Direct Methods and Structures Showing Superstructure Effects.
II. A Probabilistic Theory of Triplet Invariants

By G. Cascarano, C. Giacovazzo and M. Luić*

Dipartimento Geomineralogico, Università, Campus Universitario, Via Salvemini, 70124 Bari, Italy

Abstract

A probabilistic theory of triplet invariants is provided which may be used when pseudotranslations occur in the crystal structure. The final formula for estimating a triplet phase invariant in centrosymmetric space groups is of Cochran–Woolfson type, in noncentrosymmetric space groups of von Mises type with maximum at 2π; single phases are determined *via* a special tangent formula. Thus the usual algorithms for phase expansion and refinement can be employed with few modifications. Parameters occur in the von Mises and tangent formulae which are markedly different from the usual ones. In particular, the reliability of each triplet depends not only on $|E_h|$, $|E_k|$, $|E_{h-k}|$, but also on the actual h, k, $h-k$ indices and on the nature of the pseudotranslations. An automatic phasing procedure and some applications are also described for the solution of superstructures and other structures showing pseudotranslations.

Symbols and abbreviations

φ_h, φ_k, . . . : phase of E_h, E_k, ...
R_h, R_k, . . . : moduli of E_h, E_k, . . . respectively.
$[\sigma_r]_p$, $[\sigma_r]_q$, $[\sigma_r]_N$, . . . $= \sum Z_j^r$, where Z_j is the atomic number of the jth atom and the summation is made over the p, q, N, . . . atoms.
$C_h = \{\alpha_h[\sigma_2]_p + [\sigma_2]_q\}^{-1/2}$.
I_0: modified Bessel function of order zero.
N: number of atoms in the cell.
m: order of the space group.
p: number of atoms (symmetry-equivalent included) whose positions are related by the pseudotranslations $\mathbf{u}$.
q: number of atoms (symmetry-equivalent included) whose positions are not related by the pseudotranslations $\mathbf{u}$.
t_p: number of independent atoms which generate the p atoms when the pseudotranslations $\mathbf{u}$ and the symmetry operators C_s, $s = 1, 2, \ldots, m$, are applied.
t_q: number of independent atoms which generate the q atoms by application of the symmetry operators C_s, $s = 1, 2, \ldots, m$.
n_i: order of the pseudotranslation $\mathbf{u}_i$.

* Present address: 'Rudjer Bošković' Institute, Bijenička 54, 41000 Zagreb, Yugoslavia.

For other symbols see paper I (Cascarano, Giacovazzo & Luić, 1985*b*).

1. Introduction

Since the pioneering works by Lipson & Woolfson (1952), Rogers & Wilson (1953) and Hauptman & Karle (1953) it was clear that rational dependence of atom coordinates should affect both the statistics of normalized structure factors and the reliability of triplet phase relationships. The mere renormalization procedure proposed by Hauptman & Karle (1959) to overcome problems imposed by systematically weak and strong reflections was unsuccessful in several cases. Indeed, it implies that triplets involving reflexions from weak reflexion sets are as reliable as triplets from strong sets, which is in contradiction with Sayre's equation (Gramlich, 1975).

In some cases, it is relatively easy to determine the substructural unit: then the superstructure may be found by techniques particularly devoted to recovering the complete from the partial structure (Prick, Beurskens & Gould, 1983; Camalli, Giacovazzo & Spagna, 1985). However, the main aim of this paper is that of determining at the same time both the positions of the substructural and those of the superstructural atoms by taking into account the information (often available) on pseudotranslational symmetry. Such information may be derived by different methods: we quote, among others, that described by Cascarano, Giacovazzo & Luić (1985*a*, *b*), which suggests, *via* a statistical analysis of the diffraction intensities, the possible presence of a pseudotranslational symmetry. The method finds the nature of pseudosymmetry and estimates the fraction p/N of electrons suffering it.

In recent years more insight into the problem of using triplet relationships was achieved after the contributions by Böhme (1982, 1983), Fan Hai-fu, Yao Jia-xing, Main & Woolfson (1983), Gramlich (1984). A renormalization procedure was combined with suitable probabilistic considerations in order to provide better estimates of triplet invariants. However, some theoretical problems concerning the probabilistic estimation of triplet invariants when pseudotranslational symmetry is present still remained unsolved.

In this paper we describe a probabilistic theory of triplet relationships which explicitly takes into

account pseudotranslational symmetry. More precisely, we will calculate the conditional probability distribution of a phase φ_h given φ_k, φ_{h-k}, φ_k, $\varphi_{h-k}, \ldots$, when no correlation exists between atomic positions not related by symmetry or by pseudotranslations.

In accordance with the above observations, the first step of the mathematical approach will be the calculations of the joint probability distribution function $P(R_h, R_k, R_{h-k}, \varphi_h, \varphi_k, \varphi_{h-k})$ by taking into account: (1) the specific algebraic form of the E's, which is strongly dependent on the pseudotranslational symmetry, and (2) the reduced number of primitive random variables. Indeed, now only the $t_p + t_q$ atomic positions constitute the primitive independent random variables. According to paper I we write

$$E_h = \sum_{j=1}^{t_p+t_q} \psi_j(h) g_j(h) \qquad (1)$$

where

$$\psi_j(h) = f_j(h) \Big/ \left[\varepsilon_h \left(\alpha_h \sum_p + \sum_q \right) \right]^{1/2}.$$

The normalizing factor in (1) is

$$\langle |F_h|^2 \rangle = \varepsilon_h \left(\alpha_h \sum_p + \sum_q \right)$$

where

$$\alpha_h = (n_1 n_2 n_3 \ldots) \gamma_h / m$$

and γ_h is the number of times for which the algebraic congruences

$$h R_s u_i \equiv 0 \pmod{1} \quad \text{for } i = 1, 2, 3, \ldots$$

are simultaneously satisfied when s varies from 1 to m. If $\gamma_h = 0$ we say that F_h is a superstructure reflexion: then $\langle |F_h|^2 \rangle = \varepsilon \sum_q$. Otherwise F_h is a substructure reflexion. The maximum value of γ is m: consequently the maximum value of α is $n_1 n_2 n_3 \ldots$.

For $j < t_p$ [see equation (I.4a)],

$$g_j(h) = \sum_{s=1}^{m} \exp 2\pi i h C_s r_j \left[\prod_i \left(\frac{\sin n_i \pi h R_s u_i}{\sin \pi h R_s u_i} \right) \right.$$

$$\left. \times \exp 2\pi i h C_s \tfrac{1}{2}(n_i - 1) u_i \right];$$

for $t_p < j \leq t_p + t_q$ [see equation (I.4b)]

$$g_j(h) = \sum_{s=1}^{m} \exp 2\pi i h C_s r_j.$$

The index i varies over the independent pseudotranslations.

The mathematical model described above may be applied both when a defined substructure exists and when pseudotranslational symmetry does not give rise to a substructure (Cascarano, Giacovazzo & Luić, 1985a). On the other hand, real substructures often do not exactly comply with our model. For example,

it may occur that atoms related by pseudotranslational symmetry are not exactly located or are of different nature. Such situations involve a correlation between the superstructure and the substructure not taken into account by our model: dealing with these cases requires the modification of the theoretical background assumed in this paper. The effects of the correlation between superstructure and substructure on triplet estimation will be examined in a further paper. We only show here that our mathematical model provides probabilistic triplet estimations which can be profitably used in direct procedures for the solution of several real structures with superstructure effects.

For the sake of brevity we do not give a full account of our probabilistic approach: the reader is referred to a recent book (Giacovazzo, 1980) for the basic ideas. We give in §§ 2 and 3 only the final formulae for the cs and for the ncs cases respectively. At first sight they may appear rather unexpected. For this reason a short Appendix (Appendix A) will help the reader to justify our theoretical results. It is worthwhile mentioning that they do not hold for those special triplets described by Giacovazzo (1980, pp. 286–287) and by Pontenagel & Krabbendam (1983), for which supplementary algebraic considerations are needed.

2. Centrosymmetric space groups: the conditional probability that $E_h E_k E_{h-k}$ is positive given $R_h R_k R_{h-k}$

Denote by $P(E_h E_k E_{h-k} > 0)$ the conditional probability that the product $E_h E_k E_{h-k}$ is positive given $R_h R_k R_{h-k}$. On the assumption that all the atoms have the same unitary scattering factor we obtain (see Appendix A)

$$P(E_h E_k E_{h-k} > 0) \sim 0.5 + 0.5 \tanh A_{h,k} \qquad (2)$$

where

$$A_{h,k} = R_h R_k R_{h-k} N_{h,k}^{-1/2},$$

$$N_{h,k}^{-1/2} = C_h C_k C_{h-k} \left\{ m^{-1} [\sigma_3]_p \sum_{s=1}^{m} \left(\prod_i \frac{\tau_{s,i}}{n_i} \right) + [\sigma_3]_q \right\} \qquad (3)$$

and

$$\tau_{s,i} = \frac{\sin n_i \pi h R_s u_i}{\sin \pi h R_s u_i} \frac{\sin n_i \pi k R_s u_i}{\sin \pi k R_s u_i}$$

$$\times \frac{\sin n_i \pi (h-k) R_s u_i}{\sin \pi (h-k) R_s u_i}.$$

If $p = 0$ (no pseudotranslation occurs), then $q = N$ and

$$N_{h,k} = [\sigma_2]_N^3 / [\sigma_3]_N^2.$$

Thus (2) reduces to the well known Cochran-Woolfson formula.

If all the reflexions $\mathbf{h}$, $\mathbf{k}$, $\mathbf{h}-\mathbf{k}$ are superstructure reflexions then $(\prod_i \tau_{s,i}) = 0$ for every s, and $N_{\mathbf{h}-\mathbf{k}}$ is given by

$$(N_{\mathbf{h},\mathbf{k}})_0 = [\sigma_2]_q^3/[\sigma_3]_q^2 \sim q. \tag{4}$$

Thus the reliability of a triplet constituted by superstructure reflexions appears to be inversely proportional to $q^{1/2}$ and not to $N^{1/2}$.

It may be worth mentioning that triplets constituted by superstructure reflexions may not always be found. For example, when a pseudotranslation of order 2 occurs, the superstructure reflexion $\mathbf{h}$ satisfies the condition

$$\mathbf{h}\mathbf{u} \not\equiv 0(\bmod 1) \quad \text{and} \quad 2\mathbf{h}\mathbf{u} \equiv 0(\bmod 1).$$

Thus $\mathbf{h}\mathbf{u}$ is a semi-integer value. If $\mathbf{h}$ and $\mathbf{k}$ are superstructure reflexions, $(\mathbf{h}-\mathbf{k})\mathbf{u}$ is the sum of two semiinteger values, that is to say, $(\mathbf{h}-\mathbf{k})$ must be a substructure reflexion.

If only one of $\mathbf{h}$, $\mathbf{k}$, $\mathbf{h}-\mathbf{k}$ is a substructure reflexion (*e.g.* the reflexion $\mathbf{h}$), then $(\prod_i \tau_{s,i}) = 0$ for every s, and $N_{\mathbf{h},\mathbf{k}}$ is given by

$$(N_{\mathbf{h},\mathbf{k}})_1 \simeq \{\alpha_{\mathbf{h}}[\sigma_2]_p + [\sigma_2]_q\}[\sigma_2]_q^2/[\sigma_3]_q^2. \tag{5}$$

$(N_{\mathbf{h},\mathbf{k}})_1$ is larger than $(N_{\mathbf{h},\mathbf{k}})_0$. In order to have a simple numerical insight, let us assume that all the atoms are of the same type. Then $(N_{\mathbf{h},\mathbf{k}})_1 \simeq (\alpha_{\mathbf{h}}p + q) \geq (N_{\mathbf{h},\mathbf{k}})_0$.

If pseudotranslation symmetry gives rise to a subcell, triplets constituted of two substructure and one superstructure reflexions cannot be found. Indeed, if $\mathbf{h}$ and $\mathbf{k}$ correspond to the reciprocal lattice of the subcell, $\mathbf{h}-\mathbf{k}$ will also correspond to points of the same lattice. On the other hand, if pseudotranslational symmetry does not produce a subcell (Cascarano, Giacovazzo & Luić, 1985a), 'sub-sub-super' triplets can exist (this case has never been considered in previous literature). As an example, let $\mathbf{h} = 417$, $\mathbf{k} = \overline{3}82$ and $\mathbf{h}-\mathbf{k} = 795$: if $\mathbf{u} = \mathbf{a}/4$ in $P4$, then $\mathbf{h}$ and $\mathbf{k}$ are substructure reflexions with $\alpha_{\mathbf{h}} = \alpha_{\mathbf{k}} = 2$ while $\mathbf{h}-\mathbf{k}$ is a superstructure reflexion.

In general, if $\mathbf{h}$ and $\mathbf{k}$ are substructure reflexions and $\mathbf{h}-\mathbf{k}$ is a superstructure reflexion then

$$(N_{\mathbf{h},\mathbf{k}})_2 \simeq (\alpha_{\mathbf{h}}[\sigma_2]_p + [\sigma_2]_q)(\alpha_{\mathbf{k}}[\sigma_2]_p + [\sigma_2]_q)$$
$$\times [\sigma_2]_q/[\sigma_3]_q^2. \tag{6}$$

It is easy to see that $(N_{\mathbf{h},\mathbf{k}})_2 \geq (N_{\mathbf{h},\mathbf{k}})_1 \geq (N_{\mathbf{h},\mathbf{k}})_0$. If all the atoms are of the same type then

$$(N_{\mathbf{h},\mathbf{k}})_2 \simeq (\alpha_{\mathbf{h}}p + q)(\alpha_{\mathbf{k}}p + q)/q,$$

which is larger than $(N_{\mathbf{h},\mathbf{k}})_1$ because $(\alpha_{\mathbf{k}}p + q)/q > 1$, unless $p = 0$.

If $\mathbf{h}$, $\mathbf{k}$, $\mathbf{h}-\mathbf{k}$ are all substructure reflexions then the complete equation (3) has to be calculated, in which the factor $\sum_{s=1}^m (\prod_i \tau_{s,i}/n_i)$ is in general non-vanishing. It should be stressed that $\tau_{s,i}$ is a non-negative quantity.

Indeed: (1) $\sin n_i\pi y/\sin \pi y = 0$ if y is not an integral value [here y is one of $\mathbf{h}\mathbf{R}_s\mathbf{u}_i$, $\mathbf{k}\mathbf{R}_s\mathbf{u}_i$, $(\mathbf{h}-\mathbf{k})\mathbf{R}_s\mathbf{u}_i$]; (2) if n_i is odd then $\sin n_i\pi y/\sin \pi y = n_i$ for integral values of y; (3) if n_i is even then $\sin n_i\pi y/\sin \pi y$ equals n_i for even values of y, and it equals $-n_i$ for odd values of y. In both cases $\tau_{s,i}$ is again positive because none or two of $\mathbf{h}\mathbf{R}_s\mathbf{u}_i$, $\mathbf{k}\mathbf{R}_s\mathbf{u}_i$, $(\mathbf{h}-\mathbf{k})\mathbf{R}_s\mathbf{u}_i$ can be odd.

It may be written in a way suitable for computer calculations [see equation $(A.3)$]

$$\sum_{s=1}^m \left(\prod_i \frac{\tau_{s,i}}{n_i} \right) = \beta(n_1^2 n_2^2 n_3^2 \ldots) \tag{7}$$

where β is the number of times for which

$$\mathbf{h}\mathbf{R}_s\mathbf{u}_1 \equiv 0(\bmod 1),$$

$$\mathbf{h}\mathbf{R}_s\mathbf{u}_2 \equiv 0(\bmod 1),$$

$$\mathbf{h}\mathbf{R}_s\mathbf{u}_3 \equiv 0(\bmod 1)\ldots$$

$$\mathbf{k}\mathbf{R}_s\mathbf{u}_1 \equiv 0(\bmod 1),$$

$$\mathbf{k}\mathbf{R}_s\mathbf{u}_2 \equiv 0(\bmod 1),$$

$$\mathbf{k}\mathbf{R}_s\mathbf{u}_3 \equiv 0(\bmod 1)\ldots$$

$$(\mathbf{h}-\mathbf{k})\mathbf{R}_s\mathbf{u}_1 \equiv 0(\bmod 1),$$

$$(\mathbf{h}-\mathbf{k})\mathbf{R}_s\mathbf{u}_2 \equiv 0(\bmod 1),$$

$$(\mathbf{h}-\mathbf{k})\mathbf{R}_s\mathbf{u}_3 \equiv 0(\bmod 1)\ldots$$

are simultaneously satisfied when s varies from 1 to m. Obviously, $\beta \leq m$: furthermore, β cannot exceed the minimum among $\gamma_{\mathbf{h}}$, $\gamma_{\mathbf{k}}$, $\gamma_{\mathbf{h}-\mathbf{k}}$ (see paper I). In conclusion,

$$(N_{\mathbf{h},\mathbf{k}})_3 = \{(\alpha_{\mathbf{h}}[\sigma_2]_p + [\sigma_2]_q)(\alpha_{\mathbf{k}}[\sigma_2]_p + [\sigma_2]_q)$$
$$\times (\alpha_{\mathbf{h}-\mathbf{k}}[\sigma_2]_p + [\sigma_2]_q)\}$$
$$\times \{(\beta/m)[\sigma_3]_p(n_1^2 n_2^2 n_3^2 \ldots) + [\sigma_3]_q\}^{-2}. \tag{8}$$

It may be shown (see Appendix B) that $(N_{\mathbf{h},\mathbf{k}})_3 \lessgtr (N_{\mathbf{h},\mathbf{k}})_2$ according to different situations.

The numerical values of $(N_{\mathbf{h},\mathbf{k}})_i$ suggest which type of triplet relationships will have a marginal or a central role (see § 4) in direct procedures for phase solution. For equal-atom structures (3) reduces to

$$N_{\mathbf{h},\mathbf{k}}^{-1/2} = [(\alpha_{\mathbf{h}}p + q)(\alpha_{\mathbf{k}}p + q)(\alpha_{\mathbf{h}-\mathbf{k}}p + q)]^{-1/2}$$
$$\times \left[(1/m)p \sum_{s=1}^m \left(\prod_i \tau_{s,i}/n_i \right) + q \right]. \tag{9}$$

3. Non-centrosymmetric space groups: the conditional probability of $\Phi_{\mathbf{h},\mathbf{k}} = \varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{h}-\mathbf{k}}$ given $R_{\mathbf{h}}$, $R_{\mathbf{k}}$, $R_{\mathbf{h}-\mathbf{k}}$

Under the same conditions stated in § 2 we obtain for the ncs space groups the relation

$$P(\Phi_{\mathbf{h},\mathbf{k}}) \simeq [2\pi I_0(2A_{\mathbf{h},\mathbf{k}})]^{-1} \exp(2A_{\mathbf{h},\mathbf{k}} \cos \Phi_{\mathbf{h},\mathbf{k}}). \tag{10}$$

Equation (10) is a unimodal von Mises function, with

its maximum at $\Phi_{h,k} = 0$. In order to obtain more insight into the numerical values involved in (2) or (10) we calculate in a few cases some parameters for an equal-atom $P4$ random structure, having $t_p = 5$, $t_q = 15$, $u = a/4$. Then $p = 80$, $q = 60$, $N = 140$. According to paper I the substructure reflexions hkl have h or k congruent to zero modulo 4. If superstructure effects are not taken into account in the normalization process, we would obtain the pseudonormalized structure factors E' and the following expected averages:

$$\langle |E'_h|^2 \rangle_{sup} \simeq [\sigma_2]_q / [\sigma_2]_N \simeq q/N = 0.43$$

$$\langle |E'_h|^2 \rangle_{h=4n} \simeq 1.57$$

$$\langle |E'_h|^2 \rangle_{h=4n,k=4n} \simeq 2.71$$

$$\langle |E'_h|^2 \rangle_{h=k=0} \simeq 2.71.$$

Our normalization program (described in paper I) finds $\langle |E'_h|^2 \rangle_{sub} = 1.76$, from which p and q are estimated to be 83 and 57 respectively.

The true normalized structure factors E can be obtained via (I.22), or, more simply, by

$$E_h = E'_h \{ [\sigma_2]_N / (\alpha_h [\sigma_2]_p + [\sigma_2]_q) \}^{1/2}, \quad (11)$$

which, in our example, reduces to

$$E_h = E'_h [N/(\alpha_h p + q)]^{1/2}.$$

In Table 1 the true $N_{h,k}$ values [calculated via (9)] are shown for the various types of triplets; the values calculated by our program are in parentheses (the experimental estimates for p and q have been used). In Table 2 some pseudonormalized E' and the corresponding normalized structure factors E are shown. The values of $A_{h,k}$ according to Cochran's (1955) formula and to our procedure are given for each triplet. The tables show how different the reliability of the various triplets may be according to the two procedures. In particular the value $N = 140$ to be used in Cochran's formula is replaced in our procedure (see Table 1) by 22, 31, 38, 60, 220, 380, 807 according to the type of triplet. Clearly, triplets constituted by 3, 1 or 0 substructure reflexions may play an important role in the phasing procedures, while triplets constituted by only two substructure reflexions have a marginal role.

Besides space-group symmetry and pseudotranslational symmetry p and q values are also important for defining the specific role of a given type of triplet in the crystal-structure-determination process. Let us consider as an example a random $P4$ structure with $t_p = 15$, $t_q = 5$, $u = a/4$: then $p = 240$, $q = 20$, $N = 260$. In Table 3 the values of $N_{h,k}$ [calculated via (9)] are shown for the various types of triplets (the values calculated by our program in parentheses). From comparison of Table 3 with Table 1 it is easily seen that 'super-super-super' triplets are the most important ones. Indeed, the eight triplets with largest values

Table 1. *Values $N_{h,k}$ for the various types of triplets in a $P4$ randomly generated structure with $t_p = 5$, $t_q = 15$, $u = a/4$*

In the first column the parity of the reflexions involved in each triplet is given: f stands for 'modulo 4', n stands for 'non-modulo 4', l is either. In the second column the various values of $N_{h,k}$ are given (with the values calculated by our program in parentheses).

Type of triplet	N
$(fnl), (fnl), (fnl)$	22 (21)
$(ffl), (ffl), (ffl)$	31 (31)
$(fnl), (fnl), (ffl)$	38 (37)
$(nnl), (nnl), (nnl)$	60 (57)
$(nnl), (nnl), (nfl)$	220 (223)
$(nnl), (nnl), (ffl)$	380 (390)
$(nnl), (fnl), (fnl)$	807 (879)

Table 2. *Some data for a $P4$ randomly generated structure with $t_p = 5$, $t_q = 15$, $u = a/4$*

E' denotes a structure factor normalized without taking the pseudotranslation effects into account (pseudonormalized structure factor), E is the correct normalized structure factor. $2A'(h, k, h-k) = 2|E'_h E'_k E'_{h-k}|/N^{1/2}$ is the reliability factor of a triplet when pseudotranslation effects are not taken into account, $2A(h, k, h-k) = 2|E_h E_k E_{h-k}|/N_{h,k}^{1/2}$ is the corresponding parameter in our formula (10).

	h_1 (412)	h_2 ($\bar{8}83$)	h_3 (12,9,5)	h_4 ($\bar{3}45$)	h_5 ($7\bar{5}3$)	h_6 (231)	h_7 ($5\bar{2}4$)	h_8 ($93\bar{2}$)
E'_h	2.0	3.0	3.5	2.7	0.9	1.2	1.0	1.3
α_h	2	4	2	2	0	0	0	0
E_h	1.6	1.8	2.8	2.2	1.4	1.8	1.5	2.0

$2A'(h_1, h_2, h_3) = 3.5$ $2A(h_1, h_2, h_3) = 2.6$
$2A'(h_1, h_4, h_5) = 0.8$ $2A(h_1, h_4, h_5) = 0.3$
$2A'(h_1, h_7, h_8) = 0.66$ $2A(h_1, h_7, h_8) = 0.97$
$2A'(h_5, h_6, h_7) = 0.18$ $2A(h_5, h_6, h_7) = 0.98$

of $A(h, k)$ are all of type 'super-super-super' while the largest value of $A(h, k)$ for 'sub-sub-sub' is 2.71 (74 in the list).

Two additional points deserve to be stressed. Renormalization of structure factors proposed by Hauptman & Karle (1959) to remove in the phasing procedures problems imposed by systematically strong and weak reflexions does exploit the indirect information contained in the statistics of the structure factors, but on the other hand it ignores the usefulness of knowing the structural regularities causing those statistical effects. In the light of the above results such a procedure appears too drastic and rigid. For example, it reduces too much the reliability of the 'sub-sub-sub' triplets and overestimates by different amounts the other types of triplet. According to our theory, the information obtained from the analysis of statistics is used, not only for renormalizing structure factors, but also for evaluating suitable values of $N_{h,k}$ for each type of triplet. For example, for 'sub-sub-sub' triplets, $N_{h,k} < N$, so that those triplets are estimated to be more reliable than via the mere normalization process. Support for these conclusions is provided in Appendix C, where the Sayre (1952) equation

Table 3. *Values $N_{h,k}$ for the various types of triplets in a P4 randomly generated structure, with $t_p = 15$, $t_q = 5$,*

$$u = a/4$$

For symbols see caption to Table 1.

Type of triplet	N
$(fnl), (fnl), (fnl)$	33 (33)
$(ffl), (ffl), (ffl)$	63 (63)
$(fnl), (fnl), (ffl)$	65 (65)
$(nnl), (nnl), (nnl)$	20 (21)
$(nnl), (nnl), (nfl)$	500 (449)
$(nnl), (nnl), (ffl)$	980 (977)
$(nnl), (fnl), (fnl)$	12 500 (11 738)

is briefly discussed and compared with approaches by other authors.

The above remarks are the starting point for a more general proposition: when *a priori* information is available, it must be used in the normalization process, in the theory aiming at estimating phase relationships, and in the phasing process, otherwise systematic errors or other drawbacks will be produced during the procedure.

The second additional point to stress is that the mathematical approach described above cannot be applied rigorously when t_p or t_q are small numbers (in some structures $t_p = 1$). Situations in which the number of independent random variables is small but direct methods are successful are not infrequent: for example, very small structures or structures partially known.

The usual probabilistic theories also proved to be useful in these cases; therefore we can expect that small values of t_p or t_q are disturbing but not critical parameters for the theory developed in the present paper.

4. The conditional distribution of φ_h given R_h and more pairs φ_k, φ_{h-k}, R_k, R_{h-k}

In the cs space groups we obtain

$$P(E_h > 0) \simeq 0.5 + 0.5 \tanh \left[R_h \sum_j (N_{h,k_j}^{-1/2}) E_{k_j} E_{h-k_j} \right] \quad (12)$$

and, in ncs space groups,

$$P(\varphi_h) = [2\pi I_0(Q_h)]^{-1} \exp [Q_h \cos (\varphi_h - \theta_h)], \quad (13)$$

where

$$Q_h = \left\{ \left[\sum_j 2A_{h,k_j} \cos (\varphi_{k_j} + \varphi_{h-k_j}) \right]^2 \right.$$

$$\left. + \left[\sum_j 2A_{h,k_j} \sin (\varphi_{k_j} + \varphi_{h-k_j}) \right]^2 \right\}^{1/2}, \quad (14)$$

$$\tan \theta_h = \frac{\sum_j A_{h,k_j} \sin (\varphi_{k_j} + \varphi_{h-k_j})}{\sum_j A_{h,k_j} \cos (\varphi_{k_j} + \varphi_{h-k_j})}. \quad (15)$$

It is easily seen that the usual algorithms for phase expansion and refinement need only a few modifications for handling superstructure problems.

5. Real structures: possible sources of ambiguity and experimental results

The presence of heavy atoms in a crystal structure is not a factor limiting the effectiveness of the renormalization procedure described in paper I. Indeed, the process depends on the ratios $\sum_p/\sum_N$ and $\sum_q/\sum_N$ (scattering powers of the q and p atoms/scattering power of the complete crystal structure) which in principle are statistically estimable with identical accuracy no matter if heavy atoms occur. However, if some heavy atoms are present, the statistical analysis of intensities described in paper I does not provide the species and the number of atoms suffering pseudotranslation, but only the ratio $\sum_q/\sum_N$ which reduces to q/N only if all atoms are equal. In this case (and only in this case) (3) reduces to (9), which is therefore computable *via* the statistical analysis of diffraction data.

In general $\sum_q/\sum_N$ and q/N do not coincide.

In order to give an example, let us suppose that the crystal structure contains p atoms with atomic number Z_p and q atoms with atomic number Z_q. According to (I.13), we write

$$\langle |E_h'|^2 \rangle = (\alpha_h p Z_p^2 + q Z_q^2)/(p Z_p^2 + q Z_q^2). \quad (16)$$

Since $p = (N - q)$, (16) becomes

$$\frac{q}{N} = \frac{\langle |E_h'|^2 \rangle - \alpha_h}{\langle |E_h'|^2 \rangle - \langle |E_h'|^2 \rangle (Z_q^2/Z_p^2) - \alpha_h + (Z_q^2/Z_p^2)}$$

$$= \frac{\langle |E_h'|^2 \rangle - \alpha_h}{(\langle E_h'|^2 \rangle - 1)[1 - (Z_q^2/Z_p^2)] + 1 - \alpha_h}, \quad (17)$$

which reduces to equation (I.15) only if $Z_q = Z_p$.

For $p = 4$ atoms with $Z_p = 40$ and $q = 60$ with $Z_q = 6$ the statistical analysis of diffraction data *via* equation (I.16) should estimate $q/N \simeq 0.25$, so that p and q should be assumed proportional to 48 and 16 respectively, in evident contrast with reality.

The above is a dramatic example in which the scattering power of the q atoms constitutes a small percentage of the total scattering but the number of q atoms is a large percentage of the total number. A converse example may easily be given. The obvious consequence is that q/N or p/N or p/q have always to be interpreted as ratios between scattering powers and not between numbers of atoms.

The situation becomes more complicated when the p atoms as well as the q atoms are not of the same type. Indeed:

(*a*) If pseudotranslational symmetry relates atoms of different type then the p atoms contribute to both $\sum_p$ and $\sum_q$ (that is to say, to both substructure and superstructure reflexions). The first contribution

arises from those electrons of the p atoms which do overlap because of pseudotranslations (the maximum overlap cannot exceed the lightest atom). The contribution to $\sum_q$ arises from those electrons of the p atoms which do not satisfy pseudotranslational symmetry. Thus, $\rho_q(\mathbf{r}) = \rho(\mathbf{r}) - \rho_p(\mathbf{r})$ will contain, besides the q atoms, also some electron residuals of the p atoms (see Fig. 1). Then the overall number of atomic peaks in $\rho_q(\mathbf{r})$ is larger than $N - p$. Underestimation of this effect will produce overestimation of 'super-super-super' and 'super-super-sub' triplets with respect to the others.

(b) Statistical underestimation of $[\sigma_2]_p/[\sigma_2]_N$ is equivalent to partially neglecting information contained in pseudotranslational symmetry and causes some overestimation of 'sub-sub-sub' triplets with respect to others. However, overestimation of $[\sigma_2]_p/[\sigma_2]_N$ is more dangerous (see also Bohme, 1982) because it introduces into the mathematical model atoms with negative scattering factors (see Fig. 1). In this case, triplets of type 'super-super-super' are overestimated with respect to 'super-super-sub' which in their turn are overestimated with respect to 'sub-sub-sub' triplets.

(c) $[\sigma_3]_p$ and $[\sigma_3]_q$ are not accessible via statistical analysis so that $N_{\mathbf{h},\mathbf{k}}$ is not exactly computable via (3). Therefore, in the practical procedure, we use the approximate equation (9) instead of (3).

In spite of the above limitations our approach works well in several cases. We have applied it to the five crystal structures described in Table 4. The first three of them were originally solved via ad hoc
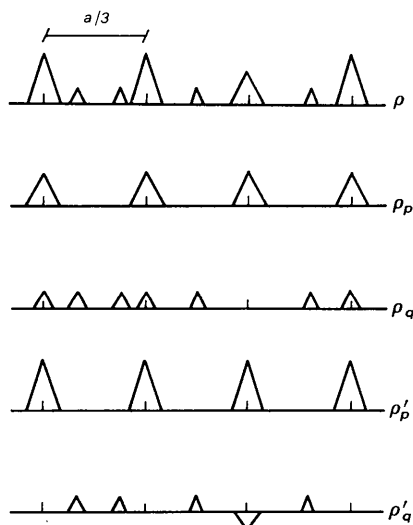

Fig. 1. ρ is the unidimensional electron density constituted by triangular 'atoms'. The pseudotranslational symmetry $\mathbf{u} = \mathbf{a}/3$ relates atoms of different type: ρ_p is the electron density suffering pseudotranslational symmetry, $\rho_q = \rho - \rho_p$ is the complementary electron distribution, ρ_p' is a wrong estimate of ρ_p (possibly suggested by statistical analysis): in this case ρ_q' contains 'negative' atoms.

Table 4. *Test structures: space group, the main pseudotranslational vector* (**u**) *and the estimated percentages of electrons* (*e*%) *suffering pseudotranslational symmetry*

Crystal structure	Space group	u	e%
Freieslebenite* PbAgSbS$_3$	$P2_1/a$ $Z = 4$	a/2+b/3	0·51
Mesolite† Na$_2$Ca$_2$Al$_6$Si$_9$O$_{30}$·8H$_2$O	$Fdd2$ $Z = 16$	a/3	0·66
Ferri‡ Fe$_2$(SO$_4$)$_3$	$P2_1/n$ $Z = 4$	(a+b)/2	0·55
Nist§ [N(C$_4$H$_9$)$_4$][Ni(C$_3$S$_5$)$_2$]	$P1$ $Z = 2$	a/2	0·37
Fega¶ Fe$_2$Ga$_2$S$_5$	$P6_3/mmc$ $Z = 2$	c/3	0·44

* Ito & Nowacki (1974).
† Adiwidjaja (1972).
‡ Christidis & Rentzeperis (1975).
§ Unpublished.
¶ Nguyen-Huy-Dung *et al.* (1986). Unpublished.

methods. Nist can also be routinely solved without introducing information on pseudotranslational symmetry. The last one was *MULTAN* resistant and has been routinely solved by the present method. In all cases our procedure revealed for each crystal structure a large percentage of atoms which, via the usual Fourier methods, led to the complete structures.

6. Concluding remarks

The mathematical model introduced in paper I has been applied to the estimation of triplet invariants. The approach is able to produce for non-centrosymmetric space groups a unique distribution for the triplet phases which is of von Mises type and whose concentration parameter changes according to the type of triplet and to the pseudotranslational symmetry. A mathematical model cannot take into account every type of pseudotranslational symmetry: thus in principle our procedure may be successful in some cases (when the actual crystal structure complies well with the model) and may be unsuccessful in others. In spite of this limitation the method described in this paper seems rather efficient: indeed, it worked on real crystal structures which do not fully comply with the mathematical model both because atoms related by pseudotranslational symmetry are not equal and because they are not exactly located.

The procedure has been introduced into the *SIR* package (Cascarano, Giacovazzo, Burla, Nunzi, Polidori, Camalli, Spagna & Viterbo, 1985). The macro-instruction *PSEUDO* compels the normalization program to carry out the statistical analysis of diffraction intensities, in order to recognize the pseudotranslational symmetry and to estimate p and q parameters. The information so obtained, together with renormalized structure factors, is passed to the Σ_2 routine, which calculates for each triplet the concentration parameter $A_{\mathbf{h},\mathbf{k}}$. In its turn this parameter

is used in the *CONVERGENCE* procedure for origin and enantiomorph definition and in the tangent process.

One of us (ML) is indebted to ICTP (International Centre for Theoretical Physics, Programme for Training and Research in Italian Laboratories, Trieste, Italy) for financial support during a three-month stay at the University of Bari.

APPENDIX A

In accordance with equation (1),

$$E_\mathbf{h}E_\mathbf{k}E_{\overline{\mathbf{h+k}}} = \sum_{j_1=1}^{t_p+t_q} \psi_{j_1}(\mathbf{h})g_{j_1}(\mathbf{h}) \sum_{j_2=1}^{t_p+t_q} \psi_{j_2}(\mathbf{k})g_{j_2}(\mathbf{k})$$
$$\times \sum_{j_3=1}^{t_p+t_q} \psi_{j_3}(\overline{\mathbf{h+k}})g_{j_3}(\overline{\mathbf{h+k}}).$$

If no correlation exists among the $t_p + t_q$ primitive random variables $\mathbf{r}_j$, then

$$\langle E_\mathbf{h}E_\mathbf{k}E_{\overline{\mathbf{h+k}}}\rangle = \sum_{j=1}^{t_p+t_q} \langle \psi_j(\mathbf{h})\psi_j(\mathbf{k})\psi_j(\overline{\mathbf{h+k}})$$
$$\times g_j(\mathbf{h})g_j(\mathbf{k})g_j(\overline{\mathbf{h+k}})\rangle. \quad (A.1)$$

If we denote

$$S(\mathbf{h}, s, i) = \frac{\sin(n_i\pi\mathbf{h}\mathbf{R}_s\mathbf{u}_i)}{\sin(\pi\mathbf{h}\mathbf{R}_s\mathbf{u}_i)},$$

$$\nu(\mathbf{h}, s) = \exp\{2\pi i\mathbf{h}\mathbf{C}_s[\mathbf{r}_j + \tfrac{1}{2}(n_1-1)\mathbf{u}_1 + \tfrac{1}{2}(n_2-1)\mathbf{u}_2$$
$$+ \tfrac{1}{2}(n_3-1)\mathbf{u}_3 + \ldots]\},$$

then

$$\langle E_\mathbf{h}E_\mathbf{k}E_{\overline{\mathbf{h+k}}}\rangle = \sum_{j=1}^{t_p} \psi_j(\mathbf{h})\psi_j(\mathbf{k})\psi_j(\overline{\mathbf{h+k}})$$
$$\times \left\langle \sum_{s_1=1}^{m} [S(\mathbf{h}, s_1, 1)S(\mathbf{h}, s_1, 2)\right.$$
$$\times S(\mathbf{h}, s_1, 3)\ldots]\nu(\mathbf{h}, s_1)$$
$$\times \sum_{s_2=1}^{m} [S(\mathbf{k}, s_2, 1)S(\mathbf{k}, s_2, 2)$$
$$\times S(\mathbf{k}, s_2, 3)\ldots]\nu(\mathbf{k}, s_2)$$
$$\times \sum_{s_3=1}^{m} [S(\overline{\mathbf{h+k}}, s_3, 1)S(\overline{\mathbf{h+k}}, s_3, 2)$$
$$\times S(\overline{\mathbf{h+k}}, s_3, 3)\ldots]\nu(\overline{\mathbf{h+k}}, s_3)\right\rangle$$
$$+ \left\langle \sum_{j=t_p+1}^{t_p+t_q} \psi_j(\mathbf{h})\psi_j(\mathbf{k})\psi_j(\overline{\mathbf{h+k}})\right.$$
$$\times \sum_{s_1,s_2,s_3=1}^{m} \exp 2\pi i[\mathbf{h}(\mathbf{C}_{s_1} - \mathbf{C}_{s_3})\mathbf{r}_j$$
$$\left.+ \mathbf{k}(\mathbf{C}_{s_2} - \mathbf{C}_{s_3})\mathbf{r}_j]\right\rangle. \quad (A.2)$$

Non-vanishing contributions to $\langle E_\mathbf{h}E_\mathbf{k}E_{\overline{\mathbf{h+k}}}\rangle$ from the right-hand side of ($A.2$) occur if

$$\langle \nu(\mathbf{h}, s_1)\nu(\mathbf{k}, s_2)\nu(\overline{\mathbf{h+k}}, s_3)\rangle \neq 0.$$

Under the conditions specified in § 1, that requires $s_1 = s_2 = s_3$. Therefore, ($A.2$) becomes

$$\langle E_\mathbf{h}E_\mathbf{k}E_{\overline{\mathbf{h+k}}}\rangle$$
$$= \left[\sum_{j=1}^{t_p} \psi_j(\mathbf{h})\psi_j(\mathbf{k})\psi_j(\overline{\mathbf{h+k}})\right]$$
$$\times \sum_{s=1}^{m} [S(\mathbf{h}, s, 1)S(\mathbf{h}, s, 2)S(\mathbf{h}, s, 3)\ldots$$
$$\times S(\mathbf{k}, s, 1)S(\mathbf{k}, s, 2)S(\mathbf{k}, s, 3)\ldots$$
$$\times S(\overline{\mathbf{h+k}}, s, 1)S(\overline{\mathbf{h+k}}, s, 2)S(\overline{\mathbf{h+k}}, s, 3)\ldots]$$
$$+ \sum_{j=t_p+1}^{t_p+t_q} m\psi_j(\mathbf{h})\psi_j(\mathbf{k})\psi_j(\overline{\mathbf{h+k}})$$
$$= \left[\sum_{j=1}^{t_p} \psi_j(\mathbf{h})\psi_j(\mathbf{k})\psi_j(\overline{\mathbf{h+k}})\right] \sum_{s=1}^{m} \left[\prod_i \tau(s, i)\right]$$
$$+ \sum_{j=t_p+1}^{t_p+t_q} m\psi_j(\mathbf{h})\psi_j(\mathbf{k})\psi_j(\overline{\mathbf{h+k}}),$$

where

$$\tau(s, i) = S(\mathbf{h}, s, i)S(\mathbf{k}, s, i)S(\overline{\mathbf{h+k}}, s, i).$$

Since

$$p = mt_p(n_1n_2n_3\ldots), \qquad q = mt_q,$$

then

$$\langle E_\mathbf{h}E_\mathbf{k}E_{\overline{\mathbf{h+k}}}\rangle = \left[\frac{1}{m}\sum_{j=1}^{p} \psi_j(\mathbf{h})\psi_j(\mathbf{k})\psi_j(\overline{\mathbf{h+k}})\right]$$
$$\times \sum_{s=1}^{m} \left[\prod_i \frac{\tau(s, i)}{n_i}\right]$$
$$+ \sum_{j=p+1}^{N} \psi_j(\mathbf{h})\psi_j(\mathbf{k})\psi_j(\mathbf{h+k}), \quad (A.3)$$

from which (3) is obtained.

APPENDIX B

Suppose all atoms are equal. Then

$$(N_{\mathbf{h,k}})_3 \geq (N_{\mathbf{h,k}})_2$$

if

$$(\alpha_{\mathbf{h+k}}p + q)/\{(\beta/m)p(n_1^2n_2^2n_3^2\ldots) + q\}^2 \geq q^{-1},$$

which is true if

$$p(p\delta^2 + 2q\delta - \alpha_{\mathbf{h+k}}q) \leq 0, \quad (B.1)$$

where $\delta = \beta(n_1^2n_2^2n_3^2\ldots)/m$.

Relation ($B.1$) may be studied as a function of δ (that is to say, p and q are fixed while the space group

and the pseudotranslations are allowed to vary) or as a function of p (the space group and the pseudotranslations are given while p is the variable). In both cases $(N_{h,k})_3 = (N_{h,k})_2$ if $p = 0$.

In the first case, $(B.1)$ is an equality if $\delta = \delta_1$ or δ_2, where

$$\delta_1 = \{-q - [q(q + \alpha_{h+k}p)]^{1/2}\}/p,$$

$$\delta_2 = \{-q + [q(q + \alpha_{h+k}p)]^{1/2}\}/p.$$

δ_2 is the unique root allowed for δ (indeed $\delta > 0$). Therefore $(N_{h,k})_3 \geq (N_{h,k})_2$ if $\delta \leq \delta_2$, $(N_{h,k})_3 \leq (N_{h,k})_2$ if $\delta \geq \delta_2$. In the second case $(N_{h,k})_3 \gtreqless (N_{h,k})_2$ if $p \lesseqgtr q(\alpha_{h+k} - 2\delta)/\delta^2$. If $2\delta > \alpha_{h+k}$ then $(N_{h,k})_3 < (N_{h,k})_2$ always.

APPENDIX C

Suppose for an equal-atom structure that the electron density $\rho(\mathbf{r})$ may be expressed as the sum of the substructure $\rho_p(\mathbf{r})$, relative to p atoms in the cell, and of $\rho_q(\mathbf{r})$, relative to q atoms not related by pseudotranslations:

$$\rho(\mathbf{r}) = \rho_p(\mathbf{r}) + \rho_q(\mathbf{r}). \qquad (C.1)$$

For the sake of simplicity let us deal with the $P1$ case. We write

$$\rho_p(\mathbf{r}) = \rho_{t_p}(\mathbf{r}) * \sum_{\nu_1=0}^{n_1-1} \delta(\mathbf{r} - \nu_1\mathbf{u}_1) * \sum_{\nu_2=0}^{n_2-1} \delta(\mathbf{r} - \nu_2\mathbf{u}_2) * \ldots,$$

$$(C.2)$$

where $\rho_{t_p}(\mathbf{r})$ is the structure corresponding to the t_p independent atoms, $*$ is the symbol of convolution, $\mathbf{u}_1, \mathbf{u}_2, \ldots$ are the independent pseudotranslations in the cell, and δ is the Dirac delta function. The Fourier transform of $(C.1)$ gives

$$F_h = T[\rho_p(\mathbf{r})] + T[\rho_q(\mathbf{r})]$$

$$= T[\rho_{t_p}(\mathbf{r})] T \left[\sum_{\nu_1=0}^{n_1-1} \delta(\mathbf{r} - \nu_1\mathbf{u}_1) \right.$$

$$\left. * \sum_{\nu_2=0}^{n_2-1} \delta(\mathbf{r} - \nu_2\mathbf{u}_2) * \ldots \right] + (F_h)_q$$

$$= (F_h)_{t_p} \prod_i \left\{ \exp\left[2\pi i \frac{\mathbf{h}(n_i - 1)\mathbf{u}}{2} \right] \right.$$

$$\left. \times \frac{\sin n_i \pi \mathbf{h}\mathbf{u}_i}{\sin \pi \mathbf{h}\mathbf{u}_i} \right\} + (F_h)_q$$

$$= (F_h)_q + (F_h)_q$$

which agrees with (1).

Squaring of $(C.1)$ according to Sayre (1952) gives

$$F_h \simeq (\theta/V)\left\{ \sum_k (F_k)_p(F_{h-k})_p + \sum_k (F_k)_q(F_{h-k})_q \right.$$

$$\left. + 2\sum_k (F_k)_p(F_{h-k})_q \right\} \qquad (C.3)$$

where θ denotes the ratio of the averaged scattering factors to the squared ones. Since ρ_p and ρ_q are supposed to be uncorrelated, in $(C.1)$ it may be assumed that $\rho_p(\mathbf{r})\rho_q(\mathbf{r}) \simeq 0$, so that in $(C.3)$ the summation $\sum_k (F_k)_p(F_{h-k})_q$ is expected to be vanishing.

Therefore, $(C.3)$ reduces to

$$F_h \simeq (\theta/V)\left\{ \sum_k (F_k)_p(F_{h-k})_p + \sum_k (F_k)_q(F_{h-k})_q \right\}.$$

$$(C.4)$$

In accordance with Jeffery (1964), the values of the superstructure reflexions depend only on ρ_q while substructure reflexions depend both on ρ_p and ρ_q. If F_h in $(C.4)$ is a superstructure reflexion, then its value is only fixed by the contribution $\sum_k (F_k)_q(F_{h-k})_q$ and $(C.4)$ reduces to

$$F_h \simeq (\theta/V) \sum_k (F_k)_q(F_{h-k})_q. \qquad (C.5)$$

Both $(F_k)_q$ and $(F_{h-k})_q$ are known in practice only when $\mathbf{k}$ and $\mathbf{h}-\mathbf{k}$ correspond to superstructure reflexions. When $\mathbf{k}$ or $\mathbf{h}-\mathbf{k}$ refer to substructure reflexions then $(F_k)_q$ and $(F_{h-k})_q$ are parts of F_k and F_{h-k}, unknown but essential for securing Sayre's equation $(C.5)$ to be satisfied. Statistically speaking the average $\langle (F_k)_q(F_{h-k})_q \rangle$ relative to 'super-super' pairs is expected to be nearly equal to the averages relative to 'super-sub' or 'sub-sub' pairs. Consequently, the contributions F_kF_{h-k} relative to 'sub-sub' pairs are expected to be more sparsely spread around F_h than the contributions F_kF_{h-k} relative to 'sub-super' pairs. In their turn these are expected to be more sparsely spread around F_h than the contributions $F_kF_{h-k} = (F_k)_q(F_{h-k})_q$ relative to 'super-super' pairs. That agrees well with our theoretical results according to which

$$(N_{h,k})_2 > (N_{h,k})_1 > (N_{h,k})_0.$$

From the point of view introduced in this Appendix, the factors $(N_{h,k})_0$, $(N_{h,k})_1$, $(N_{h,k})_2$ involved in (4), (5), (6) appear to be nothing but statistical coefficients generated by a procedure aimed at applying a Cochran-like approach to the unknown moduli $|F_h|_q$, $|F_k|_q$, $|F_{h-k}|_q$ hidden in the known moduli $|F_h|$, $|F_k|$, $|F_{h-k}|$.

References

ADIWIDJAJA, G. (1972). Strukturbeziehungen in der Natrolith-Gruppe und das Entwaesserungsverkalten des Skolezeit. Dissertation, Univ. Hamburg.

BÖHME, R. (1982). Acta Cryst. A38, 316-326.

BÖHME, R. (1983). Z. Naturforsch. 38, 304-307.

CAMALLI, M., GIACOVAZZO, C. & SPAGNA, R. (1985). Acta Cryst. A41, 605-613.

CASCARANO, G., GIACOVAZZO, C., BURLA, M. C., NUNZI, A., POLIDORI, G., CAMALLI, M., SPAGNA, R. & VITERBO, D. (1985). IXth Eur. Crystallogr. Meet., Torino. Abstract 1-046.

CASCARANO, G., GIACOVAZZO, C. & LUIĆ, M. (1985a). In Statistical Crystallography. New York: Adenine Press.

CASCARANO, G., GIACOVAZZO, C. & LUIĆ, M. (1985b). Acta Cryst. A41, 544-551.

CHRISTIDIS, P. C. & RENTZEPERIS, R. J. (1975). Z. Kristallogr. 141, 233-245.

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

FAN HAI-FU, YAO JIA-XING, MAIN, P. & WOOLFSON, M. M. (1983). Acta Cryst. A39, 566-569.

GIACOVAZZO, C. (1980). Direct Methods in Crystallography. London: Academic Press.

GRAMLICH, V. (1975). Acta Cryst. A31, S90.

GRAMLICH, V. (1984). Acta Cryst. A40, 610-616.

HAUPTMAN, H. & KARLE, J. (1953). Solution of the Phase Problem. I. The Centrosymmetric Crystal. Am. Crystallogr. Soc. Monogr. No. 3. New York: Polycrystal Book Service.

HAUPTMAN, H. & KARLE, J. (1959). Acta Cryst. 12, 846-850.

ITO, T. & NOWACKI, W. (1974). Z. Kristallogr. 139, 85-102.

JEFFERY, J. W. (1964). Acta Cryst. 17, 776-777.

LIPSON, H. & WOOLFSON, M. M. (1952). Acta Cryst. 5, 680-682.

PONTENAGEL, W. M. G. F. & KRABBENDAM, H. (1983). Acta Cryst. A39, 333-340.

PRICK, P. A. J., BEURSKENS, P. I. & GOULD, R. O. (1983). Acta Cryst. A39, 570-576.

ROGERS, D. & WILSON, A. J. C. (1953). Acta Cryst. 6, 439-449.

SAYRE, D. (1952). Acta Cryst. 5, 60-65.

Figures of Merit in Direct Methods: a New Point of View

BY G. CASCARANO AND C. GIACOVAZZO

Dipartimento Geomineralogico, Università, Campus Universitario, Via Salvemini, 70124 Bari, Italy

AND D. VITERBO

Istituto di Chimica Fisica, Università, Corso M. D'Azeglio 48, 10125 Torino, Italy

Abstract

The probabilistic approach has been extensively used for analysing the statistical meanings of some traditional figures of merit. New figures of merit have also been introduced; some exploit one-phase and two-phase structure seminvariants of the first rank, the expected negative and/or enantiomorph-sensitive triplet invariants, and the expected negative and/or enantiomorph-sensitive quartets. Other figures of merit exploit the distribution of statistical parameters connected with PSI(0) and active triplets. A new combined figure of merit is shown to be a powerful tool for selecting the correct solutions among the various sets output by multisolution methods.

1. Introduction

For the determination of very complex structures by direct methods a large initial set of known phases seems to be a basic requirement. This aim can be achieved by introducing a large number of permutable phases which are used to generate different phase sets. Magic integer sequences (Main, 1977) are often employed for phase permutations: a relatively large number of phase sets are thereby created among which the correct solutions have to be found. Figures of merit (FOM) are usually used to screen the set of solutions, prior to computing Fourier transforms (E map).

In general, FOM's are functions based on quantities which can be expected to have extreme values for the correct solution. The expectation relies on a probabilistic background and/or on algebraic properties. In the SIR program (Nunzi et al., 1984), several low-order structure seminvariants and invariants are estimated by means of representation theory (Giacovazzo, 1977, 1980a). Some of them are actively used for phase expansion and refinement, others are only employed to compute the FOM's. Since one or more FOM's are available for each type of structure seminvariant or invariant, the combined figure of merit CPHASE, based on a variety of FOM's, is expected to be effective in finding the correct solution [see Hašek, Schenk, Kiers & Schagen (1985) for some tests of distribution-fitting methods for centrosymmetric structures].

A probabilistic approach is also introduced which enables us to analyse the statistical meanings of some traditionally widely used FOM's. New effective figures of merit are devised which, combined with CPHASE, give rise to a reliable total combined figure of merit CFOM which is expected to be unity for the correct solutions.

2. The combined figure of merit CPHASE

Overbeek & Schenk (1976) first proposed a FOM based on $\sum_1$ relationships. In the SIR program the estimates of the one-phase structure seminvariants of