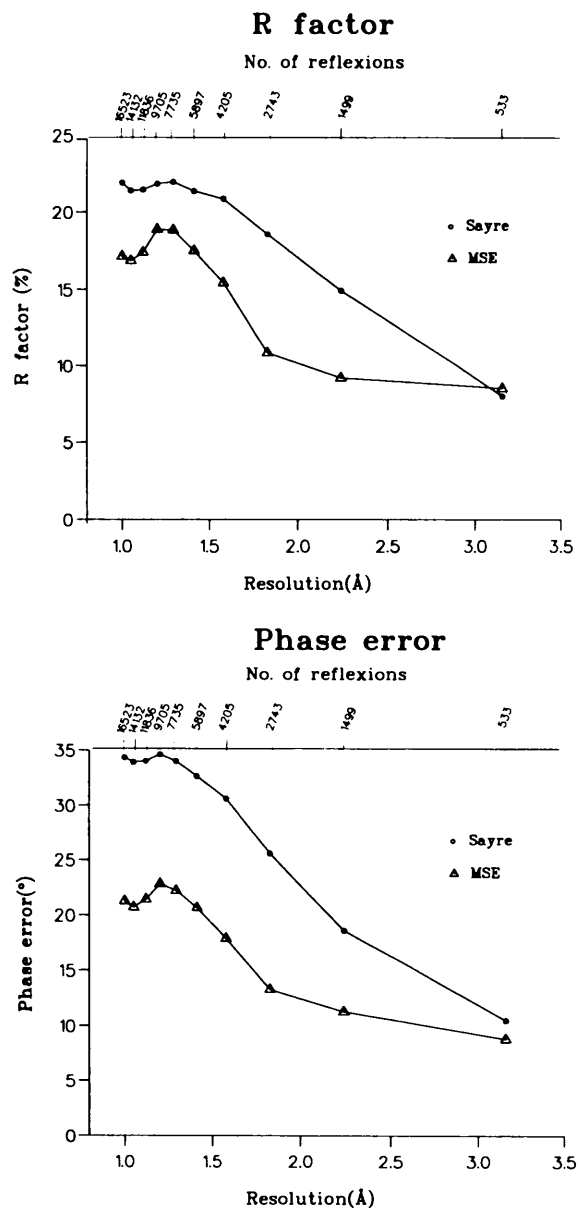


Fig. 1.  $R$  factors and phase errors of Sayre's equation and MSE with observed structure magnitudes of App. Circles are the results of Sayre's equation. Triangles are the results of MSE.

Without zinc, values of  $R$  factor and phase error generally decreased with increasing resolution as expected although there are some fluctuations and small increases persisted at the very highest resolutions. At resolution  $1.0 \text{ \AA}$  the  $R$  factor has a small value of 3.8%.

Actually, since Sayre's equation is strictly valid only for structures containing equal atoms, the  $R$  factor and mean phase error for a structure containing heavy atoms would be expected to be non-zero at any resolution, high or low.

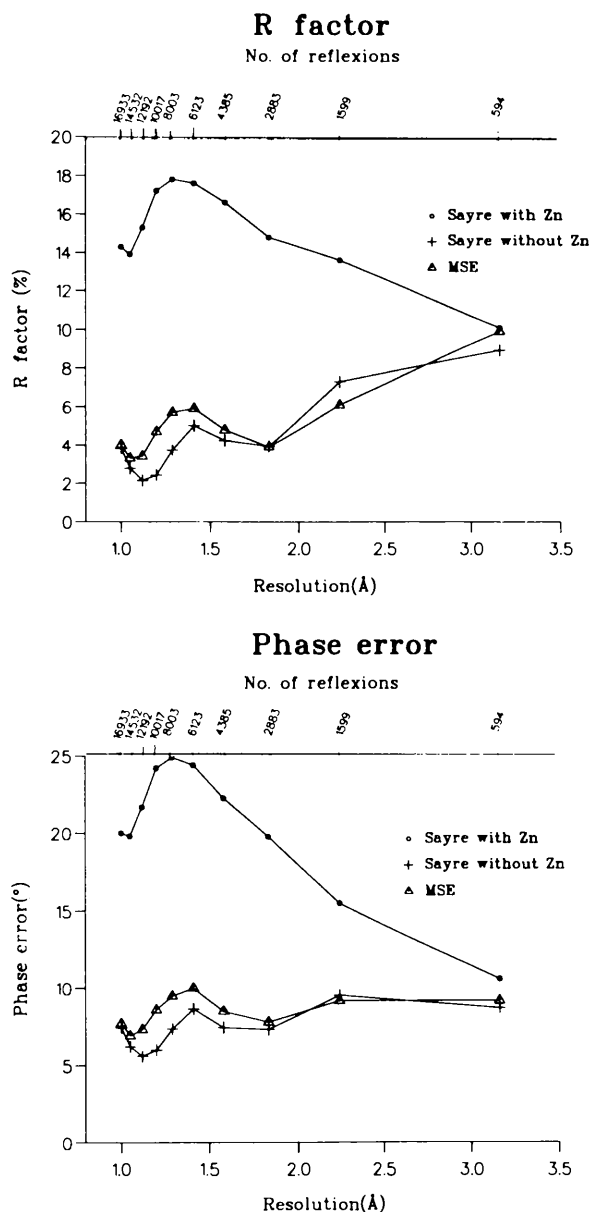


Fig. 2.  $R$  factors and phase errors of Sayre's equation and MSE with calculated structure magnitudes of App. Circles are the results of Sayre's equation. Crosses are the results of Sayre's equation on App without Zn. Triangles are the results of MSE.

### The modified Sayre's equation

Woolfson (1958) suggested the use of an equation which would be valid for structures containing two kinds of atoms. The form of the equation is

$$E(\mathbf{h}) = A(\mathbf{h})V^{-1} \sum_{\mathbf{k}} E(\mathbf{k})E(\mathbf{h}-\mathbf{k}) \\ + B(\mathbf{h})V^{-2} \sum_{\mathbf{k}} \sum_{\mathbf{l}} E(\mathbf{k})E(\mathbf{l})E(\mathbf{h}-\mathbf{k}-\mathbf{l}) \\ = A(\mathbf{h})G(\mathbf{h}) + B(\mathbf{h})H(\mathbf{h}), \quad (12)$$

where  $A(\mathbf{h})$  and  $B(\mathbf{h})$  are functions of the reciprocal-lattice vector,  $\mathbf{h}$ , and  $G(\mathbf{h})$  and  $H(\mathbf{h})$  are the structure factors for the squared and cubed  $E$  maps. In the original paper,  $F(\mathbf{h})$ 's were used, but here we shall use normalized structure factors,  $E(\mathbf{h})$ 's, and we shall call (12) the modified Sayre equation (MSE).

In order to use the MSE, we must derive the functions  $A(\mathbf{h})$  and  $B(\mathbf{h})$ . We denote the structure factors for squared and cubed  $E$  maps by  $E^{\text{sq}}$  and  $E^{\text{cu}}$ , respectively. They are written as

$$E^{\text{sq}} = [K^{\text{sq}}/g(\mathbf{h})V] \sum_{\mathbf{k}} E(\mathbf{k})E(\mathbf{h}-\mathbf{k}) \quad (13a)$$

and

$$E^{\text{cu}}(\mathbf{h}) = [K^{\text{cu}}/h(\mathbf{h})V] \sum_{\mathbf{k}} \sum_{\mathbf{l}} E(\mathbf{k})E(\mathbf{l})E(\mathbf{h}-\mathbf{k}-\mathbf{l}), \quad (13b)$$

where  $g(\mathbf{h})$  and  $h(\mathbf{h})$  are the functions which depend on the numbers of terms in the summations and  $K^{\text{sq}}$  and  $K^{\text{cu}}$  are the scaling constants. The quantity  $g(\mathbf{h})$  is proportional to (4) while  $h(\mathbf{h})$  is given as

$$h(\mathbf{h}) = V^{-2} \sum_{\mathbf{k}} \sum_{\mathbf{l}} f(\mathbf{k})f(\mathbf{l})f(\mathbf{h}-\mathbf{k}-\mathbf{l}) \\ = V^{-1} \sum_{\mathbf{k}} g(\mathbf{k})f(\mathbf{h}-\mathbf{k}) \\ \approx V^{-1} \sum_{\mathbf{k}} (\pi/12N) \{4s_{\text{max}} + s(\mathbf{k})\} \\ \times \{2s_{\text{max}} - s(\mathbf{k})\}^2 N^{-1/2} \delta(\mathbf{h}-\mathbf{k}),$$

where  $\delta(\mathbf{h})$  is the  $\delta$  function which has value 1 at the reciprocal-lattice points and zero elsewhere. This expression is replaced by an equivalent integration to give

$$h(\mathbf{h}) \approx \iiint (\pi/12N^{3/2}) \{4s_{\text{max}} + s(\mathbf{k})\} \\ \times \{2s_{\text{max}} - s(\mathbf{k})\}^2 dV^* \\ = (\pi^2/1260N^{3/2}) \\ \times \{1050s_{\text{max}}^6 - 525s_{\text{max}}^5 s(\mathbf{h}) - 315s_{\text{max}}^4 s(\mathbf{h})^2 \\ + 70s_{\text{max}}^3 s(\mathbf{h})^3 + 63s_{\text{max}}^2 s(\mathbf{h})^4 - s(\mathbf{h})^6\}, \quad (14)$$

where  $V^*$  is the reciprocal unit volume. The range of the integration is over the overlapped volume of two reciprocal spheres with radius  $s_{\text{max}}$  and the distance between their centres is  $s(\mathbf{h})$ . Now we can calculate the normalized structure factors for the squared and cubed structures with (13a) and (13b).

We have considered the case of structures which contain two kinds of atoms. Hence we have two scattering factors denoted by  $f_L$  for light atoms and  $f_H$  for heavy atoms. The MSE is derived from the Fourier transform of

$$\rho(\mathbf{u}) = A(\mathbf{h})\rho(\mathbf{u})^2 + B(\mathbf{h})\rho(\mathbf{u})^3, \quad (15)$$

where  $\rho(\mathbf{u})$  is the electron density at position  $\mathbf{u}$ . Then the scattering factors are described as linear combinations of the scattering factors for squared and cubed structures,  $(f)^{\text{sq}}$  and  $(f)^{\text{cu}}$ , as

$$f_L = a(f_L)^{\text{sq}} + b(f_L)^{\text{cu}} \quad (16a)$$

and

$$f_H = a(f_H)^{\text{sq}} + b(f_H)^{\text{cu}}. \quad (16b)$$

For ideal point atoms, the scattering factors would be given by

$$\begin{aligned} f_L &= Z_L, & (f_L)^{\text{sq}} &= Z_L^2, & (f_L)^{\text{cu}} &= Z_L^3, \\ f_H &= Z_H, & (f_H)^{\text{sq}} &= Z_H^2, & (f_H)^{\text{cu}} &= Z_H^3. \end{aligned} \quad (17)$$

Note that we usually consider carbon, nitrogen and oxygen atoms as light atoms and we compensate for this by giving average values of the scattering factors and their powers for light atoms,

$$f_L = \langle Z_L \rangle, \quad (f_L)^{\text{sq}} = \langle Z_L^2 \rangle, \quad (f_L)^{\text{cu}} = \langle Z_L^3 \rangle,$$

where

$$\langle Z_L^n \rangle = \sum_{j=1}^{N_L} Z_j^n / N_L$$

and  $N_L$  is the number of light atoms in the unit cell. Then (16a) and (16b) are written as

$$\langle Z_L \rangle = a\langle Z_L^2 \rangle + b\langle Z_L^3 \rangle$$

and

$$Z_H = aZ_H^2 + bZ_H^3.$$

By solving the simultaneous equations for  $a$  and  $b$ , we obtain

$$a = \frac{\langle Z_L \rangle Z_H^2 - \langle Z_L^3 \rangle}{(\langle Z_L^2 \rangle Z_H - \langle Z_L^3 \rangle) Z_H}, \quad (18a)$$

$$b = \frac{\langle Z_L \rangle Z_H - \langle Z_L^2 \rangle}{(\langle Z_L^2 \rangle Z_H - \langle Z_L^3 \rangle) Z_H}. \quad (18b)$$

We now write  $E$ ,  $E^{\text{sq}}$  and  $E^{\text{cu}}$  as

$$E(\mathbf{h}) = \sigma_2^{-1/2} \left\{ \langle Z_L \rangle \sum_{j=1}^{N_L} \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j) + Z_H \sum_{j=1}^{N_H} \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j) \right\}$$

$$E^{\text{sq}}(\mathbf{h}) = \sigma_4^{-1/2} \left\{ \langle Z_L^2 \rangle \sum_{j=1}^{N_L} \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j) + Z_H^2 \sum_{j=1}^{N_H} \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j) \right\},$$

$$E^{\text{cu}}(\mathbf{h}) = \sigma_6^{-1/2} \left\{ \langle Z_L^3 \rangle \sum_{j=1}^{N_L} \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j) + Z_H^3 \sum_{j=1}^{N_H} \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j) \right\},$$

where  $\sigma_n = \sum_{j=1}^{N_{\text{all}}} Z^n$  and  $N_H$  and  $N_{\text{all}}$  are the numbers of heavy atoms and all atoms in the unit cell, respectively. The factors  $\sigma_2^{-1/2}$ ,  $\sigma_4^{-1/2}$  and  $\sigma_6^{-1/2}$  ensure that each type of  $E$  is properly normalized. Now we can write the Fourier transform of (15) by applying the coefficients  $a$  and  $b$ ,

$$\sigma_2^{1/2} E(\mathbf{h}) = a\sigma_4^{1/2} E^{\text{sq}}(\mathbf{h}) + b\sigma_6^{1/2} E^{\text{cu}}(\mathbf{h})$$

or

$$E(\mathbf{h}) = a(\sigma_4^{1/2}/\sigma_2^{1/2}) E^{\text{sq}}(\mathbf{h}) + b(\sigma_6^{1/2}/\sigma_2^{1/2}) E^{\text{cu}}(\mathbf{h}).$$

Since we have already obtained the normalized structure factors,  $E^{\text{sq}}$  and  $E^{\text{cu}}$ , we can now write the MSE in its full form as

$$E(\mathbf{h}) = A(\mathbf{h}) V^{-1} \sum_{\mathbf{k}} E(\mathbf{k}) E(\mathbf{h} - \mathbf{k}) + B(\mathbf{h}) V^{-2} \sum_{\mathbf{k}} \sum_{\mathbf{l}} E(\mathbf{k}) E(\mathbf{l}) E(\mathbf{h} - \mathbf{k} - \mathbf{l}), \quad (19)$$

where

$$A(\mathbf{h}) = \frac{\langle Z_L \rangle Z_H^2 - \langle Z_L^3 \rangle}{(\langle Z_L^2 \rangle Z_H - \langle Z_L^3 \rangle) Z_H} \frac{\sigma_4^{1/2} K^{\text{sq}}}{\sigma_2^{1/2} g(\mathbf{h})}$$

and

$$B(\mathbf{h}) = \frac{\langle Z_L \rangle Z_H - \langle Z_L^2 \rangle}{(\langle Z_L^2 \rangle Z_H - \langle Z_L^3 \rangle) Z_H} \frac{\sigma_6^{1/2} K^{\text{cu}}}{\sigma_2^{1/2} h(\mathbf{h})}.$$

The MSE can be used as the basis of a new tangent formula (Appendix). However, since the quartet contribution given by the double summation in (A-5) requires all possible terms to be included, and not a convenient subset as in the Sayre-equation tangent formula, no tests of it have yet been made.

#### Application of the MSE to App

The MSE has been applied to the structure of App where the structure factors calculated from the known atomic coordinates are used.

The procedure for the calculation is as follows:

(1) Calculate the  $E$ 's from the atomic coordinates, assuming ideal point atoms.

(2) Calculate the  $E$  map.

(3) Square and cube the  $E$  map.

(4) Derive the structure factors  $G(\mathbf{h})$ 's and  $H(\mathbf{h})$ 's from the squared and cubed  $E$  maps.

(5) Calculate the right-hand side of the MSE, (19). The normalizing factors,  $K^{\text{sq}}$  and  $K^{\text{cu}}$ , are calculated by

$$K^{\text{sq}} = \frac{\sum_{\mathbf{h}} [1/g(\mathbf{h})] |E(\mathbf{h})| |G(\mathbf{h})|}{\sum_{\mathbf{h}} [1/g(\mathbf{h})^2] |G(\mathbf{h})|^2}$$

and

$$K^{\text{cu}} = \frac{\sum_{\mathbf{h}} [1/h(\mathbf{h})] |E(\mathbf{h})| |H(\mathbf{h})|}{\sum_{\mathbf{h}} [1/h(\mathbf{h})^2] |H(\mathbf{h})|^2}$$

(6) Calculate the  $R$  factor and phase error.

The result is plotted with the results of Sayre's equation in Fig. 2. We can see that the anomaly has now disappeared. The  $R$  factor and phase error generally decrease their values with increasing resolution. In Fig. 2, all three plots have a peak in the range of resolution between 1.5 and 1.25 Å in  $R$  factor and phase error and this range corresponds to that of bond lengths between atoms. The plots for the MSE have the same characteristics as for Sayre's equation on the structure of App 'without' Zn.

Next, the MSE was tested with observed structure magnitudes of App. The result is plotted in Fig. 1. While the anomaly persists, the MSE produces much lower  $R$  factors and phase errors than does Sayre's equation in the high-resolution range.

### Discussion

We have discussed the fact that heavy atoms cause the anomaly in values of  $R$  factor and phase error of Sayre's equation. This fact is not easily understandable if we just consider Sayre's equation alone but if we think in terms of the  $E$  map then it can be better understood.

We take a simple one-dimensional sample structure containing 18 carbon atoms and 2 zinc atoms in a unit cell. The result of the application of Sayre's equation is given in Table 1 and the  $E$  maps and squared  $E$  maps with number of reflexions = 10, 30 and 90 are shown in Fig. 3.

The results in Table 1 show the same character as in the three-dimensional case. The  $R$  factor and phase error have their best values for low resolution. The reason for this can be deduced when we see the  $E$  maps and squared  $E$  maps. The zinc-atom peaks are not so dominant in the low-resolution  $E$  maps; in the  $E$  map with ten reflexions, the heights of the zinc atoms are only three times higher than those of carbon atoms but in the squared  $E$  map they are nine times higher than those of the carbons. The densities of the atoms are also widely spread out, so few areas in which the density is negative are detected. Although the resolution is poor, the  $E$  map and the squared  $E$  map have very similar shapes, which is the reason for the good values of the  $R$  factor and phase errors. The  $R$  factor has the worst value with 30 reflexions; for this resolution the zinc peaks have become very dominant - about five times higher than the peaks of carbon atoms - so that in the squared map they are about 25 times higher than carbon peaks. In addition we see many negative regions in the  $E$  map which create false peaks in the squared map. Consequently

Table 1. Application of Sayre's equation to the sample structure ( $\text{C}_{18}\text{Zn}_2$ )

Number of reflexions	$R$ factor (%)	$\langle \Delta\varphi \rangle_{EG}$
10	0.4	2.2
20	6.3	4.4
30	10.9	7.2
40	8.3	7.2
50	8.8	8.4
60	8.3	9.0
70	7.9	8.9
80	8.2	8.9
90	8.6	8.9

the  $E$  map and squared map are not quite so similar in appearance. For the  $E$  map with 90 reflexions, the zinc-atom peaks become dominant and the negative regions give less-significant false peaks in the squared map.

We conclude that there may be difficulties in the use of the SETF for structures containing heavy atoms although experience with App suggests that one Zn atom in the presence of 300 light atoms can be tolerated. If a way can be found of efficiently incorporating all the terms in the MSE then it may well offer an improvement over the SETF. Actually, with the availability of hard-wired FFT facilities on computers

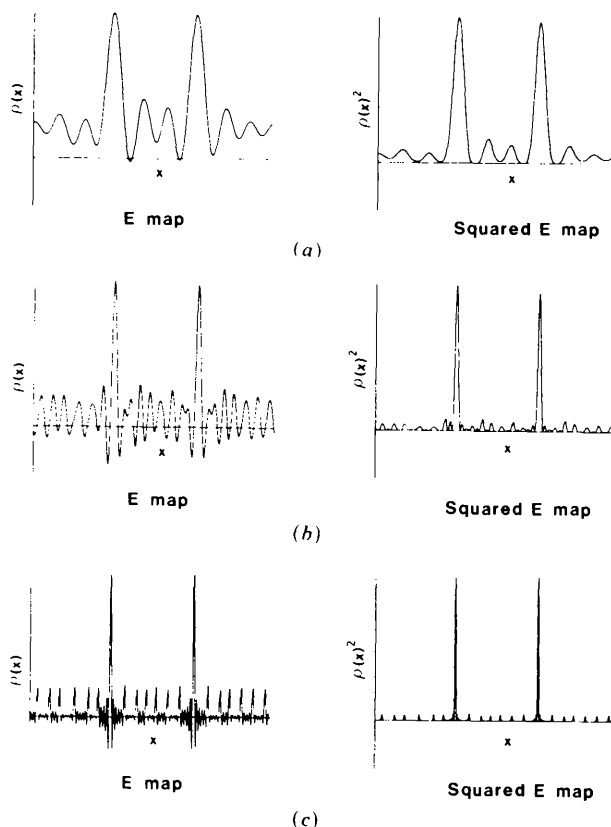


Fig. 3.  $E$  map and squared  $E$  map of the one-dimensional sample structure (a) with 10 reflexions, (b) with 30 reflexions, (c) with 90 reflexions.

able to evaluate a Fourier transform in a fraction of a second, it is possible that reciprocal-space-based direct methods may be completely replaced by real-space density-modification methods in the next decade or so. However, with the current availability of computing facilities in the crystallographic community traditional direct methods still have much to offer.

We thank Mrs E. Dodson for supplying the computer programs. We are also grateful to Dr C. Tate for helpful discussions. Thanks are also due to the Science and Engineering Research Council for the support of our projects. One of us (MS) thanks the Rigaku Corporation of Japan for the award of a research studentship.

### APPENDIX

#### Modified Sayre-equation tangent formula (MSETF)

If we define structure factors corresponding to the square and cube of the  $E$ -map density, compensated for the fall-off with the shape factors, by

$$G'(\mathbf{h}) = [1/g(\mathbf{h})] \sum_{\mathbf{k}} E(\mathbf{k})E(\mathbf{h}-\mathbf{k})$$

$$H'(\mathbf{h}) = [1/h(\mathbf{h})] \sum_{\mathbf{k}} \sum_{\mathbf{l}} E(\mathbf{k})E(\mathbf{l})E(\mathbf{h}-\mathbf{k}-\mathbf{l}), \quad (\text{A-1})$$

the  $R$  factor for the MSE is then given by

$$R = \frac{\sum_{\mathbf{h}} |E(\mathbf{h}) - E(\mathbf{h})_{\text{calc}}|^2}{\sum_{\mathbf{h}} |E(\mathbf{h})|^2}$$

$$= \frac{\sum_{\mathbf{h}} |E(\mathbf{h}) - \{PG'(\mathbf{h}) - QH'(\mathbf{h})\}|^2}{\sum_{\mathbf{h}} |E(\mathbf{h})|^2}, \quad (\text{A-2})$$

where  $P$  and  $Q$  are given by

$$P = p \frac{\sum_{\mathbf{h}} E(\mathbf{h})^* G'(\mathbf{h})}{\sum_{\mathbf{h}} |G'(\mathbf{h})|^2},$$

$$Q = q \frac{\sum_{\mathbf{h}} E(\mathbf{h})^* H'(\mathbf{h})}{\sum_{\mathbf{h}} |H'(\mathbf{h})|^2},$$

where

$$p = \frac{\langle Z_L \rangle Z_H^2 - \langle Z_L^3 \rangle}{(\langle Z_L^2 \rangle Z_H - \langle Z_L^3 \rangle) Z_H} \frac{\sigma_4^{1/2}}{\sigma_2^{1/2}}$$

and

$$q = \frac{\langle Z_L \rangle Z_H - \langle Z_L^2 \rangle}{(\langle Z_L^2 \rangle Z_H - \langle Z_L^3 \rangle) Z_H} \frac{\sigma_6^{1/2}}{\sigma_2^{1/2}}. \quad (\text{A-3})$$

Applying  $P$  and  $Q$  to (A-2), we obtain

$$R = \frac{\sum_{\mathbf{h}} \left| E(\mathbf{h}) - p \left[ \frac{\sum_{\mathbf{k}} E(\mathbf{k})^* G'(\mathbf{k})}{\sum_{\mathbf{k}} |G'(\mathbf{k})|^2} \right] G'(\mathbf{h}) \right.}{\sum_{\mathbf{h}} |E(\mathbf{h})|^2} + \frac{q \left[ \frac{\sum_{\mathbf{k}} E(\mathbf{k})^* H'(\mathbf{k})}{\sum_{\mathbf{k}} |H'(\mathbf{k})|^2} \right] H'(\mathbf{h}) \left. \right|^2}{\sum_{\mathbf{h}} |E(\mathbf{h})|^2}^{-1}. \quad (\text{A-4})$$

Following the method for deriving the Sayre-equation tangent formula (Debaerdemaker, Tate & Woolfson, 1985), we require, for all  $\mathbf{x}$ ,

$$\partial R / \partial \varphi(\mathbf{x}) = 0.$$

Application of this condition to the  $R$  factor gives a new tangent formula which should be effective for structures containing heavy atoms. The form of this modified Sayre-equation tangent formula is

$$\tan \varphi(\mathbf{x}) = [A \operatorname{Im}(t) - B \operatorname{Im}(qa1) - C \operatorname{Im}(qa2) - D \operatorname{Im}(qi) + E \operatorname{Im}(s)]$$

$$\times [A \operatorname{Re}(t) - B \operatorname{Re}(qa1) - C \operatorname{Re}(qa2) - D \operatorname{Re}(qi) + E \operatorname{Re}(s)]^{-1},$$

where

$$t = \sum_{\mathbf{h}} [1/g(\mathbf{x}) + 1/g(\mathbf{h}) + 1/g(\mathbf{x}-\mathbf{h})] E(\mathbf{h})E(\mathbf{x}-\mathbf{h})$$

$$qa1 = \sum_{\mathbf{h}} \sum_{\mathbf{k}} [1/h(\mathbf{x}) + 1/h(\mathbf{k}) + 1/h(\mathbf{x}-\mathbf{h}-\mathbf{k})]$$

$$\times E(\mathbf{h})E(\mathbf{k})E(\mathbf{x}-\mathbf{h}-\mathbf{k})$$

$$qa2 = \sum_{\mathbf{h}} E(\mathbf{x}-\mathbf{h}) [1/g(\mathbf{h})^2] \sum_{\mathbf{k}} E(\mathbf{k})E(\mathbf{h}-\mathbf{k})$$

$$qi = \sum_{\mathbf{h}} E(\mathbf{x}-\mathbf{h}) [1/g(\mathbf{h})h(\mathbf{h})]$$

$$\times \sum_{\mathbf{k}} \sum_{\mathbf{l}} E(\mathbf{k})E(\mathbf{l})E(\mathbf{h}-\mathbf{k}-\mathbf{l})$$

$$s = \sum_{\mathbf{h}} [1/h(\mathbf{h})^2] \sum_{\mathbf{k}} E(\mathbf{k})E(\mathbf{x}-\mathbf{h}-\mathbf{k})$$

$$\times \sum_{\mathbf{l}} \sum_{\mathbf{m}} E(\mathbf{l})E(\mathbf{m})E(\mathbf{h}-\mathbf{l}-\mathbf{m})$$

$$A = p(p-2) \sum_{\mathbf{h}} E(\mathbf{h})^* G'(\mathbf{h}) \sum_{\mathbf{k}} |H'(\mathbf{k})|^2$$

$$- pq \sum_{\mathbf{h}} E(\mathbf{h})^* H'(\mathbf{h}) \sum_{\mathbf{k}} G'(\mathbf{k})^* H'(\mathbf{k}) \quad (\text{A-5})$$

$$B = pq \sum_{\mathbf{h}} E(\mathbf{h})^* G'(\mathbf{h}) \sum_{\mathbf{k}} G'(\mathbf{k})^* H'(\mathbf{k})$$

$$- q(q+2) \sum_{\mathbf{h}} E(\mathbf{h})^* H'(\mathbf{h}) \sum_{\mathbf{k}} |G'(\mathbf{k})|^2$$

$$C = 2p \left\{ \sum_{\mathbf{h}} E(\mathbf{h})^* G'(\mathbf{h}) \left/ \sum_{\mathbf{h}} |G'(\mathbf{h})|^2 \right. \right\}$$

$$\times \left\{ (p-2) \sum_{\mathbf{k}} E(\mathbf{k})^* G'(\mathbf{k}) \sum_{\mathbf{l}} |H'(\mathbf{l})|^2 \right.$$

$$\left. - 2q \sum_{\mathbf{k}} E(\mathbf{k})^* H'(\mathbf{k}) \sum_{\mathbf{l}} G'(\mathbf{l})^* H'(\mathbf{l}) \right\}$$

$$D = 5pq \sum_{\mathbf{h}} E(\mathbf{h})^* G'(\mathbf{h}) \sum_{\mathbf{k}} G'(\mathbf{k})^* H'(\mathbf{k})$$

$$E = 3q \left\{ \sum_{\mathbf{h}} E(\mathbf{h})^* G'(\mathbf{h}) \left/ \sum_{\mathbf{h}} |G'(\mathbf{h})|^2 \right. \right\}$$

$$\times \left\{ 2p \sum_{\mathbf{k}} E(\mathbf{k})^* G'(\mathbf{k}) \sum_{\mathbf{l}} G'(\mathbf{l})^* H'(\mathbf{l}) \right.$$

$$\left. - (q-2) \sum_{\mathbf{k}} E(\mathbf{k})^* H'(\mathbf{k}) \sum_{\mathbf{l}} |G'(\mathbf{l})|^2 \right\}.$$

## References

- DEBAERDEMAEKER, T., TATE, C. & WOOLFSON, M. M. (1985). *Acta Cryst.* **A41**, 286–290.  
 DEBAERDEMAEKER, T., TATE, C. & WOOLFSON, M. M. (1988). *Acta Cryst.* **A44**, 353–357.  
 GLOVER, I., HANEEF, I., PITTS, J., WOOD, S., MOSS, T., TICKLE, I. & BLUNDELL, T. (1983). *Biopolymers*, **22**, 293–304.  
 SAYRE, D. (1952). *Acta Cryst.* **5**, 60–65.  
 WOOLFSON, M. M. (1958). *Acta Cryst.* **11**, 277–283, 393–397.  
 WOOLFSON, M. M. & YAO, J. (1990). *Acta Cryst.* **A46**, 409–413.

*Acta Cryst.* (1991). **A47**, 533–537

### On the Application of One-Wavelength Anomalous Scattering. III. The Wilson-Distribution and MPS Methods

BY A. C. RALPH AND M. M. WOOLFSON

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

(Received 25 June 1990; accepted 25 March 1991)

#### Abstract

This paper describes two methods which break the ambiguity associated with phase determination from one-wavelength anomalous-scattering data when the positions of the anomalous scatterers are known. In the Wilson-distribution method the magnitudes of the contributions of the light atoms are found for each of the alternative phases and the phases are then given weights according to the usual Wilson probability distribution for the magnitudes. In the MPS method the two possible magnitudes of the contributions of the light atoms to the scattering are compared with a theoretical value based on the observed structure magnitudes and the Fourier coefficient of the  $|P_s|$  function [Hao & Woolfson (1989). *Acta Cryst.* **A45**, 794–797]. Once again this leads to a weight for each alternative phase. A best-estimate phase based on the two weights is compared with true phases for two known proteins consisting of 36 and 96 amino acid residues respectively. It is concluded that the quality of the phase estimates is similar to that obtained by other previously published procedures and that the results are much more limited by the magnitude of the anomalous contribution and the data quality than by the actual method used. The methods were then applied to the smaller protein structure using calculated data both with and without added errors. It is concluded that this common procedure for the testing of methods must be done with great care, otherwise unduly optimistic conclusions may be drawn.

#### Introduction

We consider a structure in which there are  $m$  anomalous scatterers, whose positions are known, and  $n$  non-anomalous scatterers in the unit cell. From one-wavelength anomalous-scattering (OAS) data, with known positions for the anomalous scatterers,

there will be two possibilities for the phase,  $\varphi' \pm \Delta\varphi$ , as shown in Fig. 1. The various quantities shown in Fig. 1 are related by

$$|F|^2 = \frac{1}{2}(|F^+|^2 + |F^-|^2) - |F''|^2 \quad (1)$$

and

$$\cos(\Delta\varphi) = (|F^+|^2 - |F^-|^2) / 2|F||F''|. \quad (2)$$

There are various ways in which the ambiguity may be resolved or circumvented (see, for example, Okaya, Saito & Pepinsky, 1955; Kartha, 1961; Blow & Rossmann, 1961; Fan, Han, Qian & Yao, 1984). Other work, which is related to our approach but different in substance, has been done by Wang (1985), who has not so much resolved the ambiguity as solved structures *despite* the ambiguity by his solvent-flattening technique, and by Karle (1985) who has taken

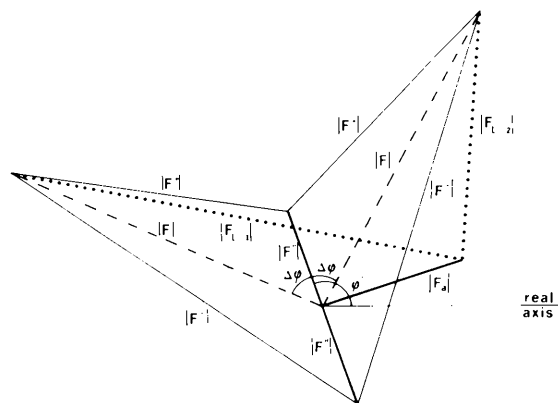


Fig. 1. The following contributions to the scattering are shown:  $|F^+|$ ,  $|F^-|$  the observed structure amplitudes of a Friedel pair;  $|F''|$  the imaginary part of the contribution of the anomalous scatterers;  $|F|$  the real part of the scattering from all scatterers;  $|F_d|$  the total real part of scattering from the anomalous scatterers;  $|F_{t,1}|$ ,  $|F_{t,2}|$  the possible contributions of the light atoms.