SAYRE, D. (1952). *Acta Cryst.* **5**, 60-65.
SAYRE, D. (1972). *Acta Cryst.* **A28**, 210-212.
SAYRE, D. (1974). *Acta Cryst.* **A30**, 180-184.
SCHEVITZ, R. W., NAVIA, M. A., BANTZ, D. A., CORNICK, G., ROSA, J. J., ROSA, M. D. H. & SIGLER, P. B. (1972). *Science*, **177**, 429-431.
SIM, G. A. (1959). *Acta Cryst.* **12**, 813-815.
STEITZ, T. A. (1968). *Acta Cryst.* **B24**, 504-507.
STEITZ, T. A., FLETTERICK, R. J. & HWANG, K. J. (1973). *J. Mol. Biol.* **78**, 551-561.
STEWART, J. M., KARLE, J., IWASAKI, H. & ITO, T. (1977). *Acta Cryst.* **A33**, 519.
SUSSMAN, J. L., HOLBROOK, S. R., CHURCH, G. M. & KIM, S.-H. (1977). *Acta Cryst.* **A33**, 800-804.
TAYLOR, D. J., WOOLFSON, M. M. & MAIN, P. (1978). *Acta Cryst.* **A34**, 870-883.
TEMPLETON, D. H., TEMPLETON, L. K., PHILLIPS, J. C. & HODGSON, K. O. (1980). *Acta Cryst.* **A36**, 436-442.
TIMKOVICH, R. & DICKERSON, R. E. (1976). *J. Biol. Chem.* **251**, 4033-4046.
TOEPLITZ, O. (1911). *Rend. Circ. Mat. Palermo*, pp. 191-192.
TSOUCARIS, G. (1970). *Acta Cryst.* **A26**, 492-499.
TSOUCARIS, G. (1980). In *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. LADD & R. A. PALMER, pp. 287-360. New York: Plenum Press.
WANG, B. C. (1985). In *Diffraction Methods in Biological Macromolecules*, edited by H. WYCKOFF, pp. 90-112. New York: Academic Press.
WEINZIERL, J. E., EISENBERG, D. & DICKERSON, R. E. (1969). *Acta Cryst.* **B25**, 380-387.
WESTBROOK, E., PIRO, O. E. & SIGLER, P. B. (1984). *J. Biol. Chem.* **259**, 9096-9103.
WILSON, K. S. (1978). *Acta Cryst.* **B34**, 1599-1608.
WOOLFSON, M. M. (1954). *Acta Cryst.* **7**, 61-64.
WOOLFSON, M. M. (1984). *Acta Cryst.* **A40**, 32-34.
WOOLFSON, M. M. & YAO JIA-XING (1988). *Acta Cryst.* **A44**, 410-413.
XU, Z. B., YANG, D. S. C., FUREY, W. JR, SAX, M., ROSE, J. & WANG, B. C. (1984). Am. Crystallogr. Assoc. Meet., 20-25 May, Abstr. PC2, p. 50.
YANG, D. S. C., XU, Z. B., FUREY, W. JR & WANG, B. C. (1984). Am. Crystallogr. Assoc. Meet., 20-25 May, Abstr. PC1, p. 50.
YAO JIA-XING (1981). *Acta Cryst.* **A37**, 642-644.
YAO JIA-XING & FAN HAI-FU (1985). *Acta Cryst.* **A41**, 284-285.

Acta Cryst. (1989). **A45**, 781-786

From a Partial to the Complete Crystal Structure. III. A New Procedure and its Applications

BY M. C. BURLA

Dipartimento di Scienze della Terra, Università, 06100 Perugia, Italy

G. CASCARANO

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

V. FARES

ITSE, CNR, CP 10, 00016, Monterotondo Stazione, Roma, Italy

C. GIACOVAZZO

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

G. POLIDORI

Dipartimento di Scienze della Terra, Università, 06100 Perugia, Italy

AND R. SPAGNA

Istituto di Strutturistica Chimica 'G. Giacomello', CNR, CP 10, 00016 Monterotondo Stazione, Roma, Italy

(*Received 23 February* 1989; *accepted 1 June* 1989)

Abstract

The procedure described in the first two papers of this series [Giacovazzo (1983). *Acta Cryst.* **A39**, 685-692; Camalli, Giacovazzo & Spagna (1985). *Acta Cryst.* **A41**, 605-613] for recovering from a partial crystal structure the complete one has been reconsidered. Several modifications have been introduced which make the procedure more efficient. Furthermore, a new method is described which is able to combine prior information on a located fragment with the second representation formula *P*10. Experimental tests show that very small fragments are sufficient to recover the complete crystal structure.

Symbols and abbreviations

Throughout the paper a number of symbols will find frequent application. For the sake of simplicity they are here listed together.

$f(\mathbf{h})$: atomic scattering factor. The thermal factor is included; anomalous dispersion is not considered.
N: number of atoms in the unit cell.

0108-7673/89/110781-06$03.00

p: number of atoms (symmetry equivalents included) whose positions are *a priori* known.
q: number of atoms (symmetry equivalents included) whose positions are unknown: $q = N - p$.
$F_{\mathbf{h}}$: structure factor with vectorial index **h**.
$\sum_p, \sum_q, \sum_N = \sum f_j^2$: the summation is extended to p, q, N atoms.
$E_{\mathbf{h}} = R_{\mathbf{h}} \exp(i\varphi_{\mathbf{h}})$: normalized structure factor in absence of any prior information; $R_{\mathbf{h}}$ is its magnitude, $\varphi_{\mathbf{h}}$ its phase.
$E_{p,\mathbf{h}} = R_{p,\mathbf{h}} \exp(i\varphi_{p,\mathbf{h}}) = F_{p,\mathbf{h}}/\sum_N^{1/2}$: normalized structure factor corresponding to the partial structure; $R_{p,\mathbf{h}}$ is its modulus, $\varphi_{p,\mathbf{h}}$ its phase.
$\Phi = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} - \varphi_{\mathbf{h}+\mathbf{k}}$.
$\Phi_p = \varphi_{p,\mathbf{h}} + \varphi_{p,\mathbf{k}} - \varphi_{p,\mathbf{h}+\mathbf{k}}$.
$E'_{\mathbf{h}}$: pseudo-normalized structure factor with vectorial index **h** defined by $E'_{\mathbf{h}} = F_{\mathbf{h}}/\sum_q^{1/2}$. $R'_{\mathbf{h}}$ is its modulus.
$E'_{p,\mathbf{h}}$: pseudo-normalized structure factor of the partial structure with p atoms in the unit cell, defined by

$$E'_{p,\mathbf{h}} = F_{p,\mathbf{h}}/\textstyle\sum_q^{1/2}.$$

$D_1(x) = I_1(x)/I_0(x)$: I_1 and I_0 are modified Bessel functions of order 1 and zero respectively.
Other locally used symbols are defined in the text.

The papers by Giacovazzo (1983) and by Camalli, Giacovazzo & Spagna (1985) will be referred to as papers I and II respectively.

1. Introduction

Recovering the complete structure from a partial one is one of the most common and critical steps in the process for crystal structure analysis. Weighted Fourier syntheses (Woolfson, 1956; Sim, 1960; Ramachandran & Srinivasan, 1970, and literature there quoted) would reveal the unknown atomic positions better than usual syntheses with $|F| \exp(i\varphi_p)$, but they work well only if a relevant part of the crystal structure is *a priori* known. Thus special procedures have been devised which may be described as follows.

(*a*) Tangent recycling methods (Karle, 1970; see also Hull & Irwin, 1978). Phases φ are accepted as good approximations of phases φ if $|F_p| > \eta|F|$, where η is the fraction of the total scattering power contained in the fragment and where $|F|$ is associated with an $|E| \geq 1{\cdot}5$. Phase expansion is obtained by application of the usual tangent formulas.

(*b*) Tangent recycling methods applied to difference structure factors (Beurskens, Prick, Doesburg & Gould, 1979). Difference structure factors $\Delta F = (|F| - |F_p|) \exp i\varphi_p$ are calculated and, in favourable cases, accepted as a first estimation of F_q. The weighted tangent formula is applied to F_q which can vary both in magnitude and phase.

(*c*) Joint probability distribution methods (Main, 1976; Heinerman, 1977). The prior information is exploited in order to obtain more accurate probabilistic formulas estimating Φ. In this method normalized structure factors $E_{\mathbf{h}}$ are defined by

$$E_{\mathbf{h}} = F_{\mathbf{h}}/\langle F_{\mathbf{h}}^2\rangle^{1/2}_{\text{p.r.v.}} = F_{\mathbf{h}}/[F^2_{p,\mathbf{h}} + \textstyle\sum_q]^{1/2},$$

where $\langle |F_{\mathbf{h}}|^2\rangle_{\text{p.r.v.}}$ denotes the average of $|F_{\mathbf{h}}|^2$ with respect to primitive random variables.

New insight into this method was afforded by paper I of this series: the coordinates of the p atoms were assumed to be fixed parameters of the distribution while the other $q = N - p$ atomic positions define the primitive random variables. A basic result of paper I was the conditional joint probability distribution function [equation (I.*B*1)]

$$P(\varphi_{\mathbf{h}}, \varphi_{\mathbf{k}}, \varphi_{\mathbf{h}-\mathbf{k}} \| E_{\mathbf{h}}|, |E_{\mathbf{k}}|, |E_{\mathbf{h}-\mathbf{k}}|, E_{p,\mathbf{h}}, E_{p,\mathbf{k}}, E_{p,\mathbf{h}-\mathbf{k}})$$

from which the conditional distribution [equation (I.24)]

$$P(\varphi_{\mathbf{h}} \| E_{\mathbf{h}}|, E_{\mathbf{k}}, E_{\mathbf{h}-\mathbf{k}}, E_{p,\mathbf{h}}, E_{p,\mathbf{k}}, E_{p,\mathbf{h}-\mathbf{k}})$$

may be derived: from (I.24) the special tangent formula (I.25) arises. The distribution (I.24) was simplified in paper II, where pseudo-normalized structure factors E' were introduced, together with a description of a procedure for the choice of the starting set of phases, of a special weighted tangent formula and of some figures of merit aimed at recovering the correct from various trial solutions. The first applications of the method were successful and the procedure was implemented in the program for automatic phase solution *SIR*85 (Cascarano, Giacovazzo, Burla, Nunzi, Polidori, Camalli, Spagna & Viterbo, 1985).

This paper describes further developments of the ideas that form the basis of papers I and II. In § 2 the procedure proposed in paper II (from now on procedure *A*) has been reconsidered and suitably modified in order to make the recovery of the complete crystal structure more efficient. In § 3 a new approach is described (from now on procedure *B*) which is able to combine prior information about the located fragment with the *P*10 (Cascarano, Giacovazzo, Camalli, Spagna, Burla, Nunzi & Polidori, 1984) formula. Applications are described in § 4.

2. A reconsideration of procedure *A*: the modified *A* procedure

The procedure *A* described in paper II may be described as follows:

(*a*) The origin and enantiomorph-defining phases are chosen among the NUMK reflexions characterized by $\alpha_{\mathbf{h}} > 10$, where $\alpha_{\mathbf{h}}$ is a reliability parameter which takes into account Sim's and triplet contributions. Values $\theta_{\mathbf{h}}$, obtained by application of (II.23), are assumed to be reliable estimates of the true (for the complete crystal structure) phases $\varphi_{\mathbf{h}}$ and constitute the starting set of the phasing process. In order to release too heavy restrictions on the refinement process the limitation NUMK ≤ 100 was introduced.

(*b*) Phases, varying according to the magic-integer method (Main, 1978) are chosen among the $\varphi_{\mathbf{h}}$ phases characterized by the largest differences $(\alpha_{\mathbf{h}})_{\max} - \alpha_{\mathbf{h}}$.

(*c*) Phase extension and refinement are made according to the weighted tangent formula (II.24). Phase indications are accepted if $\alpha^2 > \mathrm{CUT}$, where CUT is a threshold which is reduced cycle by cycle by the factor 0·65. At the end of each cycle the average $\langle\alpha^2\rangle_{\mathrm{pc}}$ on the accepted phases is calculated and used in the succeeding cycle to obtain the weight

$$w_{\mathbf{h}} = (\alpha_{\mathbf{h}}^2/\langle\alpha^2\rangle_{\mathrm{pc}})^{1/4}.$$

$w_{\mathbf{h}}$ cannot exceed 1 or be smaller than 0·15.

Even if successful in most cases (see § II.3) the procedure failed in some others. A careful reconsideration of its various steps suggested to us to introduce some modifications in order to make the procedure (from now on referred to as the modified *A* procedure) more flexible and effective.

(*a*) The core of the weighted tangent formula (II.24) is

$$\begin{aligned} q^{-1/2}\sum_{\mathbf{k}}[&Q_1 w_{\mathbf{k}} w_{\mathbf{h-k}} \exp i(\varphi_{\mathbf{k}} + \varphi_{\mathbf{h-k}}) \\ &- Q_2 w_{\mathbf{h-k}} \exp i(\varphi_{p,\mathbf{k}} + \varphi_{\mathbf{h-k}}) \\ &- Q_3 w_{\mathbf{k}} \exp i(\varphi_{\mathbf{k}} + \varphi_{p,\mathbf{h-k}}) \\ &+ Q_4 \exp i(\varphi_{p,\mathbf{k}} + \varphi_{p,\mathbf{h-k}})], \end{aligned} \tag{1}$$

where

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

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

Different values for $w_{\mathbf{k}}$ and $w_{\mathbf{h-k}}$ lead to remarkable differences not only in the $\alpha_{\mathbf{h}}$ parameter, but also in the estimated value of $\varphi_{\mathbf{h}}$. Thus severe errors in $w_{\mathbf{k}}$ and $w_{\mathbf{h-k}}$ can make (II.24) unstable. In order to reduce such a critical behaviour we have replaced in (II.25) the term (1) by

$$\begin{aligned} q^{-1/2}\sum_{\mathbf{k}} w_{\mathbf{k}} w_{\mathbf{h-k}}[&Q_1 \exp(\varphi_{\mathbf{k}} + \varphi_{\mathbf{h-k}}) \\ &- Q_2 \exp i(\varphi_{p,\mathbf{k}} + \varphi_{\mathbf{h-k}}) \\ &- Q_3 \exp i(\varphi_{\mathbf{k}} + \varphi_{p,\mathbf{h-k}}) \\ &+ Q_4 \exp i(\varphi_{p,\mathbf{k}} + \varphi_{p,\mathbf{h-k}})]. \end{aligned}$$

According to similar considerations we have replaced $Q_{\mathbf{k}}$ in (II.22) by

$$D_1(G_{\mathbf{k}})D_1(G_{\mathbf{h-k}})[Q_1 - Q_2 - Q_3 + Q_4].$$

(*b*) The threshold $\alpha_{\mathbf{h}} > 10$ which selects in the original *A* procedure the origin and enantiomorph-defining phases is too severe when the position of a small fragment (in terms of scattering power) is *a priori* known. The following criterion was then used. Up to 15 phases with $\alpha \geq 4$ are chosen to select origin and enantiomorph: if they cannot be found, up to 12 phases with $\alpha \geq 3$, or up to 8 with $\alpha > 2{\cdot}5$, or up to 4 with $\alpha > 1{\cdot}5$ are used.

(*c*) It is noted in paper II that large α values are usually involved in the process during the first cycles of the phase expansion. When several wrong phase indications are used during a given cycle then the experimental $\langle\alpha^2\rangle$ drops down. A sensitive criterion was then needed to prevent harmful subsequent phase expansion. Phases were accepted during the *n*th cycle only if

$$\alpha_{\mathrm{calc}}^2 \geq (\mathrm{CUT} + 0{\cdot}01\langle\alpha^2\rangle_{\mathrm{pc}}),$$

where $\langle\alpha^2\rangle_{\mathrm{pc}}$ is the experimental mean value obtained for the $(n-1)$th cycle. As a consequence, the modified tangent procedure stops in a smaller number of cycles: sometimes a relatively large number of reflexions have undetermined phases but determined phases are more likely correct.

(*d*) The specific figures of merit (say R_p and ψ_{op} described in § 2.5 of paper II) for recognizing the correct among the different solutions, even if successful in most cases, present two drawbacks: (i) they are intrinsically dependent on the prior information (which may be wrong) through the phases $\varphi_{p,\mathbf{h}}$. Thus good figures of merit can sometimes be obtained from wrongly positioned fragments; (ii) R_p and ψ_{op} are not on an absolute scale: *i.e.* it is not possible to judge from their experimental value the probability that a given solution is correct.

We have therefore modified the procedure in order to recover the *E* from the *E*′ values, and to calculate from them the figures of merit described in a recent paper by Cascarano, Giacovazzo & Viterbo (1987). These figures are independent of the prior structural information and provide an absolute criterion for judging the reliability of each trial.

3. An integration between the *P*10 formula and prior information: procedure *B*

The representations theory [(Giacovazzo, 1977, 1980); for a related method, see Hauptman (1976)] showed that, for any structure invariant or seminvariant Φ, the set of diffraction magnitudes may be arranged in a sequence of subsets in order of their expected effectiveness (in the statistical sense) for the estimation of Φ. Thus different formulas estimating Φ can be obtained, each exploiting a different subset of diffraction magnitudes, and therefore a different amount of prior information. For a triplet invariant the first representation formula exploits only $R_{\mathbf{h}}$, $R_{\mathbf{k}}$, $R_{\mathbf{h+k}}$, and in most cases it coincides with Cochran's (1955) formula.

The second representation formula for a triplet phase is based on the combined estimate of the special quintets

$$\psi = \{\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} - \varphi_{\mathbf{h+k}} + \varphi_{\mathbf{l}} - \varphi_{\mathbf{l}}\}$$

and therefore it exploits the larger subset of magnitudes

$$\{R_{\mathbf{h}}, R_{\mathbf{k}}, R_{\mathbf{h+k}}, R_{\mathbf{l}}, R_{\mathbf{h\pm l}}, R_{\mathbf{k\pm l}}, R_{\mathbf{h+k\pm l}}\},$$

where $\mathbf{l}$ is a free reciprocal vector.

The $P10$ formula (Cascarano *et al.*, 1984) is an effective example of a second-representation formula: it proved in practice much more powerful than Cochran's formula, so that it is the default choice of the *SIR* program for solving complex crystal structures. For such structures it is not infrequent that a molecular fragment may be located by the $P10$ formula, while Cochran's formula is unsuccessful. Thus a limit of procedure A (modified or not) may be described as follows: procedure A is, on the one hand, able to exploit prior information provided by the located fragment, but, on the other hand, it is unable to exploit the power of the $P10$ formula.

The problem may be raised now of whether the information on the located fragment may be associated with the $P10$ formula. The canonical mathematical two-step process for gaining this end may be described as follows:

(a) Calculate the joint probability distribution function

$$P(\{E_{\mathbf{h}}, E_{\mathbf{k}}, E_{\mathbf{h+k}}, E_{\mathbf{l}}, E_{\mathbf{h\pm l}}, E_{\mathbf{k\pm l}}, E_{\mathbf{h+k\pm l}}\}|\{E_{p,\mathbf{h}}, E_{p,\mathbf{k}}, \ldots, E_{p,\mathbf{h+k\pm l}}\}).$$

(b) Calculate the conditional distribution

$$P(\Phi|\{R_{\mathbf{h}}, R_{\mathbf{k}}, \ldots, R_{\mathbf{h+k\pm l}}, E_{p,\mathbf{h}}, E_{p,\mathbf{k}}, \ldots, E_{p,\mathbf{h+k\pm l}}\}). \tag{2}$$

The rigorous algebraic form of (2) is not easy to derive, and probably too time consuming for computer applications. Therefore we decided to obtain an approximate distribution according to the following considerations:

(i) Only terms up to order $1/q^0$ are retained in the joint probability distribution function (II.A1), from which (see Appendix)

$$P(\Phi|R'_{\mathbf{h}}, R'_{\mathbf{k}}, R'_{\mathbf{h+k}}, E'_{p,\mathbf{h}}, E'_{p,\mathbf{k}}, E'_{p,\mathbf{h+k}}) \sim L' \exp[S\cos(\Phi - \Phi_p)], \tag{3}$$

where

$$D_1(S) = D_1(G_{\mathbf{h}})D_1(G_{\mathbf{k}})D_1(G_{\mathbf{h-k}})$$

and

$$G = 2R'R'_p.$$

Equation (3) provides an estimate for the triplet phase ($\Phi \simeq \Phi_p$) when only Sim's contribution is available (triplet contributions of order $q^{-1/2}$ are completely neglected).

Table 1. *Crystallographic data for the twelve structures selected for study*

Structure*	Space group	Molecular formula	Z
MUNICH	$C2$	$C_{20}H_{16}$	8
SCHWZ2	$P1$	$C_{46}H_{70}O_{27}$	1
GOLDMAN2	Cc	$C_{28}H_{16}$	8
CEPHA	$C2$	$C_{18}H_{21}NO_3$	8
AZET	$Pca2_1$	$C_{21}H_{16}ClNO$	8
ERGO	$P2_12_12_1$	$C_{28}H_{44}O$	8
APAPA	$P4_12_12$	$C_{30}H_{37}N_{15}O_{16}P_2.6H_2O$	8
BED2	$I4$	$C_{26}H_{26}N_4O_4$	8
TPH	$B22_12$	$C_{24}H_{20}N_2$	12
TVAL	$P1$	$C_{54}H_{90}N_6O_{18}$	2
TPALA	$P_{\underline{2}}$	$C_{28}H_{42}N_4O_7$	2
GRA4	$P\bar{1}$	$C_{30}H_{22}N_2O_4$	2

* Complete references for test structures are not given for the sake of brevity. The reader is referred to magnetic tapes distributed by the crystallographic groups in Göttingen and York.

The asymptotical behaviour of (3) is satisfactory. Indeed, for a very small located fragment, $S = 0$ as expected. For a large fragment S is expected to be large too (in the statistical sense) so that $\Phi \simeq \Phi_p$ is well supported. If the structure is almost completely located then $S \simeq \infty$ and strictly $\Phi = \Phi_p$.

(ii) As a first approximation (3) can be considered statistically independent of the $P10$ formula. This is of von Mises type and is calculated in the absence of any information on the located fragment: *i.e.*

$$P(\Phi|\{R_{\mathbf{h}}, R_{\mathbf{k}}, R_{\mathbf{h+k}}, R_{\mathbf{l}}, R_{\mathbf{h\pm l}}, R_{\mathbf{k\pm l}}, R_{\mathbf{h+k\pm l}}\}) \sim L\exp(K\cos\Phi) \tag{4}$$

with positive, null or negative values of K. Therefore (4) may be easily combined with (3) in order to give the desired formula

$$\begin{aligned} P(\Phi|\ldots) &\sim L^{-1}\exp[S\cos(\Phi - \Phi_p) + K\cos\Phi] \\ &= L^{-1}\exp[\alpha\cos(\Phi - \Theta)], \end{aligned} \tag{5}$$

where

$$\tan\Theta = (S\sin\Phi_p)/(S\cos\Phi_p + K) = A/B \tag{6}$$

provides the estimate for Φ, and

$$\alpha = (A^2 + B^2)^{1/2} \tag{7}$$

is the concentration parameter of the distribution (5). Because of the prior information on the located fragment, Θ may lie anywhere between 0 and 2π; furthermore, the magnitude of α depends on the magnitudes of S and K, and on the sign of K.

If the ratio S/K is large (as, in a statistical sense, when a large molecular fragment is located) then $\Theta \sim \Phi_p$ as may be expected. If the ratio S/K is small (as, in the statistical sense, when only a small molecular fragment is located) then the $P10$ formula prevails: indeed Θ will be close to 0 or π according to the sign of K. Thus the full power of the $P10$ formula is preserved in (5).

It is worth mentioning that K can represent the concentration parameter of the $P10$ formula as well

Table 2. *Number of atoms found by SIR in default run and when some molecular fragment is a priori located*

The fractional scattering power of the fragment and the combined figures of merit corresponding to the most probable solutions are also given.

Structure	Default run: number of atoms/total	Mol. frag.	F.s.p. (×100)	Modified *A*: number of atoms/total	*B*: number of atoms/total
MUNICH	*	C_4	9·8	26/40 (0·386)	38/40 (0·933)
SCHWZ2	*	C_5O_3	10·8	12/73 (0·364)	54/73 (0·717)
GOLDMAN2	*	C_5	8·8	11/56 (0·44)	43/56 (0·514)
CEPHA	*	C_3O_2	13·0	14/46 (0·276)	46/46 (0·976)
AZET	16/48 (0·713)	2Cl	24·6	12/48 (0·697)	27/48 (0·696)
ERGO	*	C_5	8·1	46/58 (0·448)	54/58 (0·727)
APAPA	46/69 (0·881)	C_3N_2	5·5	29/69 (0·798)	46/69 (0·964)
BED2	20/34 (0·633)	C_6	15·3	22/34 (0·527)	23/34 (0·602)
TPH	*	C_6	14·7	38/39 (0·712)	38/39 (0·952)
TVAL	*	$C_{12}NO_3$	9·7	26/156 (0·967)	75/156 (0·984)
TPALA	*	C_4N	11·3	26/39 (0·351)	0/39 (0·359)
GRA4	32/36 (0·839)	C_6	14·8	26/36 (0·695)	32/36 (0·600)

Table 3. *The number of atoms found by procedure B when some incorrectly located atoms are added to a correctly located fragment*

Structure	Correctly located fragment	Incorrect fragment	Number of atoms/total
MUNICH	C_4	C_2	35/40 (0·928)
CEPHA	C_3O_2	O_2	44/46 (0·968)
ERGO	C_5	C_3	44/58 (0·708)
APAPA	C_3N_2	C_3	40/69 (0·975)

as that of Cochran's (1955) formula. If *K* coincides with Cochran's term then (5) can be considered roughly equivalent to the probabilistic formulas described in the papers I and II.

In the present *B* procedure parameters (6) and (7) are calculated for each triplet: for convenience the α's are rescaled on the *K* parameters and *S* has been divided by 2 (it seemed to us that the relation $\Phi \sim \Phi_p$ is rather overestimated through *S*), but that is not essential for the success of the procedure. Such information is stored in a suitably modified $\sum_2$ list. Origin and enantiomorph-defining phases are chosen exactly as in the modified *A* procedure. Accordingly, symbolic phases for the multisolution process are chosen among the $\varphi_{\mathbf{h}}$ phases characterized by the largest inconsistencies of the $(\varphi_{\mathbf{k}}+\varphi_{\mathbf{h}-\mathbf{k}})$ pairs.

4. Applications

The modified *A* and *B* procedures were both applied to 12 structures: 11 of them are in the database of difficult structures compiled by Sheldrick, the 12th (GRA4) is a difficult structure contained in a file distributed by the crystallographic group in York.

Table 1 tabulates the salient features of the structure. Table 2 tabulates the results which could be found after application of a default run of *SIR* (without any prior information), and of the modified *A* and *B* procedures. Structures starred in column 2 are not solved by default runs of *SIR* [however, it cannot be concluded that *SIR* is unable to solve them using non-default conditions: see, for example, Burla, Giacovazzo & Polidori (1987, 1989)]. Such structures were tested in order to verify whether the supplementary information provided by some located molecular fragment is sufficient for the structure solution. Non-starred structures were solved by the default run of *SIR*85 (*via* the use of the *P*10 formula): in column 2 the number of located atoms over the total is given. Such structures were tested in order to verify whether the available partial structure information makes the structure solution easier even when *P*10 information is not used.

For each structure the molecular fragment in a known position (Mol. frag.), the fractional scattering power (f.s.p.) corresponding to it and the ratios number of atoms found/total number of atoms are given. The last entries always correspond to the solution with the highest combined figures of merit CFOM, which are quoted in parentheses (correct solutions are expected to be marked by CFOM = 1).

Table 2 indicates that both procedures are powerful tools for recovering the complete from a partial structure even starting from small fragments. The procedure *B* seems remarkably more efficient in all cases except for TPALA where a failure occasionally occurs (but an additional test based on the fragment C_4N + CO correctly provides the location of 27 atoms). The modified *A* procedure also is less efficient for structures not starred in column 2 (compare columns 5 and 6). The loss of information provided by the *P*10 formula is hardly compensated by the information provided by the located fragment.

Supplementary tests were made in order to check the robustness of procedure *B*. Some atoms in wrong positions were added to the correctly located fragments shown in Table 2 and *SIR* was run in order to recover from them the complete crystal structure. The results, shown in Table 3, indicate that procedure *B* is sufficiently robust to be useful in most practical cases. We then decided to introduce it into the release

*SIR*88 (Burla, Camalli, Cascarano, Giacovazzo, Polidori, Spagna & Viterbo, 1989) of the *SIR* program.

Financial support for this work was provided by the CNR project *Metodologie Cristallografiche Avanzate.*

APPENDIX

If in the distribution (II.$A1$) only terms up to order $1/q^0$ are retained the result is

$$P(\varphi_1, \varphi_2, \varphi_3 | R'_i, R'_{pi}, \varphi_{pi}, i=1,2,3) \simeq L^{-1} \exp\left\{\sum_{i=1}^{3} G_i \cos(\varphi_i - \varphi_{pi})\right\}. \qquad (A1)$$

In this case φ_1, φ_2, φ_3 act as mutually independent variables, each φ_i being distributed according to the von Mises distribution

$$M(\varphi_i;\, \theta_i,\, G_i) = [2\pi I_0(G_i)]^{-1} \exp[G_i \cos(\varphi_i - \theta_i)],$$

where $\theta_i = \varphi_{pi}$. The following general theorem may then be applied to ($A1$) (Giacovazzo, 1979): let φ_1, $\varphi_2, \ldots, \varphi_n$ be n mutually independent variables and let any φ_i be distributed according to $M(\varphi_i;\, \theta_i,\, G_i)$. Then the variable sum $\Phi = \varphi_1 + \varphi_2 + \ldots + \varphi_n$ can be approximated by the von Mises distribution $M(\Phi;\, \theta_1 + \theta_2 + \ldots + \theta_n,\, S)$ where S satisfies

$$D_1(S) = D_1(G_1) D_1(G_2) \ldots D_1(G_n).$$

For $n=3$ this theorem gives the result provided by distribution (3).

References

Beurskens, P. T., Prick, A. J., Doesburg, H. M. & Gould, R. O. (1979). *Acta Cryst.* **A35**, 765-772.
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., Giacovazzo, C. & Polidori, G. (1987). *Acta Cryst.* **A43**, 797-802.
Burla, M. C., Giacovazzo, C. & Polidori, G. (1989). *Acta Cryst.* **A45**, 99-104.
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). IX Eur. Crystallogr. Meet., Torino, Italy. Abstract 1-046.
Cascarano, G., Giacovazzo, C., Camalli, M., Spagna, R., Burla, M. C., Nunzi, A. & Polidori, G. (1984). *Acta Cryst.* **A40**, 278-283.
Cascarano, G., Giacovazzo, C. & Viterbo, D. (1987). *Acta Cryst.* **A43**, 22-29.
Cochran, W. (1955). *Acta Cryst.* **8**, 473-478.
Giacovazzo, C. (1977). *Acta Cryst.* **A33**, 933-944.
Giacovazzo, C. (1979). *Acta Cryst.* **A35**, 757-764.
Giacovazzo, C. (1980). *Acta Cryst.* **A36**, 362-372.
Giacovazzo, C. (1983). *Acta Cryst.* **A39**, 685-692.
Hauptman, H. (1976). *Acta Cryst.* **A32**, 934-940.
Heinerman, J. J. L. (1977). *Acta Cryst.* **A33**, 100-106.
Hull, S. E. & Irwin, M. J. (1978). *Acta Cryst.* **A34**, 863-870.
Karle, J. (1970). In *Crystallographic Computing*, edited by F. R. Ahmed, pp. 155-164. Copenhagen: Munksgaard.
Main, P. (1976). In *Crystallographic Computing Techniques*, edited by F. R. Ahmed, pp. 97-105. Copenhagen: Munksgaard.
Main, P. (1978). *Acta Cryst.* **A34**, 31-38.
Ramachandran, G. N. & Srinivasan, R. (1970). *Fourier Methods in Crystallography.* New York: Wiley-Interscience.
Sim, G. A. (1960). *Acta Cryst.* **13**, 511-512.
Woolfson, M. M. (1956). *Acta Cryst.* **9**, 804-810.

Acta Cryst. (1989). A**45**, 786-793

Relativistic Hartree–Fock X-ray and Electron Atomic Scattering Factors at High Angles

By A. G. Fox* and M. A. O'Keefe

Center for Advanced Materials and National Center for Electron Microscopy, Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720, *USA*

and M. A. Tabbernor

*School of Construction, Engineering and Technology, The Polytechnic, Wulfruna Street, Wolverhampton WV*1 1*SB, England*

(*Received* 12 *December* 1988; *accepted* 21 *June* 1989)

Abstract

An enlarged set (atomic number $Z=2$ to 98) of free-atom X-ray atomic scattering (form) factors for high angles $[2 \leq (\sin\theta)/\lambda \leq 6\ \text{Å}^{-1}]$ has been calculated based on those of Doyle & Turner [*Acta Cryst.* (1968). **A24**, 390-397]. Four-parameter 'exponential polynomial' fits for these are presented which give far more accurate estimates of the scattering factors at high angles than the Gaussian fits normally used.

---

* Present address: Naval Postgraduate School, Monterey, CA 93943-5000, USA.

0108-7673/89/110786-08$03.00