Acta Cryst. (1992). A**48**, 859-865

Improved Figures of Merit for Direct Methods

By G. Cascarano, C. Giacovazzo and A. Guagliardi

Istituto di Ricerca per lo Sviluppo di Metodologie Cristallografiche CNR, c/o Dipartimento Geomineralogico, Campus Universitario, 70124 *Bari, Italy*

(*Received* 16 *December* 1991; *accepted* 13 *April* 1992)

Abstract

The figures of merit (FOMs) proposed in a recent paper [Cascarano, Giacovazzo & Viterbo (1987). *Acta Cryst.* A**43**, 22-29] are modified to improve their efficiency. It is shown that FOMs based on psi-zero triplets and FOMs based on the distribution of the concentration parameter α can be treated by the same mathematical formalism. An earlier FOM, the $R_{\rm Karle}$ criterion, is revised and suggestions are given for its efficient use. The use of the correlation coefficient is also suggested, provided large- and small-intensity reflections are used simultaneously.

1. Introduction

Multisolution methods *via* magic-integer sequences (White & Woolfson, 1975; Declercq, Germain & Woolfson, 1975; Main, 1977) or *via* random starting sets (Baggio, Woolfson, Declercq & Germain, 1978; Yao, 1981) are very efficient procedures for crystal structure determination. A relatively large number of phase sets are created among which the correct solutions are recognized by special figures of merit (FOMs). FOMs rely on a probabilistic background and/or on algebraic properties and are expected to be maxima or minima for the correct solutions.

Some FOMs are based on the agreement between the estimated and the assigned values of structure invariants and seminvariants. Accordingly, the general formula

$$\mathrm{SSFOM}(i)=\sum_j w_j G_j \cos(\Phi_j-\theta_j)\Big/\sum_j w_j D_1(G_j)$$
$$=T_i/B_i \qquad \text{for } i=1,2,3,4$$

may be used (Cascarano, Giacovazzo & Viterbo, 1987), where Φ and θ are the expected and the assigned values of the structure invariant or seminvariant, G is the reliability parameter of the phase relationship and $D_1(x)=I_1(x)/I_0(x)$, where I_0, I_1 are the modified Bessel functions of order 0 and 1, respectively. The denominator of SSFOM is the expected value of the numerator. For $i=1,2$, SSFOM denotes one-phase (Overbeek & Schenk, 1976) and two-phase (Burla, Giacovazzo & Polidori, 1989) structure seminvariants, respectively. For $i=3$, SSFOM is calculated for triplets estimated to be negative *via* their second representation (Cascarano, Giacovazzo, Camalli, Spagna, Burla, Nunzi & Polidori, 1984). For $i=4$, SSFOM involves negative-quartet invariants (Schenk, 1974; De Titta, Edmonds, Langs & Hauptman, 1975; Giacovazzo, 1976).

The information contained in the various SSFOMs can be combined into (Cascarano, Giacovazzo & Viterbo, 1987)

$$\mathrm{CPHASE}=(T_1+T_2+T_3+T_4)/(B_1+B_2+B_3+B_4). \qquad (1)$$

CPHASE automatically weights the importance of the individual FOMs. In particular, it will be dominated by the FOM(s) that involve(s) a large number of reliable phase relationships. For example, in structures for which one-phase structure seminvariants are estimated to be present in small numbers and with low reliability, the contribution of T_1 and B_1 to CPHASE will be negligible.

FOMs of a different nature (Cochran & Douglas, 1957; Karle & Karle, 1966; Declercq, Germain & Woolfson, 1979) are the traditional

$$\mathrm{ABSFOM}=\Big[\sum_{\mathbf{h}}\alpha_{\mathbf{h}}-\sum_{\mathbf{h}}(\alpha_{\mathbf{h}})_{\rm rand}\Big]$$
$$\times\Big[\sum_{\mathbf{h}}\langle(\alpha_{\mathbf{h}})\rangle-\sum_{\mathbf{h}}(\alpha_{\mathbf{h}})_{\rm rand}\Big]^{-1}, \qquad (2)$$

$$\mathrm{MABS}=\sum_{\mathbf{h}}\alpha_{\mathbf{h}}\Big/\sum_{\mathbf{h}}\langle\alpha_{\mathbf{h}}\rangle, \qquad (3)$$

$$\mathrm{RALPHA}=\sum_{\mathbf{h}}|\alpha_{\mathbf{h}}-\langle\alpha_{\mathbf{h}}\rangle|\Big/\sum_{\mathbf{h}}\langle\alpha_{\mathbf{h}}\rangle, \qquad (4)$$

$$\mathrm{PSI}(0)=\sum_{\mathbf{h}}\alpha'_{\mathbf{h}}\Big/\sum_{\mathbf{h}}v_{\mathbf{h}}^{1/2}, \qquad (5)$$

where

$$\alpha_{\mathbf{h}}=\Big[\Big(\sum_{j=1}^{r}G_j\cos\theta_j\Big)^2+\Big(\sum_{j=1}^{r}G_j\sin\theta_j\Big)^2\Big]^{1/2},$$
$$G_j=2|E_{\mathbf{h}}E_{\mathbf{k}_j}E_{\mathbf{h}-\mathbf{k}_j}|/N^{1/2},$$
$$\theta_j=\theta_{\mathbf{k}_j}+\theta_{\mathbf{h}-\mathbf{k}_j},$$

$\langle\alpha_{\mathbf{h}}\rangle$ is the expected value of the $\alpha_{\mathbf{h}}$, $(\alpha_{\mathbf{h}})_{\rm rand}$ is the expected value of $\alpha_{\mathbf{h}}$ for which the phases $\theta_{\mathbf{k}_j}$ and

0108-7673/92/060859-07$06.00

$\theta_{\mathbf{h}-\mathbf{k}_j}$ are assumed to be randomly distributed. In (5),

$$\alpha'_{\mathbf{h}}=\left[\left(\sum_{j=1}^{r}A'_j\cos\theta_j\right)^2+\left(\sum_{j=1}^{r}A'_j\sin\theta_j\right)^2\right]^{1/2},$$

where $\mathbf{h}$ is a reciprocal-lattice vector with $|E_{\mathbf{h}}|\simeq 0$,

$$v_{\mathbf{h}}=\sum_{j=1}^{r}A'^2_j,$$

$$A'_j=|E_{\mathbf{k}_j}E_{\mathbf{h}-\mathbf{k}_j}|/N^{1/2},$$

$\mathbf{k}_j$ and $\mathbf{h}-\mathbf{k}_j$ are indices of strong $|E|$ values for which phases have been determined.

A new point of view for the above FOMs was introduced by Cascarano, Giacovazzo & Viterbo (1987), who made use of the probability distribution function

$$\begin{aligned}P(\alpha_{\mathbf{h}})&\simeq N(\alpha_{\mathbf{h}};\langle\alpha_{\mathbf{h}}\rangle,\sigma^2_{\alpha_{\mathbf{h}}})\\&\simeq[(2\pi)^{1/2}\sigma_{\mathbf{h}}]^{-1}\exp[-(\alpha_{\mathbf{h}}-\langle\alpha_{\mathbf{h}}\rangle)^2/2\sigma^2_{\mathbf{h}}],\end{aligned}\qquad(6)$$

where

$$\langle\alpha_{\mathbf{h}}\rangle=\sum_{j=1}^{r}g_jD_1(G_j),$$

$$\sigma^2_{\mathbf{h}}=\tfrac{1}{2}\sum_{j=1}^{r}G_j^2[1+D_2(G_j)-2D_1^2(G_j)],$$

$$D_i(x)=I_i(x)/I_0(x)$$

and $I_i(x)$ is the modified Bessel function of order i.

The distribution (6) allowed the use of the first and of the second moments of $\alpha_{\mathbf{h}}$. Accordingly, FOMs (2)-(4) were replaced by

$$\mathrm{NALF}(1)=(nq_1)^{-1}\sum_{\mathbf{h}}[|S\alpha_{\mathbf{h}}-\langle\alpha_{\mathbf{h}}\rangle|/\sigma_{\mathbf{h}}],\qquad(7)$$

$$\mathrm{NALF}(2)=q_2^{-1}\sum_{\mathbf{h}}(S\alpha_{\mathbf{h}}-\langle\alpha_{\mathbf{h}}\rangle)^2\Big/\sum_{\mathbf{h}}\sigma^2_{\mathbf{h}},\qquad(8)$$

$$\mathrm{NALF}(3)=(nq_3)^{-1}\sum_{\mathbf{h}}[(S\alpha_{\mathbf{h}}-\langle\alpha_{\mathbf{h}}\rangle)/\sigma_{\mathbf{h}}]^2,\qquad(9)$$

$$\mathrm{NALF}(4)=(nq_4)^{-1}\sum_{\mathbf{h}}|[(S\alpha_{\mathbf{h}}-\langle\alpha_{\mathbf{h}}\rangle)/\sigma_{\mathbf{h}}]^2-1|,\qquad(10)$$

where $S=1/\mathrm{MABS}$ [see (3)], n is the number of reflections $\mathbf{h}$ involved in the summations and

$$q_1=0.798,\quad q_2=1,\quad q_3=1,\quad q_4=0.968.$$

Equations (7)-(10) can be combined into the FOM

$$\mathrm{ALFCOMB}=\tfrac{1}{4}\left\{\mathrm{NALF}(1)+\sum_{i=2}^{4}[\mathrm{NALF}(i)]^{1/2}\right\},\qquad(11)$$

which is expected to be minimum for the correct solution.

The same point of view was applied by Cascarano, Giacovazzo & Viterbo (1987) to the FOMs based on $\alpha'_{\mathbf{h}}$. The distribution

$$P(\alpha'_{\mathbf{h}})=(2\alpha'_{\mathbf{h}}/v_{\mathbf{h}})\exp(-\alpha'^2_{\mathbf{h}}/v_{\mathbf{h}})\qquad(12)$$

was used to derive the new FOMs

$$\mathrm{NPS}(0)=q_0^{-1}\sum_{\mathbf{h}}\alpha'_{\mathbf{h}}\Big/\sum_{\mathbf{h}}v_{\mathbf{h}}^{1/2},\qquad(13)$$

$$\mathrm{NPS}(1)=(nq_1)^{-1}\sum_{\mathbf{h}}\alpha'^2_{\mathbf{h}}/v_{\mathbf{h}},\qquad(14)$$

$$\mathrm{NPS}(2)=(nq_2)^{-1}\sum_{\mathbf{h}}|\alpha'^2_{\mathbf{h}}/v_{\mathbf{h}}-1|,\qquad(15)$$

$$\mathrm{NPS}(3)=(nq_3)^{-1}\sum_{\mathbf{h}}(\alpha'_{\mathbf{h}}/v_{\mathbf{h}}^{1/2}-\pi^{1/2}/2)^2,\qquad(16)$$

where, for noncentrosymmetric space groups,

$$q_0=0.886,\quad q_1=1,\quad q_2=0.736,\quad q_3=0.215,$$

and, for centrosymmetric space groups,

$$q_0=0.798,\quad q_1=1,\quad q_2=0.968,\quad q_3=0.363.$$

Equations (13)-(16) can be combined into the FOM

$$\mathrm{PSCOMB}=\tfrac{1}{4}\left\{\mathrm{NPS}(0)+\sum_{i=1}^{3}[\mathrm{NPS}(i)]^{1/2}\right\},\qquad(17)$$

which is expected to be minimum for the correct solution. It was also shown that there is strong correlation between MABS and PSCOMB, so that the final form of PSCOMB was

$$\mathrm{PSCOMB}=\frac{1}{4(\mathrm{MABS})}\left\{\mathrm{NPS}(0)+\sum_{i=1}^{3}[\mathrm{NPS}(i)]^{1/2}\right\}.\qquad(18)$$

In the program *SIR*88 (Burla, Camalli, Cascarano, Giacovazzo, Polidori, Spagna & Viterbo, 1989), PSCOMB, ALFCOMB, MABS and CPHASE may be collated to produce the overall combined figure of merit

$$\begin{aligned}\mathrm{CFOM}=\left(\sum_{i=1}^{4}w_i\right)^{-1}&\\\times\{w_1\mathrm{DABS}&+w_2\exp[-(1-\mathrm{CPHASE})^{3/2}]\\&+w_3\exp[-(\mathrm{ALFCOMB}-1)^{3/2}]\\&+w_4\exp[-(\mathrm{PSCOMB}-1)^{3/2}]\},\end{aligned}\qquad(19)$$

where $\mathrm{DABS}=1-|\mathrm{MABS}-1|$. CFOM is expected to be equal to 1 for the correct structure.

Even if (19) constitutes progress with respect to previous formulations it has a basic drawback: the weights w_i are determined by the prior confidence of the user in the various FOMs. We will show in § 2 that the structural complexity, the particular nature of the structure under investigation and the efficiency of the phasing process frequently generate situations in which application of the standard weights w_i is senseless. In § 3 we will show that PSCOMB and

Table 1. *Code name, space group and crystallochemical data for test structures*

Structure code	Space group	Molecular formula	Z
APAPA(1)	$P4_12_12$	$C_{30}H_{37}N_{15}O_{16}P_2.6H_2O$	8
QUINOL(2)	$R\bar{3}$	$C_6H_6O_2$	54
TPH(3)	$B22_12$	$C_{24}H_{20}N_2$	12
TUR10(4)	$P6_322$	$C_{15}H_{24}O_2$	12
GRA4(5)	$P\bar{1}$	$C_{30}H_{22}N_2O_4$	2
BED(6)	$I4$	$C_{26}H_{26}N_4O_4$	8
CEPHAL(7)	$C2$	$C_{18}H_{21}NO_4$	8
FEGAS(8)	$P6_3/mmc$	$Fe_2Ga_2S_5$	2
HOVI(9)	$C2/m$	$Pr_{14}Ni_{16}Si_{11}$	4
INOS(10)	$P2_1/n$	$C_6H_{12}O_6.H_2O$	8
LOGANIN(11)	$P2_12_12_1$	$C_{17}H_{26}O_{10}$	4
NO55(12)	$Fdd2$	$C_{20}H_{24}N_4$	16
POCRO(13)	$B112/m$	Cr_5KSe_8	2
BOBBY(14)	$P2_13$	$Na^+Ca^{2+}N(CH_2CO_2)_3^{3-}$	4
MGHEX(15)	$P3_1$	$C_{48}H_{68}MgN_{12}O_{12}.2ClO_4.4CH_3CN$	3
AZET(16)	$Pca2_1$	$C_{21}H_{16}ClNO$	8
DIOL(17)	$I\bar{4}2d$	$C_{10}H_{18}O_2$	16
FREIES(18)	$P2_1/a$	$PbAgSbS_3$	4

References: (1) Suck, Manor & Saenger (1976); (2) Wallwork & Powell (1980); (3) Hoekstra, Vos, Braun & Hornstra (1975); (4) Braekman, Daloze, Dupont, Tursch, Declercq, Germain & Van Meerssche (1981); (5) unpublished; (6) Sheldrick, Davison & Trotter (1978); (7) Arora, Bates, Grady, Germain, Declercq & Powell (1976); (8) Cascarano, Dogguy-Smiri & Nguyen-Huy Dung (1987); (9) Hovestreydt, Klepp & Parthé (1983); (10) Langs, Freeman, Nockolds & Oh, unpublished; (11) Jones, Sheldrick, Glüsenkamp & Tietze (1980); (12) Sheldrick & Trotter (1978); (13) Nguyen-Huy Dung, Vo-Van Tien, Behm & Beurskens (1987); (14) Barnett & Uchtman (1979); (15) Karle & Karle (1981); (16) Colens, Declercq, Germain, Putzeys & Van Meerssche (1974); (17) unpublished; (18) Ito & Novacki (1974).

ALFCOMB can be treated by a common mathematical formalism; accordingly, PSCOMB will be considered as part of a modified ALFCOMB. In § 4 the combination of ALFCOMB with CPHASE is described. In §§ 5 and 6 modified versions of the R_{Karle} criterion and the correlation coefficient are proposed as efficient figures of merit.

2. The weighting of the traditional FOMs

In Table 1 code names, space groups and crystallochemical data for the test structures are given. Among them we select for this section BOBBY, FREIES, BED, DIOLE and POCRO.

BOBBY is a structure with the pseudotranslational vector $\mathbf{u} = (\mathbf{a}+\mathbf{b}+\mathbf{c})/2$. SIR88 selected 68 (NLAR) strong reflections for active use and 39 (NSMALL) weak ones for calculating psi-zero triplets and for subsequent use in the negative-quartet routine. If prior information about pseudosymmetry is not taken into account, the *SIR*88 phasing process ends with 6 unphased strong reflections. In this situation only 11 weak reflections contribute to PSCOMB, which is therefore a highly nonselective FOM.

A limit situation occurs for FREIES if the pseudotranslational symmetry ($\mathbf{u} = \mathbf{a}/2 + \mathbf{b}/3$) is not considered as prior information. Then 76 phases (most of them correspond to substructure reflections) over NLAR = 175 are not determined. No psi-zero triplets can be used and the 'correct' solution is only determined by ALFCOMB.

For BED, *SIR*88 chooses NLAR = 286 and NSMALL = 100, for which 4000 psi-zero triplets are calculated. PSCOMB and CPHASE are highly selective while ALFCOMB is misleading: the highest values of ALFCOMB correspond to false structures for which the enantiomorph is missing.

For POCRO (pseudotranslational vector $\mathbf{u} = \mathbf{a}/6 + \mathbf{c}/2$), SIR88 chooses NLAR = 184 and NSMALL = 82. 500 highly reliable negative quartets (the average concentration parameter of their Von Mises distribution is $\langle G_4 \rangle = 2.9$) are used by CPHASE, which selects the correct structure. The opposite situation occurs for DIOLE, where 182 strong and 35 weak reflections are selected by *SIR*88. Only 84 not very reliable negative quartets ($\langle G_4 \rangle = 0.57$) are used by CPHASE, which proves to be a nonselective FOM.

The above examples show that an ideal weighting scheme for FOMs is one that automatically makes use of the various parameters available at the end of the phasing process (*i.e.* number of phased reflections, number of available phase relationships, their reliability, lack of enantiomorph *etc.*).

3. The integration of PSCOMB with ALFCOMB

ALFCOMB and PSCOMB are based on the distributions (6) and (12), which involve the variables $\alpha_{\mathbf{h}}$ and $\alpha'_{\mathbf{h}}$, respectively. The following assumptions (a) and (b) are the bases for the mathematical derivation of (6) and (12), respectively (Burla, Cascarano, Giacovazzo, Nunzi & Polidori, 1987).

(a) Calculate the distribution of the modulus $\alpha_{\mathbf{h}}$ of the resultant of r complex vectors $G_j \exp(i\theta_j)$ under the hypothesis that the θ_j are distributed according to the Von Mises function

$$M(\theta_j;\, \varphi_{\mathbf{h}},\, G_j) = [2\pi I_0(G_j)]^{-1} \exp[G_j \cos(\theta_j - \varphi_{\mathbf{h}})].$$

(b) Calculate the distribution of the modulus $\alpha'_{\mathbf{h}}$ of the resultant of r complex vectors $A'_j \exp(i\theta_j)$ under the hypothesis that the θ_j are distributed according to $M(\theta_j;\, \varphi_{\mathbf{h}}, 0)$.

Let us now introduce the variable

$$A''_j = 2|E_{\mathbf{k}_j} E_{\mathbf{h}-\mathbf{k}_j}|/N^{1/2} \qquad (20)$$

and apply to it statistical calculations similar to those described in (a) and (b). We look for the distribution of the modulus $\alpha''_{\mathbf{h}}$ of the resultant of r complex vectors $A''_j \exp(i\theta_j)$ under the hypothesis that the θ_j (strong triplets) are distributed according to $M(\theta_j;\, \varphi_{\mathbf{h}},\, G_j)$. Then we have (the derivations are omitted for brevity)

$$P(\alpha''_{\mathbf{h}}) = N(\alpha''_{\mathbf{h}};\, \langle \alpha''_{\mathbf{h}} \rangle,\, \sigma''^2_{\mathbf{h}}), \qquad (21)$$

where

$$\langle\alpha''_{\mathbf{h}}\rangle=\sum_{j=1}^{r}A''_j D_1(G_j),$$

$$\sigma''^2_{\mathbf{h}}=\tfrac{1}{2}\sum_{j=1}^{r}A''^2_j[1+D_2(G_j)-2D_1^2(G_j)].$$

The FOMs (7)-(10) can again be used provided that $\alpha_{\mathbf{h}}$, $\langle\alpha_{\mathbf{h}}\rangle$ and $\sigma_{\mathbf{h}}$ are replaced by $\alpha''_{\mathbf{h}}$, $\langle\alpha''_{\mathbf{h}}\rangle$ and $\sigma''_{\mathbf{h}}$, respectively.

The formalism described above can be applied without any modification to the so-called psi-E triplets (Altomare, Cascarano, Giacovazzo & Viterbo, 1991). In a modified version of *SIR*88, the NEXP reflections immediately following (in decreasing order of $|E|$) the NLAR reflections are selected (NEXP$\simeq$0.8 NLAR). The psi-E triplets (they relate one of the NEXP reflections to two of the NLAR reflections) are set up to allow the phasing of NEXP reflections and so reduce the truncation errors of the Fourier maps. The information (in terms of phases and of structure-factor moduli) available for the NEXP reflections is the same as that for the NLAR reflections. Thus FOMs arising from psi-E triplets can be combined with those considered in (7)-(10) simply by allowing the reflection $\mathbf{h}$ to vary over NLAR+NEXP reflections.

We now look for the distribution of the modulus $\alpha''_{\mathbf{h}}$ of the resultant of r complex vectors $A''_j\exp(i\theta_j)$ when the θ_j are distributed according to $M(\theta_j;\varphi_{\mathbf{h}},0)$ (*i.e.* randomly distributed, as for psi-zero triplets). Since $A''_j=2A'_j$, $\alpha''_{\mathbf{h}}$ will be distributed as $\alpha'_{\mathbf{h}}$ in (12): this distribution may also be approximated by a normal distribution with the same first and second moments.

Then (21) still holds provided:

(a) $\sigma''^2_{\mathbf{h}}=4v_{\mathbf{h}}$;

(b) $\langle\alpha''_{\mathbf{h}}\rangle$ is calculated as $\langle\alpha''_{\mathbf{h}}\rangle=0.886\sigma''^{1/2}_{\mathbf{h}}$ for noncentrosymmetric space groups and $\langle\alpha''_{\mathbf{h}}\rangle=0.798\sigma''^{1/2}_{\mathbf{h}}$ for centrosymmetric space groups.

Then the FOMs (7)-(10) [and consequently (11)] expressed in terms of α'' and $\sigma''_{\mathbf{h}}$ can be used for strong psi-E and psi-zero triplets simply by allowing the reflection $\mathbf{h}$ to vary over NLAR+NEXP+NSMALL. In accordance with (18), the contribution arising from the NSMALL reflections is divided by MABS before being added to the contribution from the NLAR and NEXP reflections.

It should also be noted that a maximization of the new NALF(3), say

$$\mathrm{NALF}(3)=(nq_3)^{-1}\sum_{\mathbf{h}}[(S\alpha''_{\mathbf{h}}-\langle\alpha''_{\mathbf{h}}\rangle)/\sigma_{\mathbf{h}}]^2,$$

corresponds to maximizing the joint probability distribution

$$P(\alpha''_{\mathbf{h}_1},\alpha''_{\mathbf{h}_2},\dots)=\prod_{\mathbf{h}}P(\alpha''_{\mathbf{h}}).$$

Table 2. *CFOM values*

For each test structure the values of CFOM are given for the correct solution found by the modified version of *SIR*88 (CFOM^C), for the incorrect solution with the highest value of CFOM (CFOM^I) and for the published phases (CFOM^P).

	CFOM^C	CFOM^I	CFOM^P
APAPA	0.993	0.943	0.996
QUINOL	0.981	0.202	0.982
TPH	0.987	0.382	0.987
TUR10	0.972	0.511	0.965
GRA4	0.757	0.677	0.758
BED	0.532	0.315	0.545
CEPHAL	1.000	0.563	1.000
FEGAS	0.899	0.613	0.899
HOV1	0.732	0.729	0.727
INOS	1.000	0.902	1.000
LOGANIN	0.993	0.758	0.999
NO55	0.695	0.576	0.736
POCRO	0.847	0.254	0.846
BOBBY	0.360	0.184	0.634
MGHEX	0.832	0.700	0.861
AZET	0.895	0.749	1.000
DIOL	0.838	0.651	0.797
FREIES	0.155	0.000	0.976

4. A combined FOM

CPHASE and the new ALFCOMB may be collated in the combined FOM

$$\mathrm{CFOM}=(1+w)^{-1}\{\exp[-(\mathrm{ALFCOMB}-1)^{3/2}]+w\exp[-(\mathrm{CPHASE}-1)^{3/2}]\},\qquad(22)$$

where $w=[D_1(\langle G\rangle)]^{1/2}$. $\langle G\rangle$ is the average value of G calculated for all the phase relationships (usually estimated negative-triplet and negative-quartet invariants). The confidence in CPHASE now depends on the average phase reliability: this is expected to be small for large structures.

A modified version of *SIR*88 has been used to apply (22) to the test structures quoted in Table 1. We chose NSMALL = NLAR/3 with the supplementary condition that the maximum value of $|E|$ is less than 0.4, and NEXP = NLAR/2. Such a choice proved to be reasonable: NEXP is sufficiently large to make ALFCOMB more robust but not so large as to involve in ALFCOMB too many unreliable phases. NSMALL is also sufficiently large to contribute significantly to ALFCOMB, but not so large to involve reflections with $|E|$ remarkably different from zero. In Table 2 we give for each test structure: (a) the value of CFOM for the correct structure (CFOM^C); (b) the highest value of CFOM corresponding to an incorrect solution (CFOM^I); (c) the value of CFOM calculated from the published phases (CFOM^P). We note the following.

(i) The correct solution is always identified by CFOM.

(ii) The value of CFOM, expected to be equal to unity for the correct structure, is in most cases close to unity. Remarkable deviations occur when pseudotranslational symmetry is present.

(iii) The discrimination between correct and incorrect solutions is satisfactory in most cases.

5. The use of the $R_{\rm Karle}$ figure of merit

According to Karle & Karle (1966)

$$|E_{\mathbf{h}}|_{\rm calc}\cos\varphi_{\mathbf{h}}=2n_{\mathbf{h}}^{-1}N^{-1/2}\sum_j|E_{\mathbf{k}_j}E_{\mathbf{h}-\mathbf{k}_j}|\times\cos(\varphi_{\mathbf{k}_j}+\varphi_{\mathbf{h}-\mathbf{k}_j}),\qquad(23a)$$

$$|E_{\mathbf{h}}|_{\rm calc}\sin\varphi_{\mathbf{h}}=2n_{\mathbf{h}}^{-1}N^{-1/2}\sum_j|E_{\mathbf{k}_j}E_{\mathbf{h}-\mathbf{k}_j}|\times\sin(\varphi_{\mathbf{k}_j}+\varphi_{\mathbf{h}-\mathbf{k}_j}),\qquad(23b)$$

where $n_{\mathbf{h}}$ is the number of terms in the summations. The necessary correlation between $|E_{\mathbf{h}}|_{\rm calc}$ and the observed value $|E_{\mathbf{h}}|$ suggested to Karle & Karle the following criterion:

$$R_{\rm Karle}=\sum_{\mathbf{h}}\big||E_{\mathbf{h}}|-K|E_{\mathbf{h}}|_{\rm calc}\big|\Big/\sum_{\mathbf{h}}|E_{\mathbf{h}}|=\text{minimum}\qquad(24)$$

for the correct structure. K is a scale factor defined by $K\sum_{\mathbf{h}}|E_{\mathbf{h}}|_{\rm calc}=\sum_{\mathbf{h}}|E_{\mathbf{h}}|$.

In most direct-methods packages the $R_{\rm Karle}$ criterion has been calculated from the NLAR reflections only; it was useful on several occasions but inconclusive for others. We show here that the $R_{\rm Karle}$ criterion may become more powerful when extended to NLAR+NEXP+NSMALL reflections, provided care is taken of the specific nature of the different types of reflections.

On combining (23a) with (23b) we obtain

$$\alpha''_{\mathbf{h}}=n_{\mathbf{h}}|E_{\mathbf{h}}|_{\rm calc}.$$

$E_{\mathbf{h}\,\rm calc}$ can be put on the same scale as $|E_{\mathbf{h}}|$ by using

$$\sum_{\mathbf{h}}\langle\alpha''_{\mathbf{h}}\rangle=K\sum_{\mathbf{h}}|E_{\mathbf{h}}|,$$

from which

$$K=\sum_{\mathbf{h}}\langle\alpha''_{\mathbf{h}}\rangle\Big/\sum_{\mathbf{h}}|E_{\mathbf{h}}|.\qquad(25)$$

We use two different scale factors: one for the NLAR+NEXP reflections and one for the NSMALL reflections. The $\langle\alpha''_{\mathbf{h}}\rangle$ used in the two different cases were given in § 3. Rescaled $|E_{\rm calc}|$ values can then be used in the FOM

$$R_{\rm Karle}=\sum_{\mathbf{h}}{}'\big||E_{\mathbf{h}}|-|E_{\mathbf{h}}|_{\rm calc}\big|\Big/\sum_{\mathbf{h}}{}'|E_{\mathbf{h}}|,\qquad(26)$$

where the summation $\sum'$ is made over the NLAR+NEXP+NSMALL reflections. The robustness of (26) may be deduced from the results in Table 3. For each test structure we give, as calculated by (26), the values of $R^C_{\rm Karle}$, $R^I_{\rm Karle}$ and $R^P_{\rm Karle}$, where C, I and P denote the correct solution, the incorrect solution with the smallest value of $R_{\rm Karle}$ and the published phases, respectively. The corresponding values obtained by (24) are given in parentheses in Table 3. We note that (a) use of (26) always indicates the correct solution; (b) in four cases (QUINOL, BED, CEPHAL, HOV1), use of (24) gives the smallest value of $R_{\rm Karle}$ for an incorrect solution; (c) the discriminating power of (26) is in general better than that of (24). In one case (HOV1), identical values of $R_{\rm Karle}$ are found with (26) for the correct and an incorrect solution, while use of (24) leads to identical values in two cases (GRA4, POCRO).

Table 3. *R_{Karle} values*

For each test structure the values of $R_{\rm Karle}$ calculated by (26) are given for the correct solution ($R^C_{\rm Karle}$), for the incorrect solution with the smallest value of $R_{\rm Karle}$ ($R^I_{\rm Karle}$) and for the published phases ($R^P_{\rm Karle}$). The values of $R_{\rm Karle}$ calculated according to (24) are given in parentheses.

	$R^C_{\rm Karle}$	$R^I_{\rm Karle}$	$R^P_{\rm Karle}$
APAPA	0.364 (0.356)	0.404 (0.386)	0.367 (0.351)
QUINOL	0.227 (0.232)	0.395 (0.230)	0.227 (0.232)
TPH	0.194 (0.176)	0.338 (0.306)	0.194 (0.176)
TUR10	0.369 (0.253)	0.477 (0.348)	0.368 (0.247)
GRA4	0.247 (0.225)	0.250 (0.225)	0.245 (0.225)
BED	0.318 (0.303)	0.356 (0.197)	0.334 (0.334)
CEPHAL	0.185 (0.165)	0.247 (0.142)	0.173 (0.161)
FEGAS	0.247 (0.184)	0.285 (0.215)	0.247 (0.184)
HOV1	0.244 (0.226)	0.244 (0.225)	0.244 (0.226)
INOS	0.214 (0.228)	0.274 (0.280)	0.214 (0.230)
LOGANIN	0.199 (0.188)	0.267 (0.257)	0.183 (0.177)
NO55	0.316 (0.286)	0.335 (0.298)	0.298 (0.282)
POCRO	0.338 (0.242)	0.370 (0.242)	0.342 (0.242)
BOBBY	0.220 (0.194)	0.377 (0.385)	0.232 (0.187)
MGHEX	0.297 (0.269)	0.315 (0.295)	0.331 (0.296)
AZET	0.375 (0.378)	0.419 (0.413)	0.341 (0.378)
DIOL	0.305 (0.290)	0.427 (0.414)	0.293 (0.287)
FREIES	0.177 (0.174)	0.876 (0.877)	0.226 (0.191)

6. The use of the correlation coefficient

The use of the correlation coefficient

$$\rho=\sum_{\mathbf{h}}\left[(|E_{\mathbf{h}}|_{\rm calc}-\langle|E_{\mathbf{h}}|_{\rm calc}\rangle)(|E_{\mathbf{h}}|-\langle|E_{\mathbf{h}}|\rangle)\right]\times\left[\sum_{\mathbf{h}}(|E_{\mathbf{h}}|_{\rm calc}-\langle|E_{\mathbf{h}}|_{\rm calc}\rangle)^2\sum_{\mathbf{h}}(|E_{\mathbf{h}}|-\langle|E_{\mathbf{h}}|\rangle)^2\right]^{-1/2}\qquad(27)$$

as a FOM for direct methods has never been popular.

Table 4. *ρ values for APAPA corresponding to a selection of* 112 *trials as calculated for NLAR reflections*

Trial 100 gives the correct solution.

Trial	ρ
1	−0.066
10	−0.129
20	−0.091
30	−0.041
40	−0.124
50	−0.080
60	0.000
70	−0.136
80	−0.101
90	−0.156
100	−0.096
110	−0.104
112	−0.126

Table 5. *ρ values*

For each test structure the values of ρ as calculated for the correct structure (ρ^C), for the incorrect solution with highest values of ρ (ρ^I) and for the published phases (ρ^P) are given. NLAR+NEXP+NSMALL reflections are used. The corresponding values calculated from NLAR reflections only are given in parentheses.

	ρ^C	ρ^I	ρ^P
APAPA	0.599	0.548	0.611
	(−0.087)	(0.110)	(−0.085)
QUINOL	0.828	0.596	0.828
	(0.110)	(0.166)	(0.110)
TPH	0.824	0.528	0.826
	(0.361)	(0.373)	(0.337)
TUR10	0.559	0.408	0.549
	(0.337)	(0.341)	(0.334)
GRA4	0.813	0.809	0.816
	(0.494)	(0.494)	(0.494)
BED	0.631	0.520	0.590
	(0.224)	(0.201)	(0.235)
CEPHAL	0.801	0.477	0.831
	(0.313)	(0.173)	(0.301)
FEGAS	0.820	0.710	0.820
	(0.381)	(0.386)	(0.381)
HOV1	0.754	0.754	0.757
	(0.482)	(0.482)	(0.482)
INOS	0.820	0.729	0.817
	(0.012)	(0.071)	(0.026)
LOGAN	0.808	0.664	0.837
	(0.267)	(0.303)	(0.280)
NO55	0.641	0.543	0.690
	(0.269)	(0.209)	(0.266)
POCRO	0.705	0.670	0.728
	(0.204)	(0.228)	(0.204)
BOBBY	0.673	0.436	0.811
	(0.327)	(0.254)	(0.447)
MGHEX	0.579	0.553	0.566
	(−0.016)	(0.202)	(−0.005)
AZET	0.500	0.424	0.576
	(−0.183)	(−0.107)	(−0.169)
DIOL	0.712	0.501	0.720
	(−0.008)	(0.043)	(−0.085)
FREIES	0.787	−0.414	0.860
	(0.764)	(−0.393)	(0.667)

The reason is evident in Table 4, where the value of ρ is quoted for a subset of the 112 trials calculated by *SIR*88 for APAPA when **h** in (27) varies over NLAR reflections. In most cases ρ is negative, even for the correct solution (trial 100). The correct solution cannot be found by seeking the maximum values of ρ.

The correlation coefficient may become a robust FOM if the summations in (27) are extended to NLAR+NEXP+NSMALL reflections. In Table 5, ρ^C, ρ^I and ρ^P are the values of ρ corresponding to the correct solution, to the incorrect solution with the highest value of ρ and to the published phase respectively. The corresponding values of ρ calculated *via* the NLAR reflections only are given in parentheses. We note that ρ is an unreliable FOM when calculated from the NLAR reflections only but is reliable when calculated from NLAR+NEXP+NSMALL reflections (the correct solution is always picked up).

7. Concluding remarks

Figures of merit based on the distribution of the α parameter and those involving psi-zero triplets have been reconsidered. It has been shown that the same mathematical formalism can be used with both these families of FOMs and leads to more efficient and robust figures of merit. An early FOM, the $R_{\rm Karle}$ criterion, has been revisited: again the combined use of reflections with large and small moduli of E is suggested and recipes for their efficient use are given. The correlation coefficient has also been taken into consideration as a possible FOM. Provided large and small intensity reflections are used simultaneously this FOM proved to be a powerful tool for finding the correct solution among numerous trials.

The authors thank Miss C. Chiarella for technical support.

References

Altomare, A., Cascarano, G., Giacovazzo, C. & Viterbo, D. (1991). *Acta Cryst.* A**47**, 744-748.

Arora, S. A., Bates, R. B., Grady, R. A., Germain, G., Declercq, J. P. & Powell, R. G. (1976). *J. Org. Chem.* **41**, 551-554.

Baggio, R., Woolfson, M. M., Declercq, J. P. & Germain, G. (1978). *Acta Cryst.* A**34**, 883-892.

Barnett, B. L. & Uchtman, V. A. (1979). *Inorg. Chem.* **18**, 2674-2678.

Braekman, J. C., Daloze, D., Dupont, A., Tursch, B., Declercq, J. P., Germain, G. & Van Meerssche, M. (1981). *Tetrahedron*, **37**, 179-186.

Burla, M. C., Camalli, M., Cascarano, G., Giacovazzo, C., Polidori, G., Spagna, R. & Viterbo, D. (1989). *J. Appl. Cryst.* **22**, 389-393.

Burla, M. C., Cascarano, G., Giacovazzo, C., Nunzi, A. & Polidori, G. (1987). *Acta Cryst.* A**43**, 370-374.

Burla, M. C., Giacovazzo, C. & Polidori, G. (1989). *Acta Cryst.* A**45**, 99-104.

Cascarano, G., Dogguy-Smiri, L. & Nguyen-Huy Dung (1987). *Acta Cryst.* C**43**, 2050-2053.

Cascarano, G., Giacovazzo, C., Camalli, M., Spagna, R., Burla, M. C., Nunzi, A. & Polidori, G. (1984). *Acta Cryst.* A**40**, 278-283.

Cascarano, G., Giacovazzo, C. & Viterbo, D. (1987). *Acta Cryst.* A**43**, 22-29.

Cochran, W. & Douglas, A. S. (1957). *Proc. R. Soc. London Ser. A*, **243**, 281.

COLENS, A., DECLERCQ, J. P., GERMAIN, G., PUTZEYS, J. P. & VAN MEERSSCHE, M. (1974). *Cryst. Struct. Commun.* **3**, 119-122.
DECLERCQ, J. P., GERMAIN, G. & WOOLFSON, M. M. (1975). *Acta Cryst.* **A31**, 367-372.
DECLERCQ, J. P., GERMAIN, G. & WOOLFSON, M. M. (1979). *Acta Cryst.* **A35**, 622-626.
DE TITTA, G. T., EDMONDS, J. W., LANGS, D. A. & HAUPTMAN, H. (1975). *Acta Cryst.* **A31**, 472-479.
GIACOVAZZO, C. (1976). *Acta Cryst.* **A32**, 91-99.
HOEKSTRA, A., VOS, A., BRAUN, P. B. & HORNSTRA, J. (1975). *Acta Cryst.* **B31**, 1708-1715.
HOVESTREYDT, E., KLEPP, K. & PARTHÉ, L. (1983). *Acta Cryst.* **C39**, 422-425.
ITO & NOVACKI (1974). *Z. Kristallogr.* **139**, 85-102.
JONES, P. G., SHELDRICK, G. M., GLÜSENKAMP, K.-H. & TIETZE, L. F. (1980). *Acta Cryst.* **B36**, 481-483.
KARLE, I. L. & KARLE, J. (1981). *Proc. Natl Acad. Sci. USA*, **78**, 681-685.
KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* **21**, 849-859.
MAIN, P. (1977). *Acta Cryst.* **A33**, 750-757.
NGUYEN-HUY DUNG, VO-VAN TIEN, BEHM, H. J. & BEURSKENS, P. T. (1987). *Acta Cryst.* **C43**, 2258-2260.
OVERBEEK, A. R. & SCHENK, H. (1976). *Proc. K. Ned. Akad. Wet. Ser. B*, **79**, 341-343.
SCHENK, H. (1974). *Acta Cryst.* **A30**, 477-481.
SHELDRICK, G. M., DAVISON, B. E. & TROTTER, J. (1978). *Acta Cryst.* **B34**, 1387-1389.
SHELDRICK, G. M. & TROTTER, J. (1978). *Acta Cryst.* **B34**, 3122-3124.
SUCK, D., MANOR, P. C. & SAENGER, W. (1976). *Acta Cryst.* **B32**, 1727-1737.
WALLWORK, S. C. & POWELL, H. M. (1980). *J. Chem. Soc. Perkin Trans.* 2, 641-646.
WHITE, P. & WOOLFSON, M. M. (1975). *Acta Cryst.* **A31**, 53-56.
YAO, J. (1981). *Acta Cryst.* **A37**, 642-644.

Acta Cryst. (1992). **A48**, 865-872

Modeling the Diffraction Process of Molecular Crystals: Computation of X-ray Scattering Intensities from *Ab Initio* Electron Densities

BY H. BRUNING AND D. FEIL

Chemical Physics Laboratory, University of Twente, POB 217, 7500 AE Enschede, The Netherlands

(*Received* 1 *October* 1991; *accepted* 15 *April* 1992)

Abstract

An algorithm for calculating the scattering factors of atomic fragments in molecules as defined by the Stockholder recipe is presented. This method allows the calculation, from *ab initio* molecular wave functions, of structure factors including individual anisotropic atomic temperature factors. These structure factors agree with the model used in most least-squares multipole-refinement procedures. Calculations on the H_2O molecule illustrate the method.

1. Introduction

X-ray scattering experiments can provide us with a large amount of information on the structures of molecular crystals. Accurate high-resolution experiments can even reveal details of the electron-density distribution, such as bonding densities and subtle effects of intermolecular interactions and polarization by the crystal field (Krijn, Graafsma & Feil, 1988; Krijn & Feil, 1988). This fact makes it relevant to calculate in advance the results of an X-ray scattering experiment by quantum-chemical *ab initio* methods. Firstly, it gives us the possibility to verify experimentally the approximations used in *ab initio* calculations of the electron-density distribution in a crystal. Is it necessary to use the Bloch-function approach in crystal calculations or do cluster calculations suffice? When a cluster approach is deemed suitable one can test the basis-set truncation error and the approximations that have to be used to incorporate the embedding of the cluster in the crystal, *i.e.* the polarization by the electrostatic crystal field and the effect of the exchange repulsion by the surrounding molecules. This is particularly important when hydrogen bonds and electrostatic fields are included in the calculation on molecular crystals. Secondly, the theoretically calculated X-ray intensities can be used to check the crystallographic refinement procedures that are applied to remove noise from the data and to obtain information on the electron-density distribution in analytic form. In particular, one can verify whether the structural data on which the theoretical calculations are based are reproduced by the refinement.

Most experimental X-ray diffraction data on crystals are interpreted with a model based on the assumption that the crystal is built up of atoms. The lattice vibrations of the crystal, which consist of zero-point vibrations and thermal excitations, are taken into account by the Debye-Waller factor. Widely used expressions of the Debye-Waller factor are based on the assumption that the atoms behave as coupled harmonic oscillators. This harmonic-vibration model implies that the density distribution for each nucleus is given by a three-dimensional Gaussian distribution. The adiabatic approximation leads to the model of

0108-7673/92/060865-08$06.00