



Fig. 1. Evolution of the normalized Cochran integral  $Z(\Phi)/Z_{\text{exp}}$  and of the  $\chi^2/200$  goodness of fit during the refinement of the 190 initially random phases of the largest normalized structure factors using (a) the conventional tangent formula and (b) the  $\chi^2$  tangent formula. To facilitate the comparison between the refinements, the same starting phase values have been used. While the conventional tangent formula maximizes  $Z(\Phi)$  to its global maximum (a), the addition of the  $\chi^2$  restraint succeeds in keeping  $Z(\Phi)$  at the correct local maximum (b).

is also reflected in the high  $\chi^2$  values computed at the end of each iteration.

To avoid this behaviour, 200 medium-large  $E$ 's in the range 1.70 to 1.41 have been considered in the  $\chi^2$  sum. This represents the addition of 3271 new triplets to the 3706 involving only the largest  $E$ 's. 40 sets of initially random phases were refined using the  $\chi^2$  tangent formula with  $\lambda = 0.095$ , recalculating  $\phi_{\text{H}}$  and  $G_{\text{H}}$  every 38 refined phases. From the 40 sets, two showed the images of the three symmetry-independent molecules. The evolution of  $Z(\Phi)$  and  $\chi^2$  for one of these two sets is represented in Fig. 1(b). Comparison with Fig. 1(a) shows that  $Z(\Phi)/Z_{\text{exp}}$  is here closer to 1 and that  $\chi^2/200$  gradually decreases from 4.86 at the first iteration to 0.68 at the end of the refinement. This clearly indicates that the addition of the  $\chi^2$  restraint hinders  $Z(\Phi)$  from reaching the global maximum.

Although derived in space group  $P1$ , the  $\chi^2$  tangent formula is completely general and can be applied to all space groups with only minor modifications.

This work has been supported by the DGICYT (Project PB89-0036).

#### References

- COCHRAN, W. (1952). *Acta Cryst.* **5**, 65–67.  
 CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.  
 DEBAERDEMAEKER, T., TATE, C. & WOOLFSON, M. M. (1985). *Acta Cryst.* **A41**, 286–290.  
 DEBAERDEMAEKER, T., TATE, C. & WOOLFSON, M. M. (1988). *Acta Cryst.* **A44**, 353–357.  
 GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1970). *Acta Cryst.* **B26**, 274–285.  
 HAUPTMAN, H. (1972). *Z. Kristallogr.* **135**, 1–17.  
 KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635–651.  
 POYSER, J. P., EDWARDS, R., ANDERSON, J., HURSTHOUSE, M., WALKER, N., SHELDRIK, G. & WALLEY, A. (1986). *J. Antibiot.* **39**, 167–169.  
 RIUS, J. & MIRAVITLLES, C. (1992). *Acta Cryst.* **A48**, 69–70.  
 WOOLFSON, M. M. & YAO, J. (1990). *Acta Cryst.* **A46**, 409–413.

*Acta Cryst.* (1992). **A48**, 30–36

## Integration of Patterson Information into Direct Methods. II. The First Applications

BY A. ALTOMARE, G. CASCARANO AND C. GIACOVAZZO

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

(Received 6 February 1991; accepted 21 June 1991)

#### Abstract

A probabilistic formula [Giacovazzo (1991). *Acta Cryst.* **A47**, 256–263] estimates triplet invariant phases

given prior information on a non-Harker Patterson peak  $\mathbf{u}$ . The formula requires prior information both on the coordinates of the peak and on the scattering factors of the atoms with mutual distance  $\mathbf{u}$ . Since

this second type of information is usually not available, the formula has been modified so that it depends on  $\mathbf{u}$  and on the Patterson peak intensity in  $\mathbf{u}$ . Some applications are described. The use of some algebraic relationships among structure factors arising from the positivity and atomicity of the electron density function are also considered.

### 1. Symbols and abbreviations

From now on the paper by Giacovazzo (1991) is referred to as paper I. Symbols and abbreviations coincide with those used in paper I: for the sake of brevity they are not listed here.

### 2. Introduction

In paper I it was shown that a Patterson map may be used as prior information in probabilistic approaches for the estimation of triplet invariants. Two types of formula were derived. The first one, of algebraic nature, depends on the positivity and atomicity of the electron density function. Specifically, Cochran's (1952) relationship

$$S = \int_V \rho^3(\mathbf{r}) \, d\mathbf{r} \\ = \sum'_{\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3=0} |F_{\mathbf{h}_1} F_{\mathbf{h}_2} F_{\mathbf{h}_3}| \cos \Phi = \max \quad (1)$$

was generalized into

$$S' = \int_V \rho^3(\mathbf{r}) \, d\mathbf{r} + a \left[ \sum_{s=1}^m \int_V \rho(\mathbf{r} + \mathbf{R}_s \mathbf{u}) \rho^2(\mathbf{r}) \, d\mathbf{r} \right. \\ \left. + \sum_{s=1}^m \int_V \rho(\mathbf{r} - \mathbf{R}_s \mathbf{u}) \rho^2(\mathbf{r}) \, d\mathbf{r} \right] \\ = \sum'_{\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3=0} |F_{\mathbf{h}_1} F_{\mathbf{h}_2} F_{\mathbf{h}_3}| \\ \times \cos \Phi \left\{ 3 + 2a \sum_{s=1}^m \cos 2\pi \mathbf{h}_1 \mathbf{R}_s \mathbf{u} \right. \\ \left. + \cos 2\pi \mathbf{h}_2 \mathbf{R}_s \mathbf{u} + \cos 2\pi \mathbf{h}_3 \mathbf{R}_s \mathbf{u} \right\} \\ = \max \quad (2)$$

where  $a$  is a numerical factor which is supposed to be close to unity and  $\mathbf{u}$  is a non-Harker Patterson peak. The application of (1) and (2) in direct phasing procedures requires some caution: remarks concerning this are made in § 3.

The second type of formula derived in paper I integrates the simplest type of Patterson information, the knowledge of a non-Harker peak, into the method of joint probability distribution functions of structure factors (Hauptman & Karle, 1953). The theoretical results obtained by Giacovazzo may be listed:

(a) In the absence of any Patterson information the conditional probability of the triplet phase

invariant  $\Phi$  given the moduli  $R'_{\mathbf{h}_1}, R'_{\mathbf{h}_2}, R'_{\mathbf{h}_3}$  is given (Cochran, 1952) by

$$P(\Phi | R'_{\mathbf{h}_1}, R'_{\mathbf{h}_2}, R'_{\mathbf{h}_3}) \\ \simeq [2\pi I_0(G')]^{-1} \exp(G' \cos \Phi) \quad (3)$$

where  $G' = 2|R'_{\mathbf{h}_1} R'_{\mathbf{h}_2} R'_{\mathbf{h}_3}| \sigma_3 \sigma_2^{-3/2}$ .

(b) If a Patterson peak  $\mathbf{u}$  is known, then

$$P(\Phi | R_{\mathbf{h}_1}, R_{\mathbf{h}_2}, R_{\mathbf{h}_3}, \mathbf{u}) \\ \simeq [2\pi I_0(G)]^{-1} \exp[G \cos(\Phi - \theta)], \quad (4)$$

where  $G$  and  $\theta$  are quantities determined in paper I.

(c) When several symmetry-independent interatomic peaks coincide with  $\mathbf{u}$  it is difficult to obtain information about  $\theta$ . In this case the best procedure is not to try to define the enantiomorph. Then

$$P(\Phi | R_{\mathbf{h}_1}, R_{\mathbf{h}_2}, R_{\mathbf{h}_3}, \mathbf{u}) \\ \simeq [2\pi I_0(G)]^{-1} \exp(G \cos \Phi) \quad (5)$$

where

$$G = 2R_{\mathbf{h}_1} R_{\mathbf{h}_2} R_{\mathbf{h}_3} M [ \langle |F_{\mathbf{h}_1}|^2 | \mathbf{u} \rangle \langle |F_{\mathbf{h}_2}|^2 | \mathbf{u} \rangle \langle |F_{\mathbf{h}_3}|^2 | \mathbf{u} \rangle ]^{-1/2} \quad (6)$$

$$M = \sum_3 (\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3) + \sum_{i=1}^3 Q_i \alpha_i$$

$$\alpha_i = \sum_{s=1}^m \cos 2\pi \mathbf{h}_i \mathbf{R}_s \mathbf{u},$$

$$Q_1 = \sum'_{(p,q)} [f_q(\mathbf{h}_1) f_p(\mathbf{h}_2) f_p(\mathbf{h}_3) + f_p(\mathbf{h}_1) f_q(\mathbf{h}_2) f_q(\mathbf{h}_3)]$$

$$Q_2 = \sum'_{(p,q)} [f_p(\mathbf{h}_1) f_q(\mathbf{h}_2) f_p(\mathbf{h}_3) + f_q(\mathbf{h}_1) f_p(\mathbf{h}_2) f_q(\mathbf{h}_3)]$$

$$Q_3 = \sum'_{(p,q)} [f_q(\mathbf{h}_1) f_q(\mathbf{h}_2) f_p(\mathbf{h}_3) + f_p(\mathbf{h}_1) f_p(\mathbf{h}_2) f_q(\mathbf{h}_3)]$$

$$\langle |F_{\mathbf{h}_i}|^2 | \mathbf{u} \rangle = \varepsilon_{\mathbf{h}_i} \left\{ \sum_N (\mathbf{h}_i) + 2\alpha_i \left[ \sum'_{(p,q)} f_p(\mathbf{h}_i) f_q(\mathbf{h}_i) \right] \right\}$$

$$R_{\mathbf{h}} = |F_{\mathbf{h}}| / \langle |F_{\mathbf{h}}|^2 | \mathbf{u} \rangle^{1/2}.$$

The primed summation goes over the pairs of symmetry-independent atoms that correspond to  $\mathbf{u}$ .

The quantity  $G$  cannot be directly obtained from inspection of the Patterson map. In § 4 we will show that  $G$  may be estimated in terms of Patterson peak intensities. In § 5 some applications are described and in § 6 the method is generalized to the case in which prior information on a set  $\{\mathbf{u}\}$  is available.

### 3. About the use of some algebraic relationships in direct phasing procedures

Condition (1) states that for the correct structure the integral of the cube of the electron density would be a maximum. Stanley (1979, 1986) showed that maximization of such an integral could be used both for the determination of an initial set of signs and

for the refinement of a structure. Hoppe (1963), Main & Woolfson (1963) and Allegra (1979) described algebraic methods for phase estimation which make use of the zero points in the Patterson map [*i.e.* those points for which  $P(\mathbf{u})$  is zero or expected to be zero]. Rius & Miravittles (1989) suggested that information about zero points would maximize the difference

$$\int \rho^3(\mathbf{r}) \, d\mathbf{r} - a \int \rho(\mathbf{r} + \mathbf{u})\rho^2(\mathbf{r}) \, d\mathbf{r}. \quad (7)$$

In paper I it was remarked that the above-mentioned algebraic relationships had no rigorous probabilistic background: thus their active use in the phasing process would sometimes be ineffective. We want to provide here simple tests of this statement.

The simplest way to check the efficiency of Cochran's relationship (1) is to calculate the value of  $S$  for the various trials in a multisolution approach. All the triplets which can be found among the NRIF reflections with the largest values of  $R$  contribute to  $S$ . We have used seven test structures, the code names of which, together with relevant crystallochemical data, are shown in Table 1. Default runs of *SIR88* (Burla, Camalli, Cascarano, Giacovazzo, Polidori, Spagna & Viterbo, 1989) provided data shown in Table 2, where for each structure the NRIF values, the maximum (say  $S_{\max}$ ), the minimum (say  $S_{\min}$ ) and the  $S$  values corresponding to the published phases (say  $S_{\text{cor}}$ ) are shown. It may be seen that  $S_{\text{cor}}$  is a maximum only for LOGANIN, in most cases  $S_{\text{cor}}$  is intermediate and in two cases (GOLD and SICA) it is a minimum. SICA is a striking case where the correct phases correspond to a deep minimum.

According to these results, direct-methods programs based on the maximization of  $S$  could have some difficulties in solving complex structures. This result is not completely unexpected. Indeed,  $S$  is strongly correlated (see Table 3) with the figure of merit MABS

$$\text{MABS} = \sum_h \alpha_h / \sum_h \langle \alpha_h \rangle$$

and MABS is a rather ineffective figure of merit. In *SIR88* the weight associated with MABS is 0.2 the weight associated with the so-called psi-zero triplets (Cascarano, Giacovazzo & Viterbo, 1987).

In order to define the possible role of  $S'$  in direct procedures we introduced into (2) a number of  $\mathbf{u}'$ s corresponding to real interatomic vectors and calculated the values of  $S'$  from the published phases ( $a$  is assumed to be unity). Results are shown in Table 4: the maximum and the minimum values of  $S'$  obtained from the trial solutions are also quoted in parentheses. It is easily seen that  $S'$  and  $S$  have a similar behaviour:  $S'$  is maximum for LOGANIN, minimum for GOLD and SICA, intermediate in other cases.

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

Structure code*	Space group	Molecular formula	Z
LOGANIN	$P2_12_12_1$	$C_{17}H_{26}O_{10}$	4
GOLD	$Cc$	$C_{28}H_{16}$	8
CEPHAL	$C2$	$C_{18}H_{21}NO_3$	8
AZET	$Pca2_1$	$C_{21}H_{16}ClNO$	8
MUNICH	$C2$	$C_{20}N_{16}$	8
TUR10	$P6_322$	$C_{15}H_{24}O_2$	12
SICA†	$P2_1$	$C_{20}H_{30}N_4O_8$	4

\* The complete references for such structures are not given for the sake of brevity. The reader is referred to a magnetic tape distributed by the crystallographic group in Göttingen.

† Capasso, Mazzarella, Sica, Zagari, Cascarano & Giacovazzo (1991).

Table 2. *Values of  $S_{\max}$ ,  $S_{\min}$  and  $S_{\text{cor}}$  for the seven test structures*

	NRIF	$S_{\max}$	$S_{\min}$	$S_{\text{cor}}$
LOGANIN	258	14552	9311	15627
GOLD	374	17541	16519	14634
CEPHAL	334	18945	10500	13393
AZET	342	19180	12590	14548
MUNICH	310	19504	12031	13239
TUR10	219	16161	11061	14346
SICA	406	42693	29344	13757

Table 3. *Values of MABS and  $S$  for trial solutions provided by *SIR88**

The last one corresponds to the correct solution.

	MABS	$S$
CEPHAL	1.595	18945
	0.939	11128
	0.973	11537
	0.885	10500
	1.595	18945
	1.233	14639
	1.053	12500
	1.078	12797
	1.055	12530
	1.082	13393

Table 4. *Values of  $S'$ ,  $S'_p$  and  $S''$  for the test structures*

The maximum and the minimum values of  $S'$ ,  $S'_p$  and  $S''$  obtained for the trial solutions are quoted in parentheses.

	$S'$	$S'_p$	$S''$
LOGANIN	66 177	-1990	48 870
	(60 066, 38 218)	(3843, -18578)	(53 988, 25 321)
GOLD	60 215	-7173	51 073
	(71 365, 67 997)	(797, -9819)	(61 975, 48 760)
CEPHAL	53 804	-14 492	54 673
	(80 513, 44 820)	(-7084, -19 569)	(76 407, 40 468)
AZET	183 123	20 892	22 754
	(259 229, 160 937)	(30 827, 21 432)	(29 248, 13 740)
MUNICH	54 890	-72	39 787
	(82 024, 50 537)	(5908, -2881)	(55 918, 33 013)
TUR10	69 782	-1595	44 632
	(83 974, 69 782)	(-4106, -33 141)	(81 626, 37 290)
SICA	172 172	-2817	44 091
	(518 971, 385 064)	(-1130, -19 518)	(147 513, 90 901)

Instead of (7) we calculated the (Patterson) symmetry-invariant quantity

$$S'' = \int \rho^3(\mathbf{r}) \, d\mathbf{r} - a \left[ \sum_s \int_V \rho(\mathbf{r} + \mathbf{R}_s \mathbf{u}) \rho^2(\mathbf{r}) + \sum_{s=1}^m \int_V \rho(\mathbf{r} - \mathbf{R}_s \mathbf{u}) \rho^2(\mathbf{r}) \right] \\ = S - S''_p. \quad (8)$$

where  $\mathbf{u}$  is a zero point of the Patterson. We chose, according to Rius & Miravittles (1989), a positional vector  $\mathbf{u}$  with length about 0.7 Å. The values of  $S''$  are shown in the last column of Table 4.  $S''$  is never a maximum and for SICA it is a minimum. We also quote the values of  $S''_p$  in order to demonstrate that often they are positive: in AZET,  $S''_p$  is positive both for the trial solutions and for the correct structure.

A further question could be: can the use of a larger number of Patterson peaks make the values of  $S'$  and  $S''$  more significant? The expressions of  $S'$  and  $S''$  for the cases in which several  $\mathbf{u}$ 's are known are easy to obtain: single summations over  $s$  in (2) and (8) must be replaced by a double summation over  $\mathbf{u}$  and over  $s$ . We have then selected 20 atomic Patterson vectors for each structure and calculated  $S'$  and  $S''$ . Our conclusion is that, at least for limited sets of data such as those phased in a direct procedure,  $S$ ,  $S'$  and  $S''$  are not expected to be maxima for the correct structure. That does not imply that  $S$ ,  $S'$ ,  $S''$  should be global maxima for the correct structure and for sufficient data.

#### 4. A directly applicable probabilistic formula exploiting Patterson information

We want to express the parameter  $G$  defined by (4) in terms of experimentally available quantities. We divide the numerator and denominator of (4) by  $[\sum_N(\mathbf{h}_1) \sum_N(\mathbf{h}_2) \sum_N(\mathbf{h}_3)]^{1/2}$  and we use the approximation (Cochran & Woolfson, 1955; Karle & Karle, 1966)

$$\frac{\sum_3(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)}{[\sum_N(\mathbf{h}_1) \sum_N(\mathbf{h}_2) \sum_N(\mathbf{h}_3)]^{1/2}} \\ \approx \sigma_3 / \sigma_2^{3/2} \approx N^{-1/2}.$$

Then

$$G = [2R_{\mathbf{h}_1} R_{\mathbf{h}_2} R_{\mathbf{h}_3} / (q_{\mathbf{h}_1} q_{\mathbf{h}_2} q_{\mathbf{h}_3})^{1/2}] \left\{ N^{-1/2} + [\sum_N(\mathbf{h}_1) \sum_N(\mathbf{h}_2) \sum_N(\mathbf{h}_3)]^{-1/2} \sum_{i=1}^3 Q_i \alpha_i \right\},$$

where

$$q_i = \varepsilon_{\mathbf{h}_i} \left\{ 1 + 2 \left[ \sum'_{(p,q)} f_p(\mathbf{h}_i) f_q(\mathbf{h}_i) \right] \alpha_i / \sum_N(\mathbf{h}_i) \right\}. \quad (9)$$

We introduce in (9) the approximation

$$\sum'_{(p,q)} f_p(\mathbf{h}_i) f_q(\mathbf{h}_i) / \sum_N(\mathbf{h}_i) \approx \sum'_{(p,q)} Z_p Z_q / \sum_{p=1}^N Z_p^2$$

where  $Z_p$  is the atomic number of the  $p$ th atom. Then

$$q_i = \varepsilon_{\mathbf{h}_i} \{ 1 + (2I_u / j_u I_0) \alpha_i \} = \varepsilon_{\mathbf{h}_i} \{ 1 + 2I'_u \alpha_i \} \quad (10)$$

where  $j_u$  is the ratio between  $m$  and peak multiplicity,  $I_0$  and  $I_u$  are the Patterson peak heights at the origin and at  $\mathbf{u}$  respectively.

By analogy,

$$Q_i \alpha_i / [\sum_N(\mathbf{h}_1) \sum_N(\mathbf{h}_2) \sum_N(\mathbf{h}_3)]^{1/2} \\ \approx \left[ \sum'_{(p,q)} (Z_p^2 Z_q + Z_p Z_q^2) \right] \alpha_i / \sigma_2^{3/2} \\ \approx (\sigma_3 / \sigma_2^{3/2}) \left[ \sum'_{(p,q)} Z_p Z_q (Z_p + Z_q) \right] \alpha_i / \sigma_3 \\ \approx 2N^{-1/2} \left[ \sum'_{(p,q)} Z_p Z_q \right] \alpha_i / \sigma_2 \\ \approx 2N^{-1/2} I'_u \alpha_i.$$

Then (6) may be written as

$$G \approx 2R_{\mathbf{h}_1} R_{\mathbf{h}_2} R_{\mathbf{h}_3} N^{-1/2} \left[ 1 + 2I'_u \left( \sum_{i=1}^3 \alpha_i \right) \right] \\ \times \left[ \prod_{i=1}^3 \varepsilon_{\mathbf{h}_i} (1 + 2I'_u \alpha_i) \right]^{-1/2} \quad (11)$$

where  $R_{\mathbf{h}}$  is given by (14) (see later).

It may be useful to express  $G$  in terms of normalized structure factors  $R'$  in order to see how prior information on  $\mathbf{u}$  modifies the triplet invariant reliability. We obtain

$$G = 2R'_{\mathbf{h}_1} R'_{\mathbf{h}_2} R'_{\mathbf{h}_3} N^{-1/2} C' \quad (12)$$

where

$$C' = \left[ 1 + 2I'_u \left( \sum_{i=1}^3 \alpha_i \right) \right] / \prod_{i=1}^3 \varepsilon_{\mathbf{h}_i} (1 + 2I'_u \alpha_i). \quad (13)$$

Finally, it is seen that prior Patterson information is contained in the modulating term  $C'$ ; triplet reliability is enhanced if  $C' > 1$ , reduced if  $C' < 1$ .

#### 5. First applications

$R_{\mathbf{h}}$  is referred to  $R'_h$  by the relation

$$R_{\mathbf{h}} = |F_{\mathbf{h}}| / \langle |F_{\mathbf{h}}|^2 | \mathbf{u} \rangle^{1/2} \\ = (|F_{\mathbf{h}}| / \langle |F_{\mathbf{h}}|^2 \rangle^{1/2}) (\langle |F_{\mathbf{h}}|^2 \rangle^{1/2} / \langle |F_{\mathbf{h}}|^2 | \mathbf{u} \rangle^{1/2}) \\ = R'_h [1 + 2I'_u \alpha_{\mathbf{h}}]^{1/2}. \quad (14)$$

Accordingly, in our phasing method the usual normalization procedure is run first. Then:

(a) the renormalized structure factors are obtained by application of (14);

Table 5. *SICA: section of the SIR88 output devoted to the statistical analysis of the reflection intensities*

M.f.s.p. stands for mean fractional scattering power of the electron density obeying pseudotranslational symmetry.

Classes of reflections probably affected by pseudotranslational effects:

Condition	Number of reflections	$\langle  E ^2 \rangle$	Figure of merit	(M.f.s.p.)
$2h + k + 2l = 4n$	2058	1.497	1.80	16%

Remarkable deviations (of displacive type) from ideal pseudotranslational symmetry are present:

at $(\sin \theta / \lambda)^2 = 0$	m.f.s.p. = 35%
at $(\sin \theta / \lambda)^2 = \max$	m.f.s.p. = 1%
	(m.f.s.p.) = 16%

(b) the largest  $R$  values are selected for active use in the phasing process (such a set does not in general coincide with the set of the largest  $R'$ );

(c) triplet reliability is estimated by means of (11): for the sake of simplicity the  $\varepsilon$  values are all assumed to be unitary;

(d) the usual convergence-divergence and tangent formula procedures are applied.

In order to obtain further insight into the method we describe its application to SICA. Repeated attempts at solving such a structure *via* the most powerful direct-methods packages (*SIR88* included) failed. Statistical analysis of the reflection intensity distribution by *SIR88* suggested the presence of a pseudotranslational vector  $\mathbf{u} = a/2 + b/4 + c/2$ . The calculated mean fractional scattering power of the electron density suffering from such a pseudotranslational symmetry was 16%: remarkable deviations of displacive type from ideal pseudosymmetry were also suggested (see Table 5). All subsequent trials by *SIR88* aiming at solving the structure by exploiting pseudotranslational symmetry as prior information (Casarano, Giacobozzo & Luić, 1988) were unsuccessful. An  $|E|^2$  Patterson map was then calculated which showed the following three largest peaks:

$u$	$v$	$w$	Intensity
0.00	0.00	0.00	1000
0.50	0.27	0.50	429
0.10	0.00	0.02	106

whose positions and intensities confirm the *SIR88* statistical analysis. The peak at (0.50, 0.27, 0.50) was assumed as prior information in our probabilistic approach ( $I'_u = 0.21$ ) and immediately provided a quasicomplete crystal structure. The combined figure of merit was 0.82 (the expected value for a correct structure is unity), much higher than the best CFOM values ( $\sim 0.20$ ) relative to usual trials by *SIR88*.

The role of the Patterson information may be better evaluated by a post-mortem analysis of the phasing process. If triplet invariants are arranged in increasing order of reliability according to Cochran's parameter  $G'$  then Table 6 follows. For each interval ( $G'_i - G'_{i+1}$ )

Table 6. *SICA: triplets arranged according to Cochran's reliability parameter  $G'$* 

$G'_i - G'_{i+1}$	$nr$	%	$\langle  \Phi  \rangle$ (°)
<0.8	0		
0.8-1.2	1887	63.1	76
1.2-1.6	3041	62.9	75
1.6-2.0	1543	65.2	72
2.0-2.4	789	67.0	69
2.4-3.0	497	70.2	67
3.0-3.6	157	66.9	68
3.6-4.2	48	79.2	51
4.2-4.8	22	86.4	44
4.8-5.5	14	85.7	53
5.5-6.5	1	100.0	55
6.5-9.0	1	0.0	173

Table 7. *CEPHAL: triplets arranged according to Cochran's reliability parameter  $G'$* 

$G'_i - G'_{i+1}$	$nr$	%	$\langle  \Phi  \rangle$ (°)
<0.8	152	78.3	59
0.8-1.2	1826	80.3	54
1.2-1.6	1149	87.5	45
1.6-2.0	431	91.5	42
2.0-2.4	122	97.5	34
2.4-3.0	58	94.8	28
3.0-3.6	10	100.0	14
3.6-4.2	3	100.0	32

the number of triplets ( $nr$ ) with  $G'_i < G' < G'_{i+1}$  is given, % ( $\times 100$ ) is the percentage of triplets having positive cosine sign and  $\langle |\Phi| \rangle$  is the average absolute deviation of  $\Phi$  from  $2\pi$  (the  $\Phi$ 's are calculated from the refined structure) in degrees. Table 6 clearly shows that SICA is a structure of unusual difficulty: the triplet reliability is largely overestimated, the percentage of negative cosine triplets and the  $\langle |\Phi| \rangle$  value are abnormally high for each interval of  $G'$ . The reader can usefully compare Table 6 with Table 7 where analogous data are shown for CEPHAL, a structure generally difficult to solve.

The renormalization process according to (14) radially changes the set of active reflections. For SICA, where  $\mathbf{u} = (0.50, 0.27, 0.50)$ ,

$$1 + 2I'_u \alpha_h \approx 1 + 4I'_u \cos[\pi(h+1)] \cos(\pi k/2).$$

The largest effects of the Patterson information occur for:

(a) reflections satisfying

$$\begin{cases} h+l=2n \\ k=4n \end{cases} \quad \text{or} \quad \begin{cases} h+l=2n+1 \\ k=2n(\neq 4n) \end{cases};$$

on assuming  $I'_u = 0.21$  then  $R_h \approx R'_h/1.36$ ;

(b) reflections for which

$$\begin{cases} h+l=2n \\ k=2n(\neq 4n) \end{cases} \quad \text{or} \quad \begin{cases} h+l=2n+1 \\ k=4n \end{cases};$$

on assuming  $I'_u = 0.21$  then  $R_u \approx R'_u/0.4$ .

It seems from the above examples that the estimation of  $I'_u$  is critical for the success of the method. Unfortunately, several values of  $I'_u$  are in practice available according to the coefficients used in the

Table 8. *SICA: largest peaks in  $|F|^2$  and  $|EF|$  Patterson syntheses*

$ F ^2$ Patterson				$ EF $ Patterson			
$u$	$v$	$w$	Intensity	$u$	$v$	$w$	Intensity
0.00	0.00	0.00	1000	0.00	0.00	0.00	1000
0.50	0.27	0.50	570	0.50	0.27	0.50	487
0.00	0.50	0.00	221	0.00	0.50	0.00	112

Table 9. *SICA: triplets arranged according to the parameter  $G$* 

$G_i - G_{i+1}$	Positive estimated triplets			Negative estimated triplets		
	$nr$	%	$\langle \Phi \rangle$ (°)	$nr$	%	$\langle \Phi \rangle$ (°)
<0.8	90	57.8	83	75	46.7	86
0.8-1.2	503	56.5	85	36	44.4	85
1.2-1.6	2704	62.2	76			
1.6-2.0	1533	66.2	71			
2.0-2.4	516	71.1	66			
2.4-3.0	167	66.5	68			
3.0-3.6	67	76.1	58			
3.6-4.2	19	73.7	68			
4.2-4.8	4	100.0	53			
4.8-5.5	1	0.0	173			

Patterson function. In Table 8 the positions of the three largest peaks in the  $|F|^2$  synthesis (temperature factor not removed) and  $|EF|$  Patterson synthesis are shown. For the first map,  $I'_u = 0.28$  and, for the second,  $I'_u = 0.24$ . From a numerical point of view, using  $I'_u = 0.28$  instead of  $I'_u = 0.21$  or  $I'_u = 0.24$  can sometimes provide quite different values of  $R$ . Indeed, (14) can give rise to abnormally large values of  $R$  when  $2I'_u\alpha_h$  is close to  $-1$ . Furthermore, owing to rounding errors in  $I'_u$  or  $\alpha_h$ , the quantity  $1 + 2I'_u\alpha_h$  could occasionally become negative, against any mathematical or physical expectation. In order to make less critical the estimation of  $I'_u$  we limited the values of  $2I'_u\alpha_h$  according to the following scheme:

$$\text{if } 2I'_u\alpha_h \leq -0.5 \text{ then } 2I'_u\alpha_h = -0.5. \quad (15)$$

Condition (15) is also introduced in (14).

In order to give a qualitative estimate of the influence of the Patterson information on the triplet reliability parameter we calculate here the parameter  $C$  in three typical situations: (1)  $I'_u = 0.21$ ,  $\alpha_1 = \alpha_2 = \alpha_3 \approx 2$ , then  $C \approx 0.56$ ; (2)  $I'_u = 0.21$ ,  $\alpha_1 \approx 2$ ,  $\alpha_2 = \alpha_3 = -\alpha_1$ , then  $C = 1.83$ ; (3)  $I'_u = 0.21$ ,  $\alpha_1 = \alpha_2 = \alpha_3 \approx -2$ , then  $C = -4$ .

The above examples show that  $C$  can be considerably different from unity. In Table 9 triplet invariants (found among the largest  $R$  values) are ranked according to  $G$ . It may be observed:

(1) the quality of triplets selected by the new approach is not substantially better than that of triplets ranked in Table 6. However,  $G$  is a more realistic parameter than  $G'$ .

(2) Some negative triplets have been identified: their  $\langle|\Phi|\rangle$  is close to  $\pi/2$ . Since they are excluded from any active use, the phasing procedure benefits by their omission.

## 6. The estimation of $\Phi$ given more Patterson peaks

In paper I it was shown how Cochran's estimates of triplet invariants are modified by prior information on one Patterson peak. The conclusive formulas (I.18)–(I.20) [and therefore also our formula (11)] hold even when  $\mathbf{u}$  is generated by overlapping of more symmetry-independent interatomic vectors

$$\mathbf{r}_{j_1} - \mathbf{r}_{j_2} = \mathbf{r}_{j_3} - \mathbf{r}_{j_4} = \dots = \mathbf{u}$$

provided  $j_1 \neq j_2 \neq j_3 \neq j_4 \neq \dots$ .

A problem now arises: how triplet invariants can be estimated when prior information on a set  $\{\mathbf{u}\}$  of Patterson peaks is available. If the set of atomic positions  $\{\mathbf{r}_j\}_\mu$  giving rise to the Patterson vector  $\mathbf{u}_\mu$  is disjoint from the set  $\{\mathbf{r}_j\}_\nu$  giving rise to the Patterson vector  $\mathbf{u}_\nu$ , then (I.16) may be generalized to

$$\begin{aligned} \langle|F_{\mathbf{h}}|^2|\{\mathbf{u}\}\rangle &= \varepsilon_{\mathbf{h}} \left( \sum_N (\mathbf{h}) + 2 \sum_{\mathbf{u}} \left\{ \sum_{s=1}^m \cos 2\pi \mathbf{h} \mathbf{R}_s \mathbf{u} \right. \right. \\ &\quad \left. \left. \times \left[ \sum_{p,q} f_p(\mathbf{h}) f_q(\mathbf{h}) \right] \right\} \right). \end{aligned}$$

By analogy, (I.18) still holds provided

$$G = M [\langle|F_{\mathbf{h}_1}|^2|\{\mathbf{u}\}\rangle \langle|F_{\mathbf{h}_2}|^2|\{\mathbf{u}\}\rangle \langle|F_{\mathbf{h}_3}|^2|\{\mathbf{u}\}\rangle]^{1/2} 2R_{\mathbf{h}_1} R_{\mathbf{h}_2} R_{\mathbf{h}_3}$$

$$\begin{aligned} M &= \sum_3 (\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3) + \sum_{\mathbf{u}} \left\{ \left( Q_1 \sum_{s=1}^m \cos 2\pi \mathbf{h}_1 \mathbf{R}_s \mathbf{u} \right) \right. \\ &\quad \left. + \left( Q_2 \sum_{s=1}^m \cos 2\pi \mathbf{h}_2 \mathbf{R}_s \mathbf{u} \right) \right. \\ &\quad \left. + \left( Q_3 \sum_{s=1}^m \cos 2\pi \mathbf{h}_3 \mathbf{R}_s \mathbf{u} \right) \right\}. \end{aligned}$$

Under the same hypotheses our formula (14) becomes

$$R_{\mathbf{h}} = R'_{\mathbf{h}} / [1 + \sum_{\mathbf{u}} (2I'_u \alpha_{\mathbf{h}})]^{1/2} \quad (16)$$

and (11) is transformed into

$$\begin{aligned} G &\approx 2R_{\mathbf{h}_1} R_{\mathbf{h}_2} R_{\mathbf{h}_3} N^{-1/2} \left\{ 1 + \sum_{\mathbf{u}} \left[ 2I'_u \left( \sum_{i=1}^3 \alpha_i \right) \right] \right\} \\ &\quad \times \left\{ \prod_{i=1}^3 \varepsilon_{\mathbf{h}_i} [1 + \sum_{\mathbf{u}} (2I'_u \alpha_i)] \right\}^{-1/2}. \quad (17) \end{aligned}$$

It is not easy to check (particularly for equal-atom structures) if hypotheses under which (16) and (17) hold are satisfied. Thus the application of (16) and (17) is not always justified.

## 7. Concluding remarks

Nixon (1978) recognized that overlapping vectors in the Patterson function may cause failures in direct methods. He proposed a simple removal of the overlapping vectors (which produces modified  $|E|$  values)

and successfully applied the technique to solve a structure resistant to traditional direct methods.

The probabilistic approach described in paper I aims at exploiting Patterson information, both for the renormalization of the structure factors and for modifying the phase-estimating formulae. The approach has been further developed in this paper in order to take into account experimental parameters such as position and intensity of a Patterson peak. The experimental applications described here prove that the method may be useful when the distribution of the atoms in the unit cell gives rise to a large Patterson peak originated by overlapping of several interatomic vectors.

The application of the methods to usual equal-atom structures is still questionable because supplementary information contained in one (or more) weak Patterson peak has limited influence on the triplet estimation, unless a more complete analysis of the Patterson map is made. It is in this direction that we will address our future efforts.

*Acta Cryst.* (1992). **A48**, 36–41

## Convergent-Beam RHEED Calculations of a 'Forbidden Reflection' from the Si (111) Surface

BY ANDREW E. SMITH

*Department of Physics, Monash University, Clayton, Victoria 3168, Australia*

(Received 24 August 1990; accepted 15 July 1991)

### Abstract

Results of convergent-beam reflection high-energy electron diffraction (CB-RHEED) calculations are presented for the (111) surface of silicon. These double rocking calculations are performed using a dynamical scattering approach. This is based on evaluation of the surface parallel multislice matrix for reflection with account taken of the boundary conditions. In particular, calculations are shown for intensities close to a reflection that is kinematically forbidden. Particular note is made of the computational simulation of intensity enhancements corresponding to surface wave resonance conditions.

### 1. Introduction

Reflection electron microscopy (REM) allows the direct observation of such surface topographical details as single-atom steps, dislocations and surface particles (Yagi, 1987). The best REM contrast conditions are often achieved for incident beam angles where electron reflection is not necessarily supported

### References

- ALLEGRA, G. (1979). *Acta Cryst.* **A35**, 213–220.  
 BURLA, M. C., CAMALLI, M., CASCARANO, G., GIACOVAZZO, C., POLIDORI, G., SPAGNA, R. & VITERBO, D. (1989). *J. Appl. Cryst.* **22**, 389–393.  
 CAPASSO, S., MAZZARELLA, L., SICA, F., ZAGARI, A., CASCARANO, G. & GIACOVAZZO, C. (1991). Submitted to *Acta Cryst.* B.  
 CASCARANO, G., GIACOVAZZO, C. & LUIĆ, M. (1988). *Acta Cryst.* **A44**, 176–183, 183–188.  
 CASCARANO, G., GIACOVAZZO, C. & VITERBO, D. (1987). *Acta Cryst.* **A43**, 22–29.  
 COCHRAN, W. (1952). *Acta Cryst.* **5**, 65–67.  
 COCHRAN, W. & WOOLFSON, M. M. (1955). *Acta Cryst.* **8**, 1–12.  
 GIACOVAZZO, C. (1991). *Acta Cryst.* **A47**, 256–263.  
 HAUPTMAN, H. & KARLE, J. (1953). *The Solution of the Phase Problem. I. The Centrosymmetric Crystals.* ACA Monogr. No. 3. New York: Polycrystal Book Service.  
 HOPPE, W. (1963). *Acta Cryst.* **16**, 1056–1057.  
 KARLE, J. & KARLE, I. (1966). *Acta Cryst.* **21**, 849–859.  
 MAIN, P. & WOOLFSON, M. M. (1963). *Acta Cryst.* **16**, 1046–1051.  
 NIXON, P. E. (1978). *Acta Cryst.* **A34**, 450–453.  
 RIUS, J. & MIRAVITLLES, C. (1989). *Acta Cryst.* **A45**, 490–494.  
 STANLEY, E. (1979). *Acta Cryst.* **A35**, 966–970.  
 STANLEY, E. (1986). *Acta Cryst.* **A46**, 297–299.

by strong bulk diffraction (Uchida, Lehmpfuhl & Jager, 1984). The choice of incidence angle is then guided by surface diffraction considerations and not only by supposed satisfaction of kinematical bulk Bragg reflection diffraction.

One such incident direction which can produce a strong reflection corresponds to the so-called surface wave resonance (SWR) effect. In SWR one of the diffracted beams is trapped in a state close to the surface and is not able to propagate in vacuum. This can be understood as an Ewald-sphere tangency condition and incident conditions can be determined by a simple geometric construction (Ichimiya, Kambe & Lehmpfuhl, 1980).

Strong contrast effects in REM are also achievable by illuminating at incident conditions corresponding to 'forbidden' reflections, e.g. 666 for the Si (111) surface (Uchida & Lehmpfuhl, 1987). The reflection is then not directly supported by a single bulk diffraction event, but by multiple diffraction events [*Umweganregung* (von Laue, 1948)]. However, from the viewpoint of a surface diffraction process, such