References

LENSTRA, A. T. H. (1974). *Acta Cryst.* A**30**, 363–369.

LIPSON, H. & COCHRAN, W. (1953). *The Crystalline State*, Vol. III, 1st ed., p. 250. London: Bell.

PARTHASARATHY, S. & PARTHASARATHI, V. (1972). *Acta Cryst.* A**28**, 426–432.

PARTHASARATHY, S. & PARTHASARATHI, V. (1974). *Acta Cryst.* A**30**, 310–315.

VERSICHEL, W., VAN DE MIEROOP, W. & LENSTRA, A. T. H. (1978). *Acta Cryst.* B**34**, 2643–2645.

WILSON, A. J. C. (1969). *Acta Cryst.* B**25**, 1288–1293.

WILSON, A. J. C. (1976). *Acta Cryst.* A**32**, 53–56.

WILSON, A. J. C. (1977). *Acta Cryst.* A**33**, 523–524.

Acta Cryst. (1978). A**34**, 863–870

On the Application of Phase Relationships to Complex Structures. XIV.* The Additional Use of Statistical Information in Tangent-Formula Refinement

BY S. E. HULL AND M. J. IRWIN

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

(*Received* 22 *February* 1978; *accepted* 9 *May* 1978)

Suitable weighting schemes for use in tangent-formula phase development and refinement are discussed, and a statistically based weighting scheme, which can easily be incorporated in existing computer programs, is proposed. Examples of the use of this method in structure solution and completion are presented for a previously unknown structure and for several structures which had been difficult to solve by other methods. In addition a method is given for subtracting the contribution of heavy atoms from observed $|E|$ values on a statistical basis which is very useful when the presence of such atoms leads to problems in defining the enantiomorph. These methods add greatly to the power of the *MULTAN* computer program.

Introduction

The multisolution method of phase determination using convergence mapping and the tangent formula as embodied in computer programs such as *MULTAN* (Main, 1978) and *SHELX* (Sheldrick, 1976) has been very successful in solving both centrosymmetric and noncentrosymmetric crystal structures containing up to 70 or 80 atoms in the asymmetric unit. Sometimes, however, these programs fail to yield any interpretable molecular fragments. The fact that there exist certain structures which are impossible to solve using the simple tangent formula is well known (see, for example, Lessinger, 1976). A common symptom of this inadequacy in the tangent formula is the tendency for the definition of the enantiomorph to be lost, or indeed never to be defined properly. This may lead either to a trivial solution containing both enantiomorphs or to a 'uranium-type' E map with one large peak and little else. Schenk (1972) has divided the problems caused by enantiomorph definition into two classes. Our experience with *MULTAN* is that such structures are difficult to solve for one of the following reasons.

(1) The enantiomorph is initially defined but upon phase extension or refinement is lost; for example, 3-chloro-1,3,4-triphenylazetidin-2-one (AZET) ($C_{21}H_{16}Cl$-NO, $Pca2_1$, $Z = 8$) (Colens, Declercq, Germain, Putzeys & Van Meerssche, 1974).

(2) Light-atom structures in space groups such as $C2$, $P2_1$, or $P1$ where it is often difficult to define the enantiomorph; for example, prostaglandin E_2 ($C_{20}H_{32}O_5$, $P1$, $Z = 1$) (Edmonds & Duax, 1974*a*,*b*).

(3) Structures containing one heavy atom where the squaring effect of the tangent formula leads to a trivial solution with associated symmetry about the heavy-atom position; for example, the K salt of alborixin (ALB) ($C_{48}H_{83}O_{14}^-.K^+$, $P2_1$, $Z = 2$) (Alléaume, Busetta, Farges, Gachon, Kergomard & Staron, 1975).

Only in the second of these categories is the enantiomorph not well defined at the start of phase determination: for structures in this group there exist several methods making use of enantiomorph-discriminating triple-phase invariants or quartets (Duax & Hauptman, 1972; Busetta, 1976) or enantiomorph-sensitive quartets (Gilmore, 1977), which have met with limited success.

Groups (1) and (3) consist mainly of structures for which the tangent formula is unstable even when the correct phases are used (Lessinger, 1976). This lack of stability appears to be linked to the fact that in normal tangent refinement no explicit use is made of the cosine invariant information or of Sayre's (1952) equation. The next two sections of this paper outline the development of a simple statistical weighting scheme for

---

* Part XIII: Hull (1978).

assessing the reliability of a given phase indication using this type of information. This weighting function greatly enhances the stability of the tangent-refinement process and leads to an important and powerful extension of tangent-formula recycling methods. Finally a statistical method for subtracting unwanted heavy atoms is given and used to solve the example structure of group (3) which had previously been solved only with great difficulty.

1. The tangent formula

In order to understand the limitations of the tangent formula it is necessary to review its development using the probabilistic derivation of Karle & Karle (1966). This derivation affords valuable insight into the limitations of the formula and naturally leads to a further development of its use as a refinement technique. Following Karle & Karle (1966) we consider the conditional probability distribution of the phase $\varphi_{\mathbf{h}}$ given $E_{\mathbf{h}'}$, $E_{\mathbf{h}-\mathbf{h}'}$ and $|E_{\mathbf{h}}|$ for several contributors $\mathbf{h}'$. For a single indication Cochran (1955) and later Karle & Hauptman (1956) showed that the phase $\varphi_{\mathbf{h}}$ is governed by the probability distribution

$$P_{\mathbf{h}'}(\varphi_{\mathbf{h}}) \simeq \frac{\exp[K_{\mathbf{hh}'} \cos(\varphi_{\mathbf{h}'} + \varphi_{\mathbf{h}-\mathbf{h}'} - \varphi_{\mathbf{h}})]}{2\pi I_0(K_{\mathbf{hh}'})}, \quad (1.1)$$

where I_0 is a modified Bessel function,

$$K_{\mathbf{hh}'} = \frac{2\sigma_3}{\sigma_2^{3/2}} |E_{\mathbf{h}'} E_{\mathbf{h}-\mathbf{h}'} E_{\mathbf{h}}|$$

and

$$\sigma_n = \sum_{i=1}^{N} Z_i^n,$$

the summation being over the atomic numbers Z_i of all atoms in the unit cell. For several contributors $\mathbf{h}'$, assumed to be independent, the expression for the overall probability distribution of $\varphi_{\mathbf{h}}$ becomes

$$P(\varphi_{\mathbf{h}}) \simeq \prod_{\mathbf{h}'} P_{\mathbf{h}'}(\varphi_{\mathbf{h}}) = \frac{\exp[\alpha_{\mathbf{h}} \cos(\varphi_{\mathbf{h}} - \beta_{\mathbf{h}})]}{2\pi I_0(\alpha_{\mathbf{h}})}, \quad (1.2)$$

where

$$\alpha_{\mathbf{h}}^2 = \left[\sum_{\mathbf{h}'} K_{\mathbf{hh}'} \cos(\varphi_{\mathbf{h}'} + \varphi_{\mathbf{h}-\mathbf{h}'} - \varphi_{\mathbf{h}})\right]^2 + \left[\sum_{\mathbf{h}'} K_{\mathbf{hh}'} \sin(\varphi_{\mathbf{h}'} + \varphi_{\mathbf{h}-\mathbf{h}'} - \varphi_{\mathbf{h}})\right]^2 \quad (1.3)$$

and

$$\beta_{\mathbf{h}} = \tan^{-1}\left[\frac{\sum K_{\mathbf{hh}'} \sin(\varphi_{\mathbf{h}'} + \varphi_{\mathbf{h}-\mathbf{h}'})}{\sum K_{\mathbf{hh}'} \cos(\varphi_{\mathbf{h}'} + \varphi_{\mathbf{h}-\mathbf{h}'})}\right]. \quad (1.4)$$

The tangent formula results from maximizing (1.2) with respect to $\varphi_{\mathbf{h}}$, which is equivalent to setting $\varphi_{\mathbf{h}} = \beta_{\mathbf{h}}$ in (1.4). $\alpha_{\mathbf{h}}$ is at best only a measure of the upper limit of reliability of the phase indication, as a result of the limited validity of the assertion of independent contributors. This reliability limit can be interpreted in terms of a variance or weight which can be incorporated in subsequent cycles of refinement and phase development (Germain, Main & Woolfson, 1971).

The form of this weighting function is obtained by requiring that a suitably weighted Fourier synthesis $w_{\mathbf{h}} E_{\mathbf{h}}$ give a higher signal-to-noise ratio than a normal $E_{\mathbf{h}}$ synthesis (Blow & Crick, 1959; Germain *et al.*, 1971).

Let us consider a Fourier map

$$\rho(\mathbf{r}) = \sum_{\mathbf{h}} w_{\mathbf{h}} |E_{\mathbf{h}}| \cos(2\pi\mathbf{hr} - \varphi_{\mathbf{h}}); \quad (1.5)$$

at $\mathbf{r}_j$ the centre of the *j*th atom

$$\rho(\mathbf{r}_j) = \sum_{\mathbf{h}} w_{\mathbf{h}} |E_{\mathbf{h}}| \cos(2\pi\mathbf{hr}_j - \varphi_{\mathbf{h}}) \quad (1.6)$$

and let the expected value of the electron density at this position be denoted by

$$\langle\rho(\mathbf{r}_j)\rangle = \sum_{\mathbf{h}} w_{\mathbf{h}} |E_{\mathbf{h}}| \langle\cos(2\pi\mathbf{hr}_j - \varphi_{\mathbf{h}})\rangle. \quad (1.7)$$

For errors $\Delta\varphi_{\mathbf{h}}$ in the phases $\varphi_{\mathbf{h}}$, the error in the electron density $\rho(\mathbf{r}_j)$ will be

$$\Delta\rho(\mathbf{r}_j) = \sum_{\mathbf{h}} w_{\mathbf{h}} |E_{\mathbf{h}}| [\cos(2\pi\mathbf{hr}_j - \varphi_{\mathbf{h}} + \Delta\varphi_{\mathbf{h}}) - \cos(2\pi\mathbf{h}_i \mathbf{r}_j - \varphi_{\mathbf{h}})]. \quad (1.8)$$

If the phase errors are not systematic, the error in $\varphi_{\mathbf{h}}$ will on average be symmetric about $\varphi_{\mathbf{h}}$ and the expected value of $\Delta\rho(\mathbf{r}_j)$ will be zero. However, the mean-square value of $\Delta\rho(\mathbf{r}_j)$ is non-zero and given by

$$\langle\Delta^2\rho(\mathbf{r}_j)\rangle = \sum_{\mathbf{h}} w_{\mathbf{h}}^2 |E_{\mathbf{h}}|^2 \langle\cos^2(2\pi\mathbf{hr}_j - \varphi_{\mathbf{h}}) \times (1 - \cos\Delta\varphi_{\mathbf{h}})^2 + \sin^2(2\pi\mathbf{hr}_j - \varphi_{\mathbf{h}}) \times \sin^2(\Delta\varphi_{\mathbf{h}})\rangle. \quad (1.9)$$

Cochran & Woolfson (1955) derived the probability distribution of $\cos(2\pi\mathbf{hr}_j - \varphi_{\mathbf{h}})$ given the magnitudes $|E_{\mathbf{h}}|$ and showed that the first two moments of this distribution are

$$\langle\cos(2\pi\mathbf{hr}_j - \varphi_{\mathbf{h}})\rangle = \frac{Z_j}{\sigma_2} |E_{\mathbf{h}}| \quad (1.10)$$

$$\langle\cos^2(2\pi\mathbf{hr}_j - \varphi_{\mathbf{h}})\rangle = \frac{1}{2} + \frac{Z_j^2}{4\sigma_2^2} (|E_{\mathbf{h}}|^2 - \sigma_2) \quad (1.11)$$

for structures composed of atoms having form factors of identical shape.

Substituting (1.10) and (1.11) into (1.7) and (1.9) respectively and ignoring terms of $O(1/N)$ in (1.9) results in the following equations

$$\langle\rho(\mathbf{r}_j)\rangle=\frac{Z_j}{\sigma_2}\sum_{\mathbf{h}} w_{\mathbf{h}}|E_{\mathbf{h}}|^2 \tag{1.12}$$

$$\langle\Delta^2\rho(\mathbf{r}_j)\rangle=\sum_{\mathbf{h}} w_{\mathbf{h}}^2|E_{\mathbf{h}}|^2\left(1-\frac{I_1(\alpha_{\mathbf{h}})}{I_0(\alpha_{\mathbf{h}})}\right),$$

where we have used the fact that $\langle\cos(\Delta\varphi_{\mathbf{h}})\rangle = I_1(\alpha_{\mathbf{h}})/I_0(\alpha_{\mathbf{h}})$ from the distribution (1.2). The requirement of a high signal-to-noise ratio then gives the following constraint for $w_{\mathbf{h}}$:

$$\sum w_{\mathbf{h}}\left(1-\frac{I_1(\alpha_{\mathbf{h}})}{I_0(\alpha_{\mathbf{h}})}\right)=\text{constant}. \tag{1.14}$$

Equating $1-[I_1(\alpha_{\mathbf{h}})/I_0(\alpha_{\mathbf{h}})]$ with the variance $V_{\mathbf{h}}$ of the indication shows the correspondence between this result and that of Germain *et al.* (1971), which was only valid for small phase errors. Equation (1.14) is correct for any phase error; in the limit of $\alpha_{\mathbf{h}}$ large, equivalent to small phase errors, $V_{\mathbf{h}}$ varies as $\Delta^2\varphi_{\mathbf{h}}$. It is found that if we interpret the weight $w_{\mathbf{h}}$ to be proportional to $\alpha_{\mathbf{h}}$, then (1.14) holds very well over a large range of values of $\alpha_{\mathbf{h}}$ (see Table 1). To put the weight on an absolute scale we take values for $\alpha_{\mathbf{h}} > 5$ to correspond to a weight of unity. Thus

$$\left.\begin{aligned} w_{\mathbf{h}} &= 0{\cdot}2\,\alpha_{\mathbf{h}}; & \alpha_{\mathbf{h}} &\le 5 \\ w_{\mathbf{h}} &= 1{\cdot}0; & \alpha_{\mathbf{h}} &> 5 \end{aligned}\right\}. \tag{1.15}$$

This is the weighting scheme currently used in the program *MULTAN*. Experimental trials with this and alternative reliability criteria suggest that (1.15) is indeed the optimum form to use for an upper-limit indication.

Table 1. *Values of* $\alpha_{\mathbf{h}}$, $V_{\mathbf{h}}$ *and* $\alpha_{\mathbf{h}}V_{\mathbf{h}}$

$\alpha_{\mathbf{h}}$	$V_{\mathbf{h}}$	$\alpha_{\mathbf{h}}V_{\mathbf{h}}$	$\alpha_{\mathbf{h}}$	$V_{\mathbf{h}}$	$\alpha_{\mathbf{h}}V_{\mathbf{h}}$
0·0	1·000	0·000	4·0	0·137	0·548
0·5	0·757	0·379	4·5	0·120	0·540
1·0	0·554	0·554	5·0	0·107	0·533
1·5	0·404	0·606	6·0	0·088	0·526
2·0	0·302	0·604	7·0	0·075	0·521
2·5	0·235	0·588	8·0	0·065	0·518
3·0	0·190	0·570	9·0	0·057	0·516
3·5	0·159	0·557	10·0	0·051	0·514

Several authors have noted that this type of tangent refinement makes no use of the information contained in the cosine terms of the phase distribution or of the number of contributors to the indication, or explicit use of Sayre's equation, in the algebraic sense. In §2 it is shown how to combine some aspects of this information, in a statistical manner, with a normal tangent-refinement process. This is in contrast to the least-squares methods of refinement proposed by Hauptman (1972) and Main (1977).

2. Statistically weighted tangent refinement

Each individual phase used in an iterative phase expansion or refinement will contain errors which are due to the errors in all other developed phases. This complicated pattern of error combination precludes the use of exact analysis to predict the reliability of any individual phase indication. However, it is still feasible to say something about the expected combinatorial statistics of the phasing process and hence to deduce an associated reliability criterion.

Let us again consider a phase $\varphi_{\mathbf{h}}$ generated from several contributors $E_{\mathbf{h}'}$, $E_{\mathbf{h}-\mathbf{h}'}$, and $|E_{\mathbf{h}}|$; the resultant magnitude of the combination of each $K_{\mathbf{hh}'}$ and its associated invariant phase is given by (1.3). Substituting the expectation values for $\cos(\varphi_{\mathbf{h}'}+\varphi_{\mathbf{h}-\mathbf{h}'}-\varphi_{\mathbf{h}})$ and $\sin(\varphi_{\mathbf{h}'}+\varphi_{\mathbf{h}-\mathbf{h}'}-\varphi_{\mathbf{h}})$, by making use of the result that rotation of each contributor through $-\varphi_{\mathbf{h}}$ does not alter the value of the resultant magnitude, gives for the expectation value of $\alpha_{\mathbf{h}}^2$ (Germain, Main & Woolfson, 1970)

$$\alpha^2_{\mathbf{h}\,(\mathrm{exp.})}=\sum_{\mathbf{h}'}K^2_{\mathbf{hh}'}+\sum_{\substack{\mathbf{h}'\\ \mathbf{h}'\neq\mathbf{h}''}}\sum_{\mathbf{h}''}K_{\mathbf{hh}'}K_{\mathbf{hh}''}$$

$$\times\frac{I_1(K_{\mathbf{h}'})}{I_0(K_{\mathbf{hh}'})}\,\frac{I_1(K_{\mathbf{hh}''})}{I_0(K_{\mathbf{hh}''})}\,. \tag{2.1}$$

If the phase indications are random, the second summation becomes zero and the expected value of $\alpha_{\mathbf{h}}^2$ is

$$\alpha^2_{\mathbf{h}\,(\mathrm{ran.})}=\sum_{\mathbf{h}'}K^2_{\mathbf{hh}'}. \tag{2.2}$$

Finally, the maximum possible value of $\alpha_{\mathbf{h}}^2$, corresponding to a completely consistent set of phase indications, is

$$\alpha^2_{\mathbf{h}\,(\mathrm{max.})}=\left[\sum_{\mathbf{h}'}K_{\mathbf{hh}'}\right]^2. \tag{2.3}$$

Equations (2.1), (2.2) and (2.3) now contain the necessary extra information to make effective use of the cosine-invariant estimates.

If we write Sayre's equation in terms of E's, rather than F's, and multiply both sides by $E_{\bar{\mathbf{h}}}$ we have

$$|E_{\mathbf{h}}|^2=\text{constant}\times\sum_{\mathbf{h}'}E_{\mathbf{h}'}E_{\mathbf{h}-\mathbf{h}'}E_{\bar{\mathbf{h}}}. \tag{2.4}$$

Squaring both sides and putting $K_{\mathbf{hh}'} = |E_{\mathbf{h}'}E_{\mathbf{h}-\mathbf{h}'}E_{\mathbf{h}}|$ gives

$$|E_{\mathbf{h}}|^4 = \text{constant}\left\{\left[\sum_{\mathbf{h}'} K_{\mathbf{hh}'} \cos\left(\varphi_{\mathbf{h}} + \varphi_{\mathbf{h}-\mathbf{h}'} - \varphi_{\mathbf{h}}\right)\right]^2 + \left[\sum_{\mathbf{h}'} K_{\mathbf{hh}'} \sin\left(\varphi_{\mathbf{h}} + \varphi_{\mathbf{h}-\mathbf{h}'} - \varphi_{\mathbf{h}}\right)\right]^2\right\} \quad (2.5)$$

which is similar in form to equation (1.3). We can now rewrite this equation as

$$|E_{\mathbf{h}}|^4 = \text{constant} \times \alpha^2_{\mathbf{h}}. \quad (2.6)$$

It can now be seen that a method which constrains $\alpha^2_{\mathbf{h}}$ to equal $\alpha^2_{\mathbf{h}\,(\text{exp.})}$ during tangent-formula refinement is equivalent to forcing Sayre's equation to be obeyed. This can be achieved by incorporating a weighting scheme into the tangent formula which takes account of the ratio of $\alpha^2_{\mathbf{h}}$ to $\alpha^2_{\mathbf{h}\,(\text{exp.})}$, where $\alpha^2_{\mathbf{h}\,(\text{exp.})}$ is continually recalculated by equation (2.1), taking into account those relationships in use at that stage of the refinement process.

If $\alpha^2_{\mathbf{h}}/\alpha^2_{\mathbf{h}\,(\text{exp.})} = x$ is much less than unity and of the order of $\alpha^2_{\mathbf{h}\,(\text{ran.})}/\alpha^2_{\mathbf{h}\,(\text{exp.})}$ then the reliability of this indication is low and similarly when x tends to $\alpha^2_{\mathbf{h}\,(\text{max.})}/\alpha^2_{\mathbf{h}\,(\text{exp.})}$ an over-consistent phase development is indicated and the assessment of the weight should be correspondingly reduced. For x of the order of unity the phase estimate is probably reliable and should be given a weight close to one. Fig. 1 shows an empirical weighting function developed to make use of these ideas. The form of the function close to $x = 1$ is approximately Gaussian and roughly corresponds to the expected distribution of x. The exact function used is

$$f(x) = k\, e^{-x^2} \int_0^x e^{t^2}\, \mathrm{d}t, \quad (2.7)$$

where k is a scale factor such that the maximum value of the function is unity.

The function $f(x)$ is then used in conjunction with the previously estimated upper reliability limit $w_{\mathbf{h}}$, defined by equation (1.15), to give a new weight $w_{\mathbf{h}\,(\text{stat.})}$ defined by

$$w_{\mathbf{h}\,(\text{stat.})} = \text{maximum}\,[f(x), w_{\mathbf{h}}]. \quad (2.8)$$

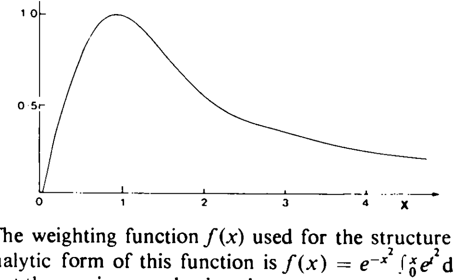

Fig. 1. The weighting function $f(x)$ used for the structure AZET. The analytic form of this function is $f(x) = e^{-x^2} \int_0^x e^{t^2} \mathrm{d}t$ scaled such that the maximum value is unity.

Table 2. *A comparison of* $w_{\mathbf{h}\,(\text{stat.})}$ *and* $w_{\mathbf{h}}$ *for varying numbers of contributors* ($K_{\mathbf{hh}'}$ *is set to* 1·5)

	Number of contributors				
	1	2	5	10	15
Random phase errors					
$w_{\mathbf{h}}$	0·30	0·42	0·67	0·95	1·00
$w_{\mathbf{h}\,(\text{stat.})}$	0·30	0·42	0·67	0·41	0·31
Expected phase errors					
$w_{\mathbf{h}}$	0·30	0·49	1·00	1·00	1·00
$w_{\mathbf{h}\,(\text{stat.})}$	0·30	0·49	1·00	1·00	1·00
Consistent phases					
$w_{\mathbf{h}}$	0·30	0·60	1·00	1·00	1·00
$w_{\mathbf{h}\,(\text{stat.})}$	0·30	0·60	0·53	0·44	0·39

Table 2 contains a comparison of $w_{\mathbf{h}\,(\text{stat.})}$ and $w_{\mathbf{h}}$ for a test case of several numbers of contributors with identical $K_{\mathbf{hh}'}$. From this it is apparent that the normal weighting function used in *MULTAN* rapidly tends to unity as the number of contributors increases, whether the phase indications are random or not, and effectively gives an unweighted refinement.

Since this type of statistically weighted tangent-formula refinement (SWTR) depends critically on the current estimates of weights and phases, these quantities are updated only at the end of each refinement step. Phase extension then proceeds in a stepwise manner. In practice it has been found that the actual size of these steps and the exact form of the function $f(x)$ are not too critical.

The exact shape of the curve used in practice depends on the size of the structure under study. For x less than unity the curve is identical for all structures since $\alpha^2_{(\text{ran.})}$ is much closer to zero than to $\alpha^2_{(\text{exp.})}$. For x greater than unity, x is replaced by

$$x' = 1 + k\,\frac{\sigma_2^{3/2}}{\sigma_3}\,(x - 1), \quad (2.9)$$

where k is chosen so that $x' = x$ for a structure with 200 equal atoms in the unit cell. This has the effect of reducing the weight more rapidly as x increases for small structures, and of reducing it less rapidly for large structures.

Examples

A number of known structures have been solved in tests with SWTR. For one of these, AZET, the best phase sets produced by the standard version of *MULTAN* give only very poor, fragmentary and unrecognizable elements of the structure (Lessinger, 1976). Furthermore, normal tangent-formula refinement with the correct phases is unstable and leads to a mirror solution about one of the Cl atoms. This loss of the enantiomorph also occurs with sets generated by the standard *MULTAN* procedure. In a straightforward run of *MULTAN* with SWTR incorporated, using 270 reflexions and 4000 triple-phase relationships, the solution having the lowest Karle residual (28%) gave an E map which showed 39 out of the total of 48 unique atoms in the top 60 peaks. This result is illustrated in Fig. 2. The absolute figure of merit (Germain *et al.*, 1971) was still rather large (1·58), but was significantly smaller than those obtained by Lessinger (1976).

In all other cases we examined for which the standard *MULTAN* program produces a solution of the structure, SWTR gave an improvement in the image of the structure in the E map.

It thus appears that for certain classes of enantiomorph problem, a statistically weighted refinement technique may sufficiently stabilize the tangent formula to prevent drifts away from the correct solution pathway.

In the tests carried out, the standard figures of merit used by *MULTAN* were used to discriminate between the phase sets. The most reliable of these figures of merit appears to be the residual (Karle & Karle, 1966), although, being based on Sayre's equation, it is linked to the statistical weighting method of phase determination.

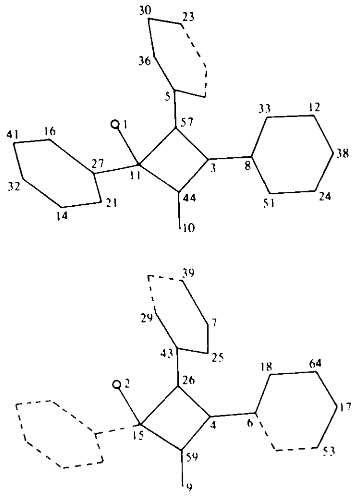

Fig. 2. The molecular structure of AZET obtained by using SWTR in *MULTAN*. The two Cl atoms are indicated by open circles. Missing atoms have their neighbouring bonds shown as broken lines. The absent six-membered ring appears as a smeared-out region of electron density which is impossible to interpret from a straightforward peak search. The numbers refer to the positions of the E-map peaks in a list ordered on height.

3. Statistically weighted Karle recycling

Recycling partial structural fragments by use of the tangent formula, as suggested by Karle (1968), has proved very successful. The technique outlined by Karle (1976) is essentially a highly constrained phase expansion based on a reliable nucleus of well estimated phases. An obvious extension of this method is to assign a weight (equation 2.8) to each phase and allow the phase to develop and refine as in normal phase-expansion procedures. A weight based on a function of $\alpha_{\mathbf{h}}$ such as is given by equation (1.15) would be inadequate, as even reflexions for which $\alpha_{\mathbf{h}}$ was close to $\alpha_{\mathbf{h}\,(\mathrm{ran.})}$ would be assigned phases with unit weights.

The following recycling procedure has been developed to make use of SWTR and has proved to be a powerful extension of Karle's original method.

(1) A nucleus of reliably estimated phases is established using the criteria given by Karle (1976).

(2) The probability distribution given by Sim (1959) is then used to estimate the reliability of the calculated phase indication for each normalized structure factor $E_{\mathbf{h}}$, given a known fragment with the corresponding coefficient $e_{\mathbf{h}}$:

$$P(\varphi_{\mathbf{h}}) = \frac{\exp\left[X_{\mathbf{h}} \cos\left(\varphi_{\mathbf{h}} - \theta_{\mathbf{h}}\right)\right]}{2\pi I_0(X_{\mathbf{h}})}, \qquad (3.1)$$

where $\theta_{\mathbf{h}}$ is the calculated phase of $e_{\mathbf{h}}$,

$$X_{\mathbf{h}} = \frac{2|E_{\mathbf{h}} e_{\mathbf{h}}|}{\left(1 - \dfrac{\sigma_1'}{\sigma_1}\right)} \qquad (3.2)$$

and σ_1' is obtained by summing over the known partial structure. The variance of the above distribution is given by Karle & Karle (1966). Phases from step (1) with $X_{\mathbf{h}}$ less than 2·4, corresponding to a phase error of approximately 45°, are rejected. In addition, any phase with $X_{\mathbf{h}}$ greater than 4·0 is also included in the starting set.

(3) The initial weight given to all phases in the starting set is $I_1(X_{\mathbf{h}})/I_0(X_{\mathbf{h}})$ after Sim (1961).

(4) Two cycles of SWTR are carried out using only the starting-set phases, as recommended by Karle (1976).

(5) Up to five tangent-formula expansion cycles are performed, the weights and phases being updated only at the end of each cycle. A new phase indication is accepted if $\alpha_{\mathbf{h}}^2$ is greater than 2·0. During these expansion cycles all phases with weights greater than 0·9 are fixed to give further stability to the procedure.

(6) Two final cycles of refinement are allowed with all phases included.

(7) Reflexions with weights less than 0·1 are not used in calculating the E map; all other reflexions are given unit weights.

In all test structures examined the SWTR recycling scheme both works more reliably and progresses to the final solution more rapidly than with Karle's method. In practice, only one cycle of SWTR recycling has been required to solve all of the known structures used in the tests, even when starting from very small known fragments. Two of the most interesting examples are discussed below.

AZET was used as a test since Karle recycling starting from the two Cl atoms failed to produce any extra readily interpretable fragments. The results for SWTR recycling are illustrated in Fig. 3; 34 of the 48 atoms were readily identifiable in the E map. Ergocalciferol (Hull, Leban, Main, White & Woolfson, 1976) ($C_{28}H_{44}O$, $P2_12_12_1$, $Z = 8$) was originally solved by recycling from a 22-atom fragment. Starting with a nine-atom fragment from one of the two molecules in the asymmetric unit and using the original E-map coordinates, SWTR recycling gave 46 out of the total of 58 atoms and Karle recycling gave 29 of the 58 atoms. Certainly in this case a subsequent cycle of Karle refinement would have revealed the remainder of the structure; however, the general advantage of the SWTR recycling is that it can allow the phases more freedom to develop, so if the input structure is basically correct the solution is reached more quickly, or if the input structure is incorrect it will soon be lost. Because of this behaviour, subsequent interpretation is much easier.

A further example of the use of SWTR is provided by the solution of a previously unknown structure (factor S from staphylomycin, $C_{46}H_{61}N_7O_{12}$, $P2_12_12_1$, $Z = 4$) (Declercq, Piret & Van Meerssche, 1971) which had resisted many attempts at solution by standard procedures both by the authors and by Germain (1978). The first E map produced by the initial run of the SWTR version of *MULTAN* showed a fragment with chemically sensible geometry from which 12 peaks were selected for use in SWTR recycling. Three cycles of the recycling process gave 16, 26 and 42 correct atoms respectively, and the structure was then completed by weighted Fourier synthesis. Full details of this structure will be published in due course.

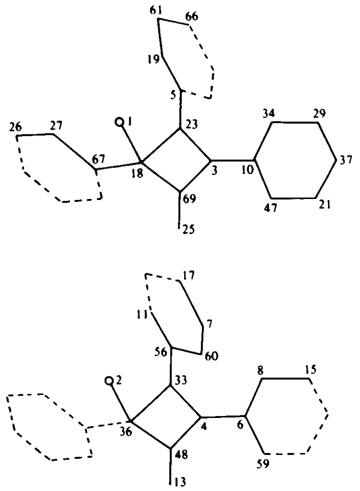

Fig. 3. The result of weighted tangent recycling for AZET starting from the two Cl atoms shown as open circles. The missing six-membered ring appears in a similar manner to that of Fig. 2. The numbers refer to the positions of the E-map peaks in a list ordered on height.

4. Statistically derived difference structures

A feature of many organic structures containing only one heavy atom per asymmetric unit is that tangent-formula refinement leads only to special phase sets consistent with the heavy-atom position, but containing no information about the light atoms (for example, ALB). For structures containing heavy atoms in which this is not a problem, there are several existing techniques for finding the light atoms: such methods include weighted Fourier synthesis, difference Fourier synthesis, and difference recycling (van den Hark, Prick & Beurskens, 1976). The problems with structures which cannot be solved by these methods arise mainly from the fact that the tangent formula implicitly incorporates electron-density squaring equations, with the result that the phases become consistent with the heavy-atom position, but all information about the light atoms is lost. Therefore instead of trying to solve the structure by methods using phases generated from the heavy-atom position, it is better to remove the contribution of the heavy atom from the magnitudes $|E_{\mathbf{h}}|$, and then attempt to solve the remaining light-atom structure.

In Fig. 4 a statistical method of removing the heavy-atom vector $e_{\mathbf{h}}$ from $E_{\mathbf{h}}$ in order to estimate the magnitude of the light-atom vector $E'_{\mathbf{h}}$ is illustrated. Use of the cosine formula gives

$$|E'_{\mathbf{h}}|^2 = |E_{\mathbf{h}}|^2 + |e_{\mathbf{h}}|^2 - 2|E_{\mathbf{h}}e_{\mathbf{h}}| \cos(\varphi_{\mathbf{h}} - \theta_{\mathbf{h}}) \quad (4.1)$$

from which the value of $E'_{\mathbf{h}}$ can be estimated by using

$$\langle\cos(\varphi_{\mathbf{h}} - \theta_{\mathbf{h}})\rangle = \frac{I_1(X_{\mathbf{h}})}{I_0(X_{\mathbf{h}})} \quad (4.2)$$

where $X_{\mathbf{h}}$ is defined by equation (3.2). This gives the estimate

$$\langle|E'_{\mathbf{h}}|^2\rangle = |E_{\mathbf{h}}|^2 + |e_{\mathbf{h}}|^2 - 2|E_{\mathbf{h}}e_{\mathbf{h}}| \frac{I_1(X_{\mathbf{h}})}{I_0(X_{\mathbf{h}})}. \quad (4.3)$$

Reflexions in centrosymmetric zones are treated in the same way as general reflexions since this avoids practical complications and is found to work well in practice. The new structure factors are renormalized and the structure solved by the SWTR version of *MULTAN*.

ALB was used as a test structure since it had remained insoluble even with the SWTR version of *MULTAN*. The structure was originally solved by Busetta (1976) using quartet and triplet structure invariants to reduce the enantiomorph problem, and least-squares refinement (Hauptman, Fisher, Hancock & Norton, 1969) to generate initial phases, followed by a modified tangent-formula refinement. The E map generated by this procedure still contained both enantiomorphs, but could be interpreted and led to the complete structure. The final phases are quoted as being within 30° of the trivial solution.

The statistical difference method was used to generate a set of structure factors for ALB, the potassium ion coordinates being determined by an ordinary run of *MULTAN*, which was then run through the SWTR version of *MULTAN*. Out of the 50 phase sets produced there were several easily interpretable E maps; the best of these showed 52 out of the 63 atoms in the molecule amongst the top 75 peaks. The highest spurious peak was fourteenth in the list produced by the peak-search program. The potassium ion still appeared in the E map, although somewhat reduced in relative size. This result is illustrated in Fig. 5.

Because the heavy-atom contribution to $|E_{\mathbf{h}}|$ is less likely to be overestimated than in the method of van den Hark *et al.* (1976), the possibility that the difference coefficients $|E'_{\mathbf{h}}|$ correspond to a structure with negative electron density in place of the original heavy atom, for which the tangent formula may be unstable, is reduced. Van den Hark *et al.* (1976) assumed that the phases for certain of the reflexions in the light-atom structure are given approximately by the phases calculated from the heavy-atom position and used these phases as a starting point for refinement: the method described in this paper makes no such assumptions, and therefore succeeds in solving structures, such as ALB, for which the earlier method fails because the heavy-atom positions do not define the origin and enantiomorph.

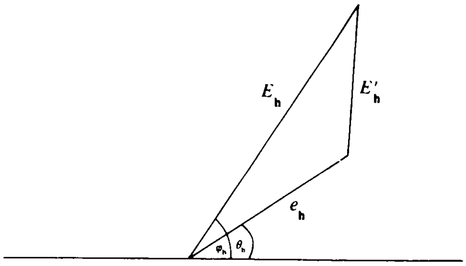

Fig. 4. Removal of the heavy-atom vector.

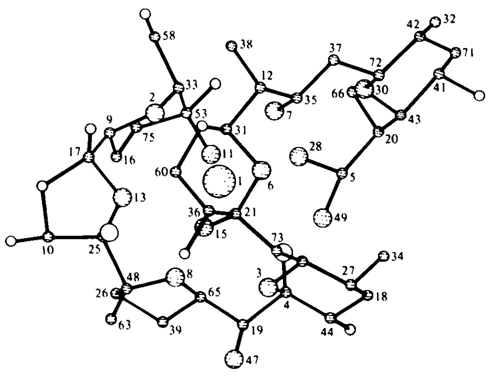

Fig. 5. The molecular structure of ALB. The atoms labelled by the cross-hatched circles were given from SWTR in *MULTAN* operating on the difference structure. The remaining atoms, open circles, were generated from a weighted Fourier synthesis based on the partial structure. The numbers refer to the positions of the E-map peaks in a list ordered on height.

Conclusion

The effect of incorporating a statistical form of Sayre's equation into the tangent formula as a reliability criterion or weight has several theoretical advantages: (1) Random phase indications are given low weights; (2) Phase sets are stopped from becoming too consistent; (3) The difficulties caused by the squaring effect of the tangent formula are reduced; (4) The refinement is stabilized by having an effective weight throughout; (5) In some cases enantiomorph definition is improved. However, this technique has the added advantage of making use of existing computer programs for crystal structure determination and can easily be included in them. The extension of the recycling method of Karle now makes it possible to solve structures such as the smaller nucleotides in one or two cycles, once the phosphorus positions have been determined. Light-atom structures can usually be obtained in one cycle from small structural fragments containing approximately 15% of the scattering matter in the unit cell.

We would like to thank Dr Peter Main and Professor Michael Woolfson for their encouragement of this work. We would also like to thank Drs G. Germain, J. P. Declercq and B. Busetta for supplying data for use in the tests. Finally we gratefully acknowledge financial support from the Science Research Council.

References

Alléaume, M., Busetta, B., Farges, C., Gachon, P., Kergomard, A. & Staron, T. (1975). *Chem. Commun.* pp. 411–412.

Blow, D. M. & Crick, F. (1959). *Acta Cryst.* **12**, 794–802.

Busetta, B. (1976). *Acta Cryst.* A**32**, 139–143.

Cochran, W. (1955). *Acta Cryst.* **8**, 473–478.

Cochran, W. & Woolfson, M. M. (1955). *Acta Cryst.* **8**, 1–12.

Colens, A., Declercq, J. P., Germain, G., Putzeys, J. P. & Van Meerssche, M. (1974). *Cryst. Struct. Commun.* **3**, 119–122.

DECLERCQ, J. P., PIRET, P. & VAN MEERSSCHE, M. (1971). *Acta Cryst.* B**27**, 1276.
DUAX, W. & HAUPTMAN, H. (1972). *Acta Cryst.* B**28**, 2912–2916.
EDMONDS, J. W. & DUAX, W. L. (1974*a*). *Prostaglandins*, **5**, 275–281.
EDMONDS, J. W. & DUAX, W. L. (1974*b*). Abstracts. Am. Crystallogr. Assoc., Spring Meeting, p. 34.
GERMAIN, G. (1978). Private communication.
GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1970). *Acta Cryst.* B**26**, 274–285.
GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1971). *Acta Cryst.* A**27**, 368–376.
GILMORE, C. J. (1977). *Acta Cryst.* A**33**, 712–716.
HARK, TH. E. M. VAN DEN, PRICK, P. & BEURSKENS, P. T. (1976). *Acta Cryst.* A**32**, 816–821.
HAUPTMAN, H. (1972). *Crystal Structure Determination: The Role of the Cosine Seminvariants*, pp. 232–236. New York: Plenum.
HAUPTMAN, H., FISHER, J., HANCOCK, H. & NORTON, D. (1969). *Acta Cryst.* B**25**, 811–814.
HULL, S. E. (1978). *Acta Cryst.* A**34**, 38–42.
HULL, S. E., LEBAN, I., MAIN, P., WHITE, P. S. & WOOLFSON, M. M. (1976). *Acta Cryst.* B**32**, 2374–2381.
KARLE, J. (1968). *Acta Cryst.* B**24**, 182–186.
KARLE, J. (1976). *Crystallographic Computing Techniques*, edited by F. R. AHMED, pp. 155–164. Copenhagen: Munksgaard.
KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635–651.
KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* **21**, 849–859.
LESSINGER, L. (1976). *Acta Cryst.* A**32**, 538–550.
MAIN, P. (1977). Private communication.
MAIN, P. (1978). *MULTAN 78. A Program System for the Automatic Solution of Crystal Structures from X-ray Diffraction Data.* Univ. of York, England.
SAYRE, D. (1952). *Acta Cryst.* **5**, 60–65.
SCHENK, H. (1972). *Acta Cryst.* A**28**, 412–422.
SHELDRICK, G. M. (1976). *SHELX*. Program for Crystal Structure Determination. Univ. of Cambridge, England.
SIM, G. A. (1959). *Acta Cryst.* **12**, 813–815.
SIM, G. A. (1961). *Computing Methods and the Phase Problem in X-Ray Crystal Analysis*, edited by R. PEPINSKY, J. M. ROBERTSON & J. C. SPEAKMAN, pp. 227–235. Oxford: Pergamon Press.

Acta Cryst. (1978). A**34**, 870–883

On the Application of Phase Relationships to Complex Structures.
XV. Magic Determinants

BY D. J. TAYLOR, M. M. WOOLFSON AND P. MAIN

*Department of Physics, University of York, Heslington, York YO*1 5*DD, England*

(*Received* 15 *March* 1978; *accepted* 2 *May* 1978)

A magic determinant is a Karle–Hauptman determinant in which the phases of reflexions forming the elements are expressed in magic-integer representation. By means of appropriate algorithms, which are described, magic determinants may be found, up to order 30, which have a high content of large E's. The value of the magic determinant depends on the values of the independent variables (usually two) in terms of which the magic-integer-represented phases are defined. Maxima are sought in a map of magic-determinant values and these give trial values for the constituent phases. These phases may be refined by maximizing one or more of the largest eigenvalues of the matrix of elements. Applications to five different structures are described and the usefulness and limitations of the magic-determinant concept are discussed. It is concluded that Karle–Hauptman determinants, chosen in the way described, provide a very discriminating figure of merit. However, the structure factors they contain usually provide a rather poor base for subsequent phase development and this limits the usefulness of the magic-determinant approach.

Introduction

Phase-determining methods based on magic integers were described in papers VII and VIII of this series (White & Woolfson, 1975; Declercq, Germain & Woolfson, 1975). Further developments in the use of magic integers were described in paper X where a general overview of the *MAGLIN* system was presented (Woolfson, 1977). A theory of magic integers and a critical examination of the use of magic integers in phase determination were given by Main (1977, 1978) in papers XI and XII respectively. This paper describes in detail the part of *MAGLIN* which makes use of the properties of Karle–Hauptman determinants. Magic