Table 1. *Special relationships between dispersion of specimen crystal and dispersion of monochromator*

θ_c	$\Delta\omega$	$\Delta 2\theta$	Comments
$\arctan(0{\cdot}5 \tan\theta_M)$	$-\frac{1}{2}k\tan\theta_M$	0	All wavelengths enter the detector in parallel but are not diffracted simultaneously
$\arctan(0{\cdot}6 \tan\theta_M)$	$-\frac{2}{5}k\tan\theta_M$	$\frac{1}{5}k\tan\theta_M$	Minimum wavelength dispersion in $\Delta\omega$, $\Delta 2\theta$ space
θ_M	0	$k\tan\theta_M$	All wavelengths enter the detector simultaneously but as a divergent beam

This corresponds to the 'focusing' condition (Arndt & Willis, 1966) where all wavelengths, λ_2 to λ_1, diffract simultaneously. Note, however, that while the dispersion is zero in respect of $\Delta\omega$ it is not zero in respect of $\Delta 2\theta$. When one goes beyond s_2s_1, the sequence of intersection with the reflecting circle inverts to t_1t_2, *i.e.* first t_1 then t_2. The special relationships are summarized in Table 1.

I am grateful to Drs S. L. Mair and A. W. Stevenson for critical and helpful comments on the manuscript. Also to Dr B. T. M. Willis for very considerable contributions to textual clarification.

References

ARNDT, U. W. & WILLIS, B. T. M. (1966). *Single Crystal Diffractometry.* Cambridge Univ. Press.
MATHIESON, A. MCL. (1985). *Acta Cryst.* A41, 309-316.
SCHOENBORN, B. P. (1983). *Acta Cryst.* A39, 315-321.
ZACHARIASEN, W. H. (1945). *Theory of X-ray Diffraction in Crystals.* New York: John Wiley.

Acta Cryst. (1985). A41, 605-613

From a Partial to the Complete Crystal Structure.
II. The Procedure and Its Applications

BY M. CAMALLI

Istituto di Strutturistica Chimica 'G. Giacomello', Area della Ricerca del CNR, CP 10,
00016 *Monterotondo Stazione, Roma, Italy*

C. GIACOVAZZO

Dipartimento Geomineralogico, Università di Bari, Palazzo Ateneo, 70121 *Bari, Italy*

AND R. SPAGNA

Istituto di Strutturistica Chimica 'G. Giacomello', Area della Ricerca del CNR, CP 10,
00016 *Monterotondo Stazione, Roma, Italy*

(*Received* 14 *January* 1985; *accepted* 25 *June* 1985)

Abstract

A multisolution procedure, based on the probabilistic formulas obtained by Giacovazzo [*Acta Cryst.* (1983). A39, 685-692] is described, which aims at recovering the complete crystal structure from a partial one. A new weighted tangent formula develops starting phases: the correct solution among others is found by means of two revised figures of merit. The procedure is successfully applied to some practical cases.

Symbols and abbreviations

Throughout the paper a number of symbols will find frequent application. For most of them the reader is referred to the first paper of this series (Giacovazzo, 1983), from now on referred to as paper I. Other symbols not used in I are listed below.

$\Sigma_p, \Sigma_q, \Sigma_N = \sum f_i^2$	The summation is extended to the p, q, N atoms. Atomic thermal factors are included.
$\Sigma_p^0, \Sigma_q^0, \Sigma_N^0 = \sum^0 f_i^2$	The summation is extended to the p, q, N atoms. Atomic thermal factors are excluded.
$F_{p,\mathbf{h}}^0$	Structure factor for the partial structure. Atomic thermal factors are not considered.
$I_\mathbf{h}$	$\lvert F_\mathbf{h}\rvert^2$ on an arbitrary scale.
$E'_\mathbf{h}$	Pseudo-normalized structure factor with vectorial index $\mathbf{h}$ defined by $E'_\mathbf{h} = F_\mathbf{h}/\Sigma_q^{1/2}$.
$E'_{p,\mathbf{h}}$	Pseudo-normalized structure factor of the partial structure with p atoms in the unit cell, defined by $E'_{p,\mathbf{h}} = F_{p,\mathbf{h}}/\Sigma_q^{1/2}$.

0108-7673/85/060605-09$01.50

$R'_\mathbf{h}$, $R'_{p,\mathbf{h}}$	Moduli of $E'_\mathbf{h}$ and $E'_{p,\mathbf{h}}$ respectively.
$D_1(x)=I_1(x)/I_0(x)$	$I_i(x)$ is the modified Bessel function of order i.
$G_\mathbf{h}$	$2R'_\mathbf{h}R'_{p,\mathbf{h}}$.
cs	Centrosymmetric.
ncs	Non-centrosymmetric.

Other locally used symbols are defined in the text.

1. The probabilistic background

In paper I a probabilistic theory is described aimed at recovering the full structure from a partial one. A group of p atoms was assumed to be correctly positioned while the other $q=N-p$ atomic positions define the primitive random variables. According to Main (1976) and Heinerman (1977), the normalized structure factor was defined by

$$E_\mathbf{h}=F_\mathbf{h}/\langle|F_\mathbf{h}|^2\rangle^{1/2}_{\text{p.r.v.}},$$

where $\langle|F_\mathbf{h}|^2\rangle_{\text{p.r.v.}}$ denotes the average of $|F_\mathbf{h}|^2$, the variable being the primitive random variables. Then [see (I.2)]

$$E_\mathbf{h}=F_\mathbf{h}/[|F_{p,\mathbf{h}}|^2+\Sigma_q]^{1/2}.$$

By analogy it was assumed in I that

$$E_{p,\mathbf{h}}=F_{p,\mathbf{h}}/[|F_{p,\mathbf{h}}|^2+\Sigma_q]^{1/2}$$
$$E_{q,\mathbf{h}}=F_{q,\mathbf{h}}/[|F_{p,\mathbf{h}}|^2+\Sigma_q]^{1/2}.$$

It was explicitly noted that $E_{p,\mathbf{h}}$ is not a variable and $|E_{p,\mathbf{h}}|\le 1$ always, while $E_{q,\mathbf{h}}$ is a variable but it is not a normalized structure factor. Indeed, $\langle|E_{q,\mathbf{h}}|^2\rangle\le 1$.

A basic result of I was (I.B1), from which (I.23) and (I.24) were derived, estimating $\varphi_\mathbf{h}$ in $P\bar{1}$ and $P1$ respectively *via* one or more triplet terms. The use of these equations is rather difficult because they require, for each triplet, the calculation of the coefficients $b(\mathbf{h})$, $b(\mathbf{k})$, $b(\mathbf{h}+\mathbf{k})$, c. A remarkable simplification in (I.B1) is obtained by replacing E by the pseudo-normalized structure factor E'. If $\mathbf{k}$ is changed into $-\mathbf{k}$ (to be consistent with standard notation) and E is replaced by E' then (I.B1) reduces to (A1). Subsequent calculations lead to new expressions for (I.23) and (I.24). In particular, the conditional probability that the sign of $E'_\mathbf{h}$ is plus in cs space groups is given by

$$P^+(E'_\mathbf{h})\simeq 0\cdot5+0\cdot5\tanh\left\{R'_\mathbf{h}\left[E'_{p,\mathbf{h}}+q^{-1/2}\sum_\mathbf{k}(E'_\mathbf{k}-E'_{p,\mathbf{k}})(E'_{\mathbf{h}-\mathbf{k}}-E'_{p,\mathbf{h}-\mathbf{k}})\right]\right\}. \qquad (1)$$

For the ncs case we obtain

$$P(\varphi_\mathbf{h}|\cdots)=M[\varphi_\mathbf{h};\,\theta_\mathbf{h},\alpha_\mathbf{h}]\equiv[2\pi I_0(\alpha)]^{-1}\exp[\alpha\cos(\varphi_\mathbf{h}-\theta_\mathbf{h})], \qquad (2)$$

where

$$\alpha_\mathbf{h}^2=\alpha_1'^2+\alpha_2'^2, \qquad (3)$$

$$\begin{aligned}\alpha'_1=2R'_\mathbf{h}\Big\{&R'_{p,\mathbf{h}}\cos\varphi_{p,\mathbf{h}}\\&+q^{-1/2}\sum_\mathbf{k}[R'_\mathbf{k}R'_{\mathbf{h}-\mathbf{k}}\cos(\varphi_\mathbf{k}+\varphi_{\mathbf{h}-\mathbf{k}})\\&-R'_{p,\mathbf{k}}R'_{\mathbf{h}-\mathbf{k}}\cos(\varphi_{p,\mathbf{k}}+\varphi_{\mathbf{h}-\mathbf{k}})\\&-R'_\mathbf{k}R'_{p,\mathbf{h}-\mathbf{k}}\cos(\varphi_\mathbf{k}+\varphi_{p,\mathbf{h}-\mathbf{k}})\\&+R'_{p,\mathbf{k}}R'_{p,\mathbf{h}-\mathbf{k}}\cos(\varphi_{p,\mathbf{k}}+\varphi_{p,\mathbf{h}-\mathbf{k}})]\Big\}\\=2R'_\mathbf{h}\Big\{&\mathscr{R}\Big[E'_{p,\mathbf{h}}+q^{-1/2}\sum_\mathbf{k}(E'_\mathbf{k}-E'_{p,\mathbf{k}})\\&\times(E'_{\mathbf{h}-\mathbf{k}}-E'_{p,\mathbf{h}-\mathbf{k}})\Big]\Big\},\end{aligned} \qquad (4)$$

$$\begin{aligned}\alpha'_2=2R'_\mathbf{h}\Big\{&R'_{p,\mathbf{h}}\sin\varphi_{p,\mathbf{h}}\\&+q^{-1/2}\sum_\mathbf{k}[R'_\mathbf{k}R'_{\mathbf{h}-\mathbf{k}}\sin(\varphi_\mathbf{k}+\varphi_{\mathbf{h}-\mathbf{k}})\\&-R'_{p,\mathbf{k}}R'_{\mathbf{h}-\mathbf{k}}\sin(\varphi_{p,\mathbf{k}}+\varphi_{\mathbf{h}-\mathbf{k}})\\&-R'_\mathbf{k}R'_{p,\mathbf{h}-\mathbf{k}}\sin(\varphi_\mathbf{k}+\varphi_{p,\mathbf{h}-\mathbf{k}})\\&+R'_{p,\mathbf{k}}R'_{p,\mathbf{h}-\mathbf{k}}\sin(\varphi_{p,\mathbf{k}}+\varphi_{p,\mathbf{h}-\mathbf{k}})]\Big\}\\=2R'_\mathbf{h}\Big\{&\mathscr{I}\Big[E'_{p,\mathbf{h}}+q^{-1/2}\sum_\mathbf{k}(E'_\mathbf{k}-E'_{p,\mathbf{k}})\\&\times(E'_{\mathbf{h}-\mathbf{k}}-E'_{p,\mathbf{h}-\mathbf{k}})\Big]\Big\},\end{aligned} \qquad (5)$$

$$\tan\theta_\mathbf{h}=\alpha'_2/\alpha'_1. \qquad (6)$$

In (4) and (5) $\mathscr{R}$ and $\mathscr{I}$ stand for 'real part of' and 'imaginary part of' respectively.

A more complete insight into the approach may be obtained by analysing the formulas in the light of one of the most used procedures aimed at recovering the complete structure from a partial one, the *DIRDIF* method by Prick, Beurskens & Gould (1983). Analogies and differences may be schematically described in the following way:

(1) *DIRDIF* assumes that the structure factor of the unknown part of the structure may be estimated by a (sometimes rough) difference of moduli: $|F_q|\simeq||F|-|F_p||$. A phase is associated with each $(|F|-|F_p|)$, which is refined cycle by cycle. In our procedure we do not need to estimate $|F_q|$: in fact the formulas depend on both F and F_p. In particular, the phase φ is refined cycle by cycle while φ_p is a constant.

(2) *DIRDIF* procedure aims to estimate $\varphi_{q,\mathbf{h}}$ whose reliability is given by the parameter (function

of moduli only)

$$K_{\mathbf{h},\mathbf{k}}=2N^{-1/2}|(|E_{\mathbf{h}}|-|E_{p,\mathbf{h}}|)(|E_{\mathbf{k}}|-|E_{p,\mathbf{k}}|)$$
$$\times(|E_{\mathbf{h}-\mathbf{k}}|-|E_{p,\mathbf{h}-\mathbf{k}}|).$$

Our method estimates $\varphi_{\mathbf{h}}$ whose reliability parameter is

$$2|E_{\mathbf{h}}||E'_{p,\mathbf{h}}+q^{-1/2}(E'_{\mathbf{k}}-E'_{p,\mathbf{k}})(E'_{\mathbf{h}-\mathbf{k}}-E'_{p,\mathbf{h}-\mathbf{k}})|,$$

which depends both on the moduli $|E'_{\mathbf{h}}|$, $|E'_{p,\mathbf{h}}|$, $|E'_{\mathbf{k}}|$, $|E'_{p,\mathbf{k}}|$, $|E'_{\mathbf{h}-\mathbf{k}}|$, $|E'_{p,\mathbf{h}-\mathbf{k}}|$ and on the phases $\varphi_{p,\mathbf{h}}$, $\varphi_{\mathbf{k}}$, $\varphi_{p,\mathbf{k}}$, $\varphi_{\mathbf{h}-\mathbf{k}}$, $\varphi_{p,\mathbf{h}-\mathbf{k}}$. Using $\varphi_{\mathbf{h}}$ instead of $\varphi_{q,\mathbf{h}}$ offers a further advantage: at the end of the phasing process the inspection of the E map immediately indicates if prior information has been lost during phase extension and refinement.

(3) When several triplets are available the phase $\varphi_{q,\mathbf{h}}$ is estimated in *DIRDIF* by the parameter

$$\alpha_{\mathbf{h}}^2=\left\{\sum_{\mathbf{k}} K_{\mathbf{h},\mathbf{k}}\cos\left(\varphi_{q,\mathbf{k}}+\varphi_{q,\mathbf{h}-\mathbf{k}}\right)\right\}^2$$
$$+\left\{\sum_{\mathbf{k}} K_{\mathbf{h},\mathbf{k}}\sin\left(\varphi_{q,\mathbf{k}}+\varphi_{q,\mathbf{h}-\mathbf{k}}\right)\right\}^2.$$

During phase refinement the values φ_q vary but moduli $K_{\mathbf{h},\mathbf{k}}$ remain fixed.

According to (3) the reliability parameter of $\varphi_{\mathbf{h}}$ depends on

$$2|E_h|\left|E_{p,\mathbf{h}}+\sum_{\mathbf{k}} q^{-1/2}(E_{\mathbf{k}}-E_{p,\mathbf{k}})(E_{\mathbf{h}-\mathbf{k}}-E_{p,\mathbf{h}-\mathbf{k}})\right|.$$

During the phase refinement, the phased difference terms $(E_{\mathbf{k}}-E_{p,\mathbf{k}})$ and $(E_{\mathbf{h}-\mathbf{k}}-E_{p,\mathbf{h}-\mathbf{k}})$ change both their moduli and their phases.

Further analogies and differences will appear when the weighted tangent procedure and the new figures of merit are described.

The new forms of the distributions allow the immediate understanding of their properties. We note:

(*a*) the coefficients $b(\mathbf{h})$, $b(\mathbf{k})$, $b(\mathbf{h}-\mathbf{k})$ do not appear in the new expressions, which are now easily applicable.

(*b*) the triplet contribution is of order $q^{-1/2}$ [it is replaced by $(\sigma_3/\sigma_2^{3/2})_q$ if atoms are not equal]. It is worthwhile observing that q is just the number of independent random variables in our problem [by analogy, the triplet contribution in Cochran's (1955) relationship is of order $N^{-1/2}$, and N is there the number of independent random variables].

(*c*) E' and E'_p are not normalized structure factors: indeed (the averages are assumed to be made over $\mathbf{h}$)

$$\langle|E'|^2\rangle=1+\Sigma_p/\Sigma_q,\quad \langle|E'_p|^2\rangle=\Sigma_p/\Sigma_q,$$

so that $\langle|E'|^2\rangle$ is always larger than unity, while $\langle|E'_p|^2\rangle$ is larger or smaller than unity according to whether Σ_p is larger or smaller than Σ_q. The larger the *a priori* information, the smaller $\langle|E'-E'_p|^2\rangle$ is.

(*d*) equation (2) reduces to Sim's (1959) formula if triplet contribution is not taken into account.

(*e*) the best estimate for $\varphi_{\mathbf{h}}$, *i.e.* $\theta_{\mathbf{h}}$, is the phase of the complex vector

$$E'_{p,\mathbf{h}}+q^{-1/2}\Sigma_{\mathbf{k}}(E'_{\mathbf{k}}-E'_{p,\mathbf{k}})(E'_{\mathbf{h}-\mathbf{k}}-E'_{p,\mathbf{h}-\mathbf{k}}). \quad (7)$$

The larger its modulus α is, the larger the expected accuracy of the estimation. According to (7), the vectorial differences $(E'_{\mathbf{k}}-E'_{p,\mathbf{k}})$ and $(E'_{\mathbf{h}-\mathbf{k}}-E'_{p,\mathbf{h}-\mathbf{k}})$ do influence the value of $\theta_{\mathbf{h}}$ [not only the scalar differences $(R'_{\mathbf{k}}-R'_{p,\mathbf{k}})$ and $(R'_{\mathbf{h}-\mathbf{k}}-R'_{p,\mathbf{h}-\mathbf{k}})$].

(*f*) if $p\rightarrow 0$, then $q\rightarrow N$, $E'_p\rightarrow 0$ and (1) and (2) reduce to the classical Cochran-Woolfson's (1955) and Karle & Karle's (1966) relationships respectively.

(*g*) the larger is p, the less important in the average is the triplet contribution compared with Sim's contribution. In particular, because of point (*c*), the triplet contribution vanishes when $q\rightarrow 0$. Then $R'_{\mathbf{h}}$ and $R'_{p,\mathbf{h}}$ are both infinite and (2) approximates the Dirac δ function $\delta(\varphi_{\mathbf{h}}-\varphi_{p,\mathbf{h}})$.

(*h*) From (2)–(6) the following probabilistic relation is suggested:

$$E'_{\mathbf{h}}\simeq E'_{p,\mathbf{h}}+q^{-1/2}\sum_{\mathbf{k}}(E'_{\mathbf{k}}-E'_{p,\mathbf{k}})(E'_{\mathbf{h}-\mathbf{k}}-E'_{p,\mathbf{h}-\mathbf{k}}), \quad (8)$$

which may be considered as a generalized Sayre's (1952) equation emphasizing the fact that part of the structure is known. When $q=0$ (8) reduces to the trivial identity $E'_{\mathbf{h}}=E'_{p,\mathbf{h}}=\infty$. When $q=N$ then $E'_p=0$ and (8) reduces to the classical Sayre's equation. When $q\neq 0$, N the prior information introduces new algebraic or probabilistic constraints so as to recentre $E'_{\mathbf{h}}$ around $E'_{p,\mathbf{h}}$. Algebraic reasons supporting (8) and the new generalized Sayre's equation in terms of F are given in Appendix *B*.

In § 2 a practical procedure is described aiming at exploiting in an effective way (2)–(6). In § 3 practical applications are described.

2. The automatic procedure

The *SIR* program (Cascarano, Giacovazzo, Polidori, Spagna & Viterbo, 1982) has been modified in order to take advantage of the available *a priori* information. The basic steps are described below.

2.1. *Normalization and pseudo-normalization*

If all the atoms of the structure are supposed to have the same isotropic temperature factor B, then

$$\langle I_{\mathbf{h}}\rangle_{\text{shell}}=K\{\langle|F^0_{p,\mathbf{h}}|^2+\Sigma^0_q\rangle_{\text{shell}}\}\exp\left(-2Bs^2\right), \quad (9)$$

where $s=\sin\theta/\lambda$. In (9) 'shell' denotes any reciprocal-space region chosen for the Wilson method and K is the scale factor. From (9) the Wilson plot is obtained *via*

$$\log\{\langle I_{\mathbf{h}}\rangle_{\text{shell}}/\langle|F^0_{p,\mathbf{h}}|^2+\Sigma^0_q\rangle_{\text{shell}}\}=\log K-2Bs^2, \quad (10)$$

giving the K and B values.

Often, estimates of the temperature factors of the atoms in the partial structure are available. If the observed amplitudes are on the absolute scale (for example, *via* a previous Wilson plot), we could write

$$\langle|F_{\mathbf{h}}|^2\rangle_{\rm shell}=\langle|F_{p,\mathbf{h}}|^2+\Sigma_q^0\exp(-2B_qs^2)\rangle_{\rm shell},$$

from which

$$\log\{\langle|F_{\mathbf{h}}|^2-|{\rm F}_{\rm p,\mathbf{h}}|^2\rangle_{\rm shell}/\langle\Sigma_q^0\rangle_{\rm shell}\}=-2B_qs^2 \tag{11}$$

is obtained, giving the average temperature factor B_q of the unknown part of the structure. The E values are then obtained by means of

$$|E_{\mathbf{h}}|^2=I_{\mathbf{h}}/\{K[|F_{p,\mathbf{h}}|^2+\Sigma_q]\} \tag{12}$$

and the E' values by

$$|E'_{\mathbf{h}}|^2=I_{\mathbf{h}}/(K\Sigma_q). \tag{13a}$$

Accordingly,

$$|E'_{p,\mathbf{h}}|^2=I_{p,\mathbf{h}}/(K\Sigma_q). \tag{13b}$$

In (12), (13*a*) and (13*b*), $K=1$ if (11) is used.

A further method for obtaining E and E' is that described by Gould, van den Hark & Beurskens (1975), which minimizes the quantity

$$\sum(G-G_p-G_q)^2, \tag{14}$$

where

$$G=I/\Sigma_N^0,$$

$$G_p=K\langle|F_p^0|^2\exp(-2B_ps^2)/\Sigma_N^0\rangle,$$

$$G_q=K\langle\Sigma_q^0\exp(-2B_qs^2)/\Sigma_N^0\rangle.$$

Equation (14) is particularly useful when: (*a*) B_p is expected to be markedly different from B_q (*i.e.* for p heavy atoms); (*b*) all the p atoms are expected to have a nearly equal temperature factor.

On the other hand, (11) allows p different temperature factors for the p atoms but uses the scale factor provided by a previous Wilson plot under different conditions (*i.e.* an overall isotropic temperature factor). Such an assumption proved not critical in our experimental tests. Therefore, we implemented in the *SIR* package the method using (11).

2.2. *Starting estimates of the phases*

If $\varphi_{p,\mathbf{h}}$ is known then the distribution of $\varphi_{\mathbf{h}}$ is the Von Mises distribution (Sim, 1959)

$$M[\varphi_{\mathbf{h}};\,\varphi_{p,\mathbf{h}},\,G_{\mathbf{h}}],\quad\text{where } G_{\mathbf{h}}=2R'_{\mathbf{h}}R'_{p,\mathbf{h}}. \tag{15}$$

Because of the prior information the complete set $\{R',R'_p,\varphi_p\}$ is known. It is then possible to integrate (A1) with respect to φ_2 and φ_3 and calculate the most probable value $\theta_{\mathbf{h}}$ of $\varphi_{\mathbf{h}}$ given $\{R',R'_p,\varphi_p\}$. We obtain

$$P(\varphi_{\mathbf{h}}|\{R'_{\mathbf{h}},R'_k,R'_{\mathbf{h}-\mathbf{k}},\varphi_{p,\mathbf{k}},\varphi_{p,\mathbf{h}-\mathbf{k}}\})$$
$$=M[\varphi_{\mathbf{h}};\,\theta_{\mathbf{h}},\,X_{\mathbf{h}}], \tag{16}$$

where

$$\tan\theta_{\mathbf{h}}=\sum_j P_j\sin\theta_j/\sum_j P_j\cos\theta_j=A_{\mathbf{h}}/B_{\mathbf{h}}, \tag{17}$$

$$X_{\mathbf{h}}=\{A_{\mathbf{h}}^2+B_{\mathbf{h}}^2\}^{1/2}, \tag{18}$$

$$P_1=G_{\mathbf{h}},$$

$$\theta_1=\varphi_{p,\mathbf{h}},$$

$$\underset{j\neq 1}{P_j}=2q^{-1/2}R'_{\mathbf{h}}R'_{p,\mathbf{k}_j}R'_{p,\mathbf{h}-\mathbf{k}_j},$$

$$\underset{j\neq 1}{\theta_j}=\varphi_{p,\mathbf{k}_j}+\varphi_{p,\mathbf{h}-\mathbf{k}_j}.$$

As we see, the knowledge of the complete set $\{\varphi_p\}$ generates, in addition to Sim's contribution, a non-negligible contribution of order $q^{-1/2}$ arising from triplets of type $(\varphi_{\mathbf{h}}-\varphi_{p,\mathbf{k}}-\varphi_{p,\mathbf{h}-\mathbf{k}})$.

The above procedure deals with $\varphi_{\mathbf{k}}$ and $\varphi_{\mathbf{h}-\mathbf{k}}$ as they were randomly distributed around 0 and 2π, in spite of the fact that, according to Sim, $\varphi_{p,\mathbf{k}}$ and $\varphi_{p,\mathbf{h}-\mathbf{k}}$ are the expected values of $\varphi_{\mathbf{k}}$ and $\varphi_{\mathbf{h}-\mathbf{k}}$ respectively. This suggests that $\varphi_{\mathbf{h}}$ may be estimated by introducing into (2) the assumptions

$$\varphi_{\mathbf{k}}\simeq\varphi_{p,\mathbf{k}},\qquad\varphi_{\mathbf{h}-\mathbf{k}}\simeq\varphi_{p,\mathbf{h}-\mathbf{k}}. \tag{19}$$

Then (2) reduces to

$$P(\varphi_{\mathbf{h}}|\{R',R'_p,\varphi_p,\varphi\equiv\varphi_p\})=M[\varphi_{\mathbf{h}};\,\theta_{\mathbf{h}},\,\beta_{\mathbf{h}}], \tag{20}$$

where

$$\beta_{\mathbf{h}}^2=\beta_1^2+\beta_2^2$$

$$\beta_1=2R'_h\left\{R'_{p,\mathbf{h}}\cos\varphi_{p,\mathbf{h}}\right.$$
$$+q^{-1/2}\sum_{\mathbf{k}}(R'_{\mathbf{k}}-R'_{p,\mathbf{k}})(R'_{\mathbf{h}-\mathbf{k}}-R'_{p,\mathbf{h}-\mathbf{k}})$$
$$\left.\times\cos(\varphi_{p,\mathbf{k}}+\varphi_{p,\mathbf{h}-\mathbf{k}})\right\}$$

$$\beta_2=2R'_{\mathbf{h}}\left\{R'_{p,\mathbf{h}}\sin\varphi_{p,\mathbf{h}}\right.$$
$$+q^{-1/2}\sum_{\mathbf{k}}(R'_{\mathbf{k}}-R'_{p,\mathbf{k}})(R'_{\mathbf{h}-\mathbf{k}}-R'_{p,\mathbf{h}-\mathbf{k}})$$
$$\left.\times\sin(\varphi_{p,\mathbf{k}}+\varphi_{p,\mathbf{h}-\mathbf{k}})\right\}$$

$$\tan\theta_{\mathbf{h}}=\beta_2/\beta_1.$$

However, (19) is allowed only for a limited number of phases. In order to avoid the numerous unreliable assumptions (19) destroying the information provided by the subset of the more reliable ones, the Sim weight is associated with each relation (19). Then (2) reduces to

$$M[\varphi_{\mathbf{h}};\,\theta_{\mathbf{h}},\,\alpha_{\mathbf{h}}], \tag{21}$$

where

$$\alpha_{\mathbf{h}}^2 = \alpha_1^2 + \alpha_2^2 \qquad (22)$$

$$\alpha_1 = G_{\mathbf{h}} \cos \varphi_{p,\mathbf{h}} + q^{-1/2} \sum_{\mathbf{k}} Q_{\mathbf{k}} \cos (\varphi_{p,\mathbf{k}} + \varphi_{p,\mathbf{h}-\mathbf{k}})$$

$$\alpha_2 = G_{\mathbf{h}} \sin \varphi_{p,\mathbf{h}} + q^{-1/2} \sum_{\mathbf{k}} Q_{\mathbf{k}} \sin (\varphi_{p,\mathbf{k}} + \varphi_{p,\mathbf{h}-\mathbf{k}})$$

$$Q_{\mathbf{k}} = Q_1 D_1(G_{\mathbf{k}}) D_1(G_{\mathbf{h}-\mathbf{k}}) - Q_2 D_1(G_{\mathbf{h}-\mathbf{k}}) - Q_3 D_1(G_{\mathbf{k}}) + Q_4$$

$$Q_1 = 2R'_{\mathbf{h}} R'_{\mathbf{k}} R'_{\mathbf{h}-\mathbf{k}} \qquad Q_2 = 2R'_{\mathbf{h}} R'_{p,\mathbf{k}} R'_{\mathbf{h}-\mathbf{k}}$$

$$Q_3 = 2R'_{\mathbf{h}} R'_{\mathbf{k}} R'_{p,\mathbf{h}-\mathbf{k}} \qquad Q_4 = 2R'_{\mathbf{h}} R'_{p,\mathbf{k}} R_{p,\mathbf{h}-\mathbf{k}}$$

$$\tan \theta_{\mathbf{h}} = \alpha_2 / \alpha_1. \qquad (23)$$

The $\theta_{\mathbf{h}}$ values provided by (23) and characterized by large values of $\alpha_{\mathbf{h}}$ are expected to estimate the true $\varphi_{\mathbf{h}}$ value more accurately than Sim's estimates. Thus they may act as a starting set for subsequent phase expansions.

On the other hand, phases $(\theta_{\mathbf{k}} + \theta_{\mathbf{h}-\mathbf{k}})$ arising from the partial structure are mutually inconsistent when $\theta_{\mathbf{h}}$ is characterized by a small $\alpha_{\mathbf{h}}$. Then $\theta_{\mathbf{h}}$ is expected to differ markedly from the true value $\varphi_{\mathbf{h}}$.

The above considerations suggest that:

(*a*) the original- and enantiomorph-defining phases should be chosen among the $\varphi_{\mathbf{h}}$ values with largest $\alpha_{\mathbf{h}}$;

(*b*) symbolic phases, varying according to the magic integer method (Main, 1978) may be chosen among the $\theta_{\mathbf{h}}$ phases characterized by the largest inconsistency of the $\theta_{\mathbf{k}} + \theta_{\mathbf{h}-\mathbf{k}}$ pairs. In this way information additional to that provided by partial structure is used in view of: (1) enlarging the starting set; (2) making easier the phase refinement procedure in difficult cases (*e.g.* when the partial structure is centrosymmetric and the complete structure is not, or when the partial structure suffers pseudotranslational symmetry *etc.*).

2.3. *The starting set*

The R' and R'_p values are obtained according to (13*a*) and (13*b*). Then the values $\alpha_{\mathbf{h}}$ given by (22) are calculated for each $\mathbf{h}$ and arranged in decreasing order of α. The NUMK θ values given by (23) and characterized by $\alpha \geq 10$ are assumed to be reliable estimates of the true phases φ and constitute the starting set. In order to avoid loss of the origin they are not allowed to vary during the phasing procedure except in the last two cycles. On the other hand, the condition NUMK$\leq$100 is introduced in order to release too heavy restrictions in the refinement process.

If the origin and the enantiomorph are not fixed by the NUMK reflexions then suitable additional phases with large α values are introduced in the starting set. Specifically, the enantiomorph is assumed to be fixed if at least one triplet is found among the NUMK reflexions whose $(\theta_{\mathbf{h}} - \theta_{\mathbf{k}} - \theta_{\mathbf{h}-\mathbf{k}})$ value differs by more than 30° from 0 or π. The magic integer symbolic phases to be introduced in the starting set (there are four in default conditions) are chosen among those having the largest values of (see Prick *et al.*, 1983, for a related procedure)

$$\Delta\alpha_{\mathbf{h}} = (\alpha_{\mathbf{h}})_{\max} - \alpha_{\mathbf{h}},$$

where

$$(\alpha_{\mathbf{h}})_{\max} = G_{\mathbf{h}} + q^{-1/2} \sum_{\mathbf{k}} |Q_{\mathbf{k}}|.$$

$\alpha_{\mathbf{h}} = (\alpha_{\mathbf{h}})_{\max}$ only when in (22) $\varphi_{p,\mathbf{k}_j} + \varphi_{p,\mathbf{h}-\mathbf{k}_j} \equiv \varphi_{p,\mathbf{k}_i} + \varphi_{p,\mathbf{h}-\mathbf{k}_i} \equiv \varphi_{p,\mathbf{h}}$, whatever i and j may be.

The criterion of maximizing $\Delta\alpha$ instead of minimizing α is suggested by the following considerations. The reflexions with the smallest values of α are likely to have true phases markedly different from θ. However, even if this information is very precious it is of limited usefulness in the phasing procedure because these reflexions are not likely to be closely connected with the phases to be determined. It is therefore preferable to choose as variable phases those having the largest values of $\Delta\alpha$ because they are likely to have a sufficiently large number of contributions and relatively unreliable estimates.

2.4. *The weighted tangent formula*

In order to ensure that poorly determined phases have little effect on the determination of other phases the weighted tangent formula (24) has been formulated:

$$\tan \theta_{\mathbf{h}} = \alpha''_2 / \alpha''_1, \qquad (24)$$

where

$$\begin{aligned}\alpha''_1 = G_1 \cos \varphi_{p,\mathbf{h}} + q^{-1/2} \sum_{\mathbf{k}} \{ & Q_1 w_{\mathbf{k}} w_{\mathbf{h}-\mathbf{k}} \cos (\varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}}) \\ & - Q_2 w_{\mathbf{h}-\mathbf{k}} \cos (\varphi_{p,\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}}) \\ & - Q_3 w_{\mathbf{k}} \cos (\varphi_{\mathbf{k}} + \varphi_{p,\mathbf{h}-\mathbf{k}}) \\ & + Q_4 \cos (\varphi_{p,\mathbf{k}} + \varphi_{p,\mathbf{h}-\mathbf{k}}) \} \qquad (25)\end{aligned}$$

$$\begin{aligned}\alpha''_2 = G_1 \sin \varphi_{p,\mathbf{h}} + q^{-1/2} \sum_{\mathbf{k}} \{ & Q_1 w_{\mathbf{k}} w_{\mathbf{h}-\mathbf{k}} \sin (\varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}}) \\ & - \ldots \} \qquad (26)\end{aligned}$$

and

$$\alpha_{\mathbf{h}}^2 = \alpha''^2_2 + \alpha''^2_1.$$

A phase indication is accepted and incorporated in the known set if $\alpha^2 >$ CUT, where CUT is a threshold varying cycle by cycle. At the beginning CUT = CUTIN, where CUTIN is the minimum α^2 value associated by (22) with the starting set. Cycle by cycle CUT is reduced by the factor 0·65. At the end of each cycle the average $\langle \alpha^2 \rangle$ for the accepted phases is calculated and used in the succeeding cycle to obtain the weight

$$w_{\mathbf{h}}=(\alpha_{\mathbf{h}}^2/\langle\alpha^2\rangle)^{1/4}. \tag{27}$$

w cannot be larger than unity and smaller than 0·15. The above scheme has been introduced because of the large α values usually involved in the process from the first cycles. While usual weighting schemes could lead to an unweighted situation, with $w=1$ for any reflexions, in our scheme w does not depend on α but on a ratio whose denominator changes cycle by cycle. The effect is that the weights of the phases are more conveniently arranged in the range 0·15–1.

2.5. *The figures of merit*

The correct one among other solutions may be found by appropriate figures of merit. A revision of the traditional figures is needed according to the following considerations:

(*a*) criteria based on the maximization of $\sum_{\mathbf{h}}\alpha_{\mathbf{h}}$ usually are not convenient here. To show that, let us observe that $\alpha_{\mathbf{h}}$ is a maximum if the phase of the triplet contribution $\sum_{\mathbf{k}} E_{q,\mathbf{k}}E_{q,\mathbf{h}-\mathbf{k}}$ coincides with $\varphi_{p,\mathbf{h}}$. Since $q^{-1/2}\sum_{\mathbf{k}} E_{q,\mathbf{k}}E_{q,\mathbf{h}-\mathbf{k}}$ is expected to converge (see Appendix *B*) to $E_{q,\mathbf{h}}$, $\sum_{\mathbf{h}}\alpha_{\mathbf{h}}$ is then a maximum when, for every $\mathbf{h}$, $\varphi_{q,\mathbf{h}}\simeq\varphi_{p,\mathbf{h}}$, which cannot correspond to the true structure.

(*b*) a residue criterion may be introduced according to the formula

$$R'_{\mathbf{h}\mathrm{calc}}\cos\varphi_{\mathbf{h}}-R'_{p,\mathbf{h}}\cos\varphi_{p,\mathbf{h}}\simeq Sq^{-1/2}\langle U\rangle \tag{28}$$

$$R'_{\mathbf{h}\mathrm{calc}}\sin\varphi_{\mathbf{h}}-R'_{p,\mathbf{h}}\sin\varphi_{p,\mathbf{h}}\simeq Sq^{-1/2}\langle V\rangle, \tag{29}$$

where U and V are the summations of order $q^{-1/2}$ that appear in (25) and (26) respectively, and

$$\langle U\rangle=U/n,\ \langle V\rangle=V/n.$$

n is the number of contributions in U and V.

The scaling factor S is found by adding the squares of (28) and (29):

$$R'^2_{\mathbf{h}\mathrm{calc}}+R'^2_{p,\mathbf{h}}-2R'_{\mathbf{h}\mathrm{calc}}R'_{p,\mathbf{h}}\cos(\varphi_{\mathbf{h}}-\varphi_{p,\mathbf{h}})$$
$$\simeq S^2q^{-1}(\langle U\rangle^2+\langle V\rangle^2).$$

Now we write (28) and (29) as

$$R'_{\mathbf{h}\mathrm{calc}}\cos\varphi_{\mathbf{h}}=R'_{p,\mathbf{h}}\cos\varphi_{p,\mathbf{h}}+Sq^{-1/2}\langle U\rangle,$$

$$R'_{\mathbf{h}\mathrm{calc}}\sin\varphi_{\mathbf{h}}=R'_{p,\mathbf{h}}\sin\varphi_{p,\mathbf{h}}+Sq^{-1/2}\langle V\rangle,$$

from which $R'^2_{\mathbf{h}\mathrm{calc}}$ may be obtained. The residue

$$R_p=\sum_{\mathbf{h}}|R'_{\mathbf{h}}-R'_{\mathbf{h}\mathrm{calc}}|/\sum_{\mathbf{h}}R'_{\mathbf{h}} \tag{30}$$

is then used as a figure of merit.

(*c*) the ψ_0 criterion, which in the standard version may be written as

$$\psi_0=\sum_{\mathbf{h}}\left|\sum_{\mathbf{k}}E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}\right|, \tag{31}$$

in our case becomes

$$\psi_{0p}=\sum_{\mathbf{h}}\left|E'_{p,\mathbf{h}}+q^{-1/2}\sum_{\mathbf{k}}(E'_{\mathbf{k}}-E'_{p,\mathbf{k}})(E'_{\mathbf{h}-\mathbf{k}}-E'_{p,\mathbf{h}-\mathbf{k}})\right|.$$

Table 1. *Crystal data for test structures*

$N_{\mathrm{a.u.}}$ is the number of non-hydrogen atoms in the asymmetric unit. N_{refl} is the number of independent reflexions (with largest values of R') introduced in the procedure.

Compound	PROL	RIFOL	NAFTO
Chemical formula	$C_{26}H_{40}N_4O_7$	$C_{39}H_{49}NO_{13}$	$C_{42}H_{50}ClNO_9.2H_2O$
$N_{\mathrm{a.u.}}$	37	53	55
Z	2	2	4
Space group	$P2_1$	$P2_1$	$P2_12_12_1$
a(Å)	12·768	11·860	12·565
b(Å)	10·924	9·140	14·410
c(Å)	10·064	20·423	24·239
α(°)	90	90	90
β(°)	95·70	90·72	90
γ(°)	90	90	90
N_{refl}	300	400	400

(*d*) standard figures of merit based on negative quartets and one-phase and two-phase semivariants can usefully be applied without modifications. Their power will certainly be magnified when formulas are available explicitly exploiting prior information on the partial structure.

3. Applications

All calculations were performed by the procedure described in § 2 using default executional parameters. The test structures are listed in Table 1: they are denoted by the code names PROL (Colapietro, De Santis, Nocilli, Palleschi & Spagna, 1985), RIFOL (Brufani *et al.*, 1985), NAFTO (Keller-Shierlein *et al.*, 1985). All these structures were *MULTAN* resistant and were solved by application of the *SIR* package (Nunzi *et al.*, 1984). PROL and RIFOL were chosen as examples of equal-atom structures, NAFTO is taken as an example of an organic molecule with one medium-heavy atom (chlorine) whose position in favourable circumstances may be found by a sharpened Patterson synthesis. Displays of the molecules of PROL, RIFOL and NAFTO are shown in Figs. 1, 2 and 3 respectively. Descriptions of the molecular fragments in known positions for the three structures are given in Table 2. Usual techniques of successive Fourier syntheses do not allow the recovery from the fragment of the complete crystal structure.

Pseudo-normalized structure factors were calculated according to (11) and (13) and starting estimates of the phases were obtained according to (21). The use of (21) may be appreciated by looking at Table 3. The entries of the column 'overall' give the values in degrees of $\langle|\Delta\varphi|_1\rangle=\langle|\varphi^0_{\mathbf{h}}-\varphi^0_{p,\mathbf{h}}|\rangle$ and $\langle|\Delta\varphi|_2\rangle=\langle|\varphi^0_{\mathbf{h}}-\theta^0_{\mathbf{h}}|\rangle$ calculated over all the N_{refl} reflexions. $\theta_{\mathbf{h}}$ is seen to estimate $\varphi_{\mathbf{h}}$ more accurately that $\varphi_{p,\mathbf{h}}$.

Table 2. *Description of the starting fragments for PROL, RIFOL and NAFTO*

er is the ratio (number of *a priori* located electrons)/(total number of electrons).
$R = [\sum ||F_{\text{obs}}| - |F_{\text{calc}}||]/[\sum |F_{\text{obs}}|]$ gives the discrepancy index calculated for the given fragments.

Compound	Starting fragment	*er*	*R*
PROL	O(3), N(1), C(2), C(3), C(4), C(5), C(6)	0·16	0·670
RIFOL	O(1), O(12), C(1), C(2), C(3), C(4), C(9), C(10)	0·11	0·668
NAFTO	Cl	0·04	0·674

However, the most important effect of (21) is that reflexions are ranked according to $\alpha_{\mathbf{h}}$ in (22) more conveniently than according to $G_{\mathbf{h}}$ in (15). In Table 3 the values of $\langle|\Delta\varphi|_1\rangle$ and $\langle|\Delta\varphi|_2\rangle$ are shown for the first 50, first 100, last 100, last 50 reflexions ranked according to $G_{\mathbf{h}}$ and $\alpha_{\mathbf{h}}$. It is seen that the origin and enantiomorph phases have to be chosen among the reflexions with the largest values of α.

The introduction of variable phases in the starting set according to § 2.3 and the application of the weighted tangent formula described in § 2.4 lead to several possible solutions from which the correct one has to be selected. In this view we used the figures of merit (30) and (21). Their effectiveness clearly stands out in Table 4, where the R_p, ψ_{0p} and $\langle|\Delta\varphi|\rangle = \langle|\varphi^0_{\text{true}} - \varphi^0_{\text{calc}}|\rangle$ are given for each solution for RIFOL.

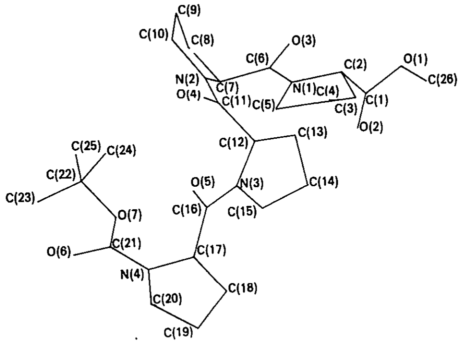

Fig. 1. Schematic display of the PROL molecule.

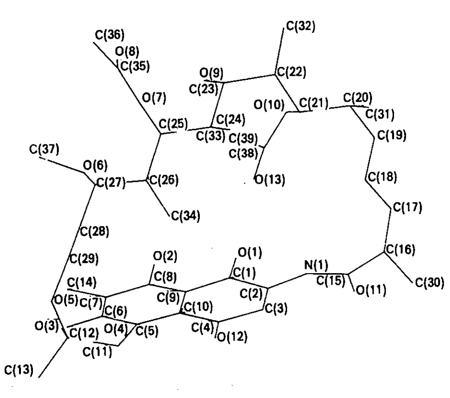

Fig. 2. Schematic display of the RIFOL molecule.

Table 3. *The values* $\langle|\Delta\varphi|_1\rangle = \langle|\varphi_{\mathbf{h}} - \varphi_{p,\mathbf{h}}|\rangle$ *and* $\langle|\Delta\varphi|_2\rangle = \langle|\varphi_{\mathbf{h}} - \theta_{\mathbf{h}}|\rangle$ *calculated for PROL, RIFOL and NAFTO in various circumstances (see text)*

Compound	Overall	First 50	First 100	Last 100	Last 50
PROL	$\langle\|\Delta\varphi\|_1\rangle$ 43	21	24	64	72
	$\langle\|\Delta\varphi\|_2\rangle$ 35	20	20	55	71
RIFOL	$\langle\|\Delta\varphi\|_1\rangle$ 43	21	20	77	91
	$\langle\|\Delta\varphi\|_2\rangle$ 42	15	18	83	94
NAFTO	$\langle\|\Delta\varphi\|_1\rangle$ 48	25	34	65	70
	$\langle\|\Delta\varphi\|_2\rangle$ 45	23	31	64	78

Table 4. *Serial number n,* $\langle|\Delta\varphi|\rangle = \langle|\varphi^0_{true} - \varphi^0_{calc}|\rangle$, $100R_p$ *and* ψ_0 *values for the various solutions obtained for RIFOL*

n	$\langle\|\Delta\varphi\|\rangle$	R_p	ψ_{0p}	*n*	$\langle\|\Delta\varphi\|\rangle$	R_p	ψ_{0p}
1	47	31	190	2	16	26	157
3	48	31	189	4	16	25	157
5	48	31	195	6	47	31	189
7	48	31	192	8	49	31	195
9	48	31	193	10	16	26	156
11	16	26	157	12	16	25	157
13	49	31	195	14	48	31	194
15	46	31	187	16	15	26	157
17	15	26	156	18	48	31	190
19	49	32	195	20	48	31	196
21	41	33	179	22	47	31	182
23	47	31	187	24	49	32	202
25	15	26	155	26	46	31	187
27	47	31	175	28	48	31	194
29	47	31	191	30	16	26	154
31	16	26	155	32	48	31	191
33	48	31	195	34	47	31	191
35	45	31	182	36	15	26	156
37	15	26	155	38	47	31	189
39	48	31	195	40	48	31	195

It is also seen that several good solutions are obtained in the 40, indicating that a reserve of power is still disposable.

For PROL, RIFOL and NAFTO the most probable solution was chosen whose Fourier transform gave (see Figs. 1, 2, 3):

(*a*) for PROL: all non-hydrogen atoms except O(6) and C(23);

(*b*) for RIFOL: 33 additional atoms. The other ones [O(11), C(30), C(13), C(17), C(29), C(37), C(35), O(8), C(36), C(32), C(33), C(34)] were easily

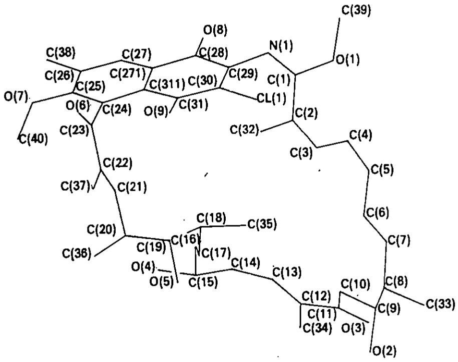

Fig. 3. Schematic display of the NAFTO molecule.

provided by standard techniques.

(*c*) for NAFTO: a fragment of 20 additional atoms easily recognizable by a routine peak search program [C(22), C(23), C(24), C(25), C(40), O(7), C(27), C(271), C(28), C(29), C(30), C(31), C(311), O(9), N(1), C(32), C(2), C(1), O(1), C(39)]. At this stage the discrepancy index was $R=0{\cdot}51$ and an observed Fourier synthesis did not give additional atoms. The 21-atom fragment was used as prior information for a new cycle of our procedure and a new fragment of 12 atoms was immediately recognized by the peak search program [C(26), C(4), C(5), C(6), C(7), C(8), C(9), C(10), C(11), O(2), O(3), C(33)]. Recovering the complete structure was then a trivial task.

4. Conclusions

The procedure described above may be considered as an image reconstruction process working in reciprocal space. Inded, prior information on the position of a structural fragment is transferred into the reciprocal space in order to obtain the F_p factors, which, together with diffraction intensities, are exploited for recovering the complete crystal structure. The procedure has a probabilistic nature but agrees well with an algebraic modified form of Sayre's equation described in Appendix *B*. The method has been applied to cases in which recovering of the crystal structure is not trivial. In particular, fragments containing electron fractions of 0·16, 0·11 and 0·04 (corresponding to relative scattering power of 0·18, 0·15 and 0·11) proved to be sufficient input for the method. Since more solutions in the multisolution process contemporaneously contain additional structural information we feel that the method may work also in more critical situations.

APPENDIX *A*

By introducing the change of variables $R\rightarrow R'$ and by assuming $\varphi_2=\varphi_{\mathbf{k}}$, $\varphi_3=\varphi_{\mathbf{h}-\mathbf{k}}$, $\varphi_{p2}=\varphi_{p,\mathbf{k}}$, $\varphi_{p3}=\varphi_{p,\mathbf{h}-\mathbf{k}}$, (I.*B*1) becomes

$$\begin{aligned}
&P(\varphi_1,\varphi_2,\varphi_3|R'_i,R'_{pi},\varphi_{pi},i=1,2,3)\\
&\quad\simeq L^{-1}\exp\Big\{\sum_{i=1}^{3}2R'_iR'_{pi}\cos(\varphi_i-\varphi_{pi})\\
&\qquad+2q^{-1/2}[R'_1R'_2R'_3\cos(\varphi_1-\varphi_2-\varphi_3)\\
&\qquad-R'_{p1}R'_2R'_3\cos(\varphi_{p1}-\varphi_2-\varphi_3)\\
&\qquad-R'_1R'_{p2}R'_3\cos(\varphi_1-\varphi_{p2}-\varphi_3)\\
&\qquad-R'_1R'_2R'_{p3}\cos(\varphi_1-\varphi_2-\varphi_{p3})\\
&\qquad+R'_{p1}R'_{p2}R'_3\cos(\varphi_{p1}-\varphi_{p2}-\varphi_3)\\
&\qquad+R'_{p1}R'_2R'_{p3}\cos(\varphi_{p1}-\varphi_2-\varphi_{p3})\\
&\qquad+R'_1R'_{p2}R'_{p3}\cos(\varphi_1-\varphi_{p2}-\varphi_{p3})]\Big\}. \qquad (A1)
\end{aligned}$$

By the same change of variable (I.*A*1) reduces to

$$\begin{aligned}
&P(E'_1|E'_2,E'_3,E'_{p1},E'_{p2},E'_{p3})\\
&\quad\simeq(2\pi)^{-1/2}\exp\{-\tfrac{1}{2}(E'_1-E'_{p1})^2\\
&\qquad+q^{-1/2}(E'_1-E'_{p1})(E'_2-E'_{p2})(E'_3-E'_{p3})\}. \qquad (A2)
\end{aligned}$$

APPENDIX *B*

For equal atoms fully resolved from one another the following relations hold:

$$T[\rho_q(\mathbf{r})]=F_{q,\mathbf{h}}$$

$$T[\rho_q^2(\mathbf{r})]=\theta_{\mathbf{h}}F_{q,\mathbf{h}}=V^{-1}\sum_{\mathbf{k}}F_{q,\mathbf{k}}F_{q,\mathbf{h}-\mathbf{k}},$$

where *T* stands for 'Fourier transform of'. Then

$$F_{q,\mathbf{h}}=(V\theta_{\mathbf{h}})^{-1}\sum_{\mathbf{k}}F_{q,\mathbf{k}}F_{q,\mathbf{h}-\mathbf{k}}. \qquad (B1)$$

Since

$$\begin{aligned}
F_{q,\mathbf{h}}&=F_{\mathbf{h}}-F_{p,\mathbf{h}}\\
F_{q,\mathbf{k}}&=F_{\mathbf{k}}-F_{p,\mathbf{k}} \qquad (B2)\\
F_{q,\mathbf{h}-\mathbf{k}}&=F_{\mathbf{h}-\mathbf{k}}-F_{p,\mathbf{h}-\mathbf{k}},
\end{aligned}$$

(*B*1) becomes

$$F_{\mathbf{h}}=F_{p,\mathbf{h}}+(V\theta_{\mathbf{h}})^{-1}\sum_{\mathbf{k}}(F_{\mathbf{k}}-F_{p,\mathbf{k}})(F_{\mathbf{h}-\mathbf{k}}-F_{p,\mathbf{h}-\mathbf{k}}), \qquad (B3)$$

which is the new form of the Sayre equation. Equation (*B*3) should not be confused with results by Fan Hai-fu (1965) and Krabbendam & Kroon (1971) whose formulas modify Sayre's equation for the presence of heavy atoms. Our equation (8) may be considered as the probabilistic counterpart of (*B*3). We note:

(*a*) (*B*3) strictly holds even when the known part of the structure contains unequal atoms. This condition does not coincide with Sayre's condition, according to which all the atoms in the structure have to be equal;

(*b*) the relations (*B*2) can be considered as constraints for the F_q factors. Indeed the moduli and phases of the factors F_p and the moduli of the factors *F* are *a priori* known. Such constraints make (*B*3) more useful than the classical Sayre's equation when prior information is available.

References

Brufani, M., Cellai, L., Cerrini, S., Lamba, D., Segre, A., Nunzi, A., Burla, M. C. & Polidori, G. (1985). To be published.

Cascarano, G., Giacovazzo, C., Polidori, G., Spagna, R. & Viterbo, D. (1982). *Acta Cryst.* A**38**, 663-670.

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

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

Colapietro, M., De Santis, P., Nocilli, F., Palleschi, A. & Spagna, R. (1985). *Acta Cryst.* C**41**, 126-130.

Fan Hai-fu (1965). *Acta Phys. Sin.* **21**, 1105-1113.
Giacovazzo, C. (1983). *Acta Cryst.* A**39**, 685-692.
Gould, R. O., van den Hark, Th. E. M. & Beurskens, P. T. (1975). *Acta Cryst.* A**31**, 813-817.
Heinerman, J. J. L. (1977). *Acta Cryst.* A**33**, 100-106.
Karle, J. & Karle, I. (1966). *Acta Cryst.* **21**, 849-859.
Keller-Schierlein, W., Meyer, M., Cellai, L., Cerrini, S., Lamba, D., Segre, A., Fedeli, W. & Brufani, M. (1985). *J. Antibiot.* In the press.
Krabbendam, H. & Kroon, J. (1971). *Acta Cryst.* A**27**, 48-53.
Main, P. (1976). *Crystallographic Computing Techniques*, edited by F. R. Ahmed, pp. 97-105. Copenhagen: Munksgaard.
Main, P. (1978). *Acta Cryst.* A**34**, 31-38.
Nunzi, A., Burla, M. C., Polidori, G., Giacovazzo, C., Cascarano, G., Viterbo, D., Camalli, M. & Spagna, R. (1984). *Acta Cryst.* A**40**, C425.
Prick, P. A. J., Beurskens, P. I. & Gould, R. O. (1983). *Acta Cryst.* A**39**, 570-576.
Sayre, D. (1952). *Acta Cryst.* **5**, 60-65.
Sim, G. A. (1959). *Acta Cryst.* **12**, 813-815.

Acta Cryst. (1985). A**41**, 613-617

New Methods for Deriving Joint Probability Distributions of Structure Factors. I

By J. Brosius*

Katholieke Universiteit Leuven, Departement Wiskunde, Belgium

(*Received* 23 *January* 1984; *accepted* 3 *June* 1985)

Abstract

With new probabilities, based on the Patterson function, for the 'atomic' random variables $\mathbf{x}_1, \ldots, \mathbf{x}_N$ in $P\bar{1}$, it is shown that an improved estimate can be obtained for the sign of the seminvariant $E_{2\mathbf{h}}$ in $P\bar{1}$. Two probability measures are considered. A method is also given for the case of a known Patterson vector of the form $2\mathbf{r}_1$, giving an estimate for the sign of any structure factor $E_\mathbf{h}$ by using its first neighborhood.

1. Introduction

For deriving joint probability distributions of structure factors one has used up to now two conceptually different approaches. One is to consider the structure factor

$$E_\mathbf{h} = \left(\sum_{i=1}^{N} f_i^2\right)^{-1/2} \sum_{i=1}^{N} f_i \exp(2\pi i \mathbf{h} \cdot \mathbf{x}_i) \qquad (1)$$

as a function of the random variables $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$; the other consists in regarding $E_\mathbf{h}$ as a function of the random variable $\mathbf{h}$. The first method consists in letting the random variables $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ range uniformly and independently over the unit cell, which may be represented mathematically by $[0, 1[^3$ [the set of all triples (u, v, w) where $0 \leq u, v, w < 1$]. In this paper other probability measures are considered for the random variables $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ based on the Patterson function. In particular, we study the seminvariant $E_{2\mathbf{h}}$ in $P\bar{1}$.

2. The probability distribution of $E_{2\mathbf{h}}$ in $P\bar{1}$ for different probabilities for $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$

Several probability measures for $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ will be considered and used to determine the sign of $E_{2\mathbf{h}}$ for its first neighborhood. In order to simplify calculations we shall treat the case of N equal atoms for which the structure factor $E_\mathbf{h}$ is given by

$$E_\mathbf{h} = 2N^{-1/2} \sum_{i=1}^{t} \cos(2\pi \mathbf{r}_i \cdot \mathbf{h})$$

($\mathbf{r}_i \in [0, 1[^3$ and $t = N/2$). The function Q defined on $[0, 1[^3$ by

$$\mathbf{u} \in [0, 1[^3 \rightarrow Q(\mathbf{u}) = \langle (E_\mathbf{k}^2 - 1) \exp(-2\pi i \mathbf{k} \cdot \mathbf{u}) \rangle_\mathbf{k} \qquad (2)$$

(where $\langle . \rangle_\mathbf{k}$ means the average over all reciprocal-lattice vectors) gives

$$Q(\mathbf{u}) = \begin{cases} N^{-1} & \text{if } \mathbf{u} = [2\mathbf{r}_i] \text{ or } \mathbf{u} = [-2\mathbf{r}_i] \quad (1 \leq i \leq t) \\ 2N^{-1} & \text{if } \mathbf{u} = [\mathbf{r}_i - \mathbf{r}_j] \text{ or } \mathbf{u} = [\mathbf{r}_i + \mathbf{r}_j] \text{ or} \\ & \quad \mathbf{u} = [-\mathbf{r}_i - \mathbf{r}_j] \quad (1 \leq i, j \leq t \text{ and } i \neq j) \\ 0 & \text{elsewhere,} \end{cases} \qquad (3)$$

where $[\mathbf{x}]$ for $\mathbf{x} \in \mathbb{R}^3$ denotes the unique vector in $[0, 1[^3$, which differs from $\mathbf{x}$ by some vector (p, q, r), where p, q and r are integer numbers.

This function Q will be used to construct several probability measures on the 'atomic' random variables $\mathbf{x}_i (1 \leq i \leq t)$. The simplest probability measure is obtained as follows. The random variables $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ will be taken to be independent. They are defined on $[0, 1[^3$, equipped with its usual collection of Borel sets, by $\mathbf{u} \in [0, 1[^3 \rightarrow \mathbf{x}_i(\mathbf{u}) = \mathbf{u}$ $(1 \leq i \leq t)$.

---

* Present address: Université du Burundi, Département de Máthematiques, BP 2700 Bujumbura, Burundi.

0108-7673/85/060613-05$01.50