

(4) ATBs were preferred in the polycrystal rather than STBs [except for the (111) $\Sigma = 3$ STB] or TGBs. In particular, ATBs in the $\Sigma = 3$ system in the (011) zone and closer to the (111) STB than the (112) STB occurred frequently. These occurrences could be explained in terms of grain-boundary energy considerations.

References

- BRANDON, D. G. (1966). *Acta Metall.* **14**, 1479–1490.
 CRAWFORD, D. C. & WAS, G. (1991). *J. Electron Microsc. Tech.* **19**, 345–352.
 DASH, S. & BROWN, N. (1963). *Acta Metall.* **11**, 1067.
 FULLMAN, R. L. (1951). *J. Appl. Phys.* **22**, 456–460.
 FURLEY, J. & RANDLE, V. (1991). *Mater. Sci. Technol.* **7**, 12–19.
 OMAR, R. (1987). PhD thesis, Univ. of Warwick, England.
 PAIDAR, V. (1987). *Acta Metall.* **35**, 2035–2048.
 PALUMBO, G. & AUST, K. T. (1990). *Acta Metall.* **38**, 2343–2352.
 RANDLE, V. (1989). *Scr. Metall.* **23**, 773–778.
 RANDLE, V. (1991). *Mater. Sci. Technol.* **7**, 985–990.
 RANDLE, V. (1992). *Microtexture Determination and its Applications*. London: Institute of Materials.
 RANDLE, V. (1993a). *Grain Boundary Geometry in Cubic Polycrystals*. Bristol: Institute of Physics.
 RANDLE, V. (1993b). Proceedings of Conference on Grain Boundary Engineering, Kingston, Canada. In the press.
 RANDLE, V. & DINGLEY, D. J. (1989). *Scr. Metall.* **23**, 1565–1570.
 RANDLE, V. & DINGLEY, D. J. (1990). *Proceedings of EUROMAT 89, Advanced Materials and Processes*, edited by H. E. EXNER & V. SCHUMACHER. Oberusel: DGM.
 SARGENT, C. M. (1968). *Trans. Metall. Soc. AIME*, **242**, 1188–1191.
 VAUGHAN, D. (1970). *Philos. Mag.* **40**, 1003–1011.
 WOLF, D. (1985). *J. Phys.* **46**, Colloq. C4, 197–208.
 WOLF, U., ERNST, F., MUSCHIK, T., FINNIS, M. W. & FISCHMEISTER, H. F. (1992). *Philos. Mag.* **A66**, 991–1016.

Acta Cryst. (1994). **A50**, 595–601

Probabilistic Estimation of the Structure Seminvariants in the Monoclinic and Orthorhombic Systems when Anomalous Scatterers are Present

BY YONG-SHENG LIU AND NING-HAI HU

Applied Spectroscopy Laboratory, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun 130022, People's Republic of China

(Received 29 October 1993; accepted 15 March 1994)

Abstract

The estimate formulas for the two-phase structure seminvariants (TPSSs) in the presence of anomalous scattering are obtained from the estimate of the two-phase structure invariants [Hauptman (1982). *Acta Cryst.* **A38**, 632–641; Giacovazzo (1983). *Acta Cryst.* **A39**, 585–592] and the conditional probability distribution of the TPSSs is derived from the neighborhood theory [Hauptman (1975). *Acta Cryst.* **A31**, 680–687]. A procedure for estimating the one-phase structure seminvariants near 0 or π (OPSSs) is proposed, based on the TPSS estimates. Test results for known structures with error-free data show accurate estimates for the TPSSs, even in the macromolecular case. For the OPSSs, the accuracy is related to the strength of anomalous scattering by heavy atoms in the macromolecular case. The incorporation of the heavy-atom information improves the results.

1. Introduction

In about the past ten years, the combination of traditional direct methods with anomalous dispersion has been extensively studied to improve the

methods for phase determination of macromolecular structures. Hauptman (1982b) proposed the probabilistic theory of the two- and three-phase structure invariants by integrating anomalous-dispersion effects into the neighborhood concept. Giacovazzo (1983) also obtained a similar result by a different route. Fortier, Fraser & Moore (1986) later re-examined the theoretical basis for this approach and obtained more-accurate three-phase-invariant estimates by the use of anomalous-scatterer substructure information. Subsequently, Hao & Fan (1988) presented a method for the individual phase estimates by the incorporation of heavy-atom information into Hauptman's distribution.

The estimate of the two-phase structure invariants (TPSIs) arose from the joint probability distribution of the Friedel pair $E_{\mathbf{H}}$ and $E_{\bar{\mathbf{H}}}$ in the presence of anomalous scattering. The practical application of this result has been suggested by Cascarano & Giacovazzo (1984). Further studies of the TPSIs can be found in other publications (Guo & Hauptman, 1989; Guo, 1990; Guo, Blessing & Hauptman, 1991).

In research into structure seminvariants, Velmurugan & Hauptman (1989) derived the conditional probability distribution for the OPSSs having

Table 1. The types of \bar{H}_i and n values for the monoclinic and orthorhombic systems

H-K group	\bar{H}_i	Space group and n										
Monoclinic system												
$(h,k,l)-P(2,0,2)$	$h \bar{k} l$	$P2$	$P2_1$									
		0	k									
$(h,k,l)-P(0,2,0)$	$\bar{h} k \bar{l}$	Pm	Pc									
		0	l									
Orthorhombic system												
		$P222$	$P222_1$	$P2_12_12$	$P2_12_12_1$							
$(h,k,l)-P(2,2,2)$	$\left\{ \begin{array}{l} \bar{h} k l \\ h \bar{k} \bar{l} \\ h k \bar{l} \end{array} \right.$	0	0	$h+k$	$h+k$							
		0	l	$h+k$	$k+l$							
		0	l	0	$l+h$							
$(h,k,l)-P(2,2,0)$	$\left\{ \begin{array}{l} h k \bar{l} \\ h \bar{k} \bar{l} \\ \bar{h} k \bar{l} \end{array} \right.$	$Pmm2$	$Pmc2_1$	$Pcc2$	$Pma2$	$Pca2_1$	$Pnc2$	$Pmn2_1$	$Pba2$	$Pna2_1$	$Pnn2$	
		0	l	0	0	l	0	$l+h$	0	l	0	
		0	0	l	h	$l+h$	$k+l$	0	$h+k$	$h+k+l$	$h+k+l$	
		0	l	l	h	h	$k+l$	$l+h$	$h+k$	$h+k$	$h+k+l$	

values near 0 or π in the presence of anomalous scattering by embedding the OPSSs into the three-phase structure invariants and gave the results of its applications (Velmurugan, Hauptman & Potter, 1989). These works clearly indicate that the fusion of direct methods with anomalous dispersion will facilitate the solution of those crystal structures that contain one or more anomalous scatterers. In this paper, the estimate for the TPSIs developed by Hauptman is extended to the TPSSs and, further, the OPSS (near 0 or π) estimates are realized by combining the formulas for the TPSSs with Cochran's distribution for the triplet phase relationship (Cochran, 1955). The estimate formulas for both the TPSSs and OPSSs were tested with error-free diffraction data.

2. Theoretical basis

2.1. The formulas for the TPSS estimate

According to Hauptman (1982b), the probabilistic estimate of the TPSI for a crystal with anomalous scatterers at one wavelength is

$$\varphi_{\mathbf{H}} + \varphi_{\bar{\mathbf{H}}} = -\xi. \quad (1)$$

The quantity ξ is defined by

$$\tan \xi = -S_{\mathbf{H}}/C_{\mathbf{H}} \quad (2)$$

with

$$C_{\mathbf{H}} = (1/\alpha_{\mathbf{H}}) \sum_{j=1}^N |f_{j\mathbf{H}}|^2 \cos 2\delta_{j\mathbf{H}}, \quad (3)$$

$$S_{\mathbf{H}} = (1/\alpha_{\mathbf{H}}) \sum_{j=1}^N |f_{j\mathbf{H}}|^2 \sin 2\delta_{j\mathbf{H}}, \quad (4)$$

$$\alpha_{\mathbf{H}} = \sum_{j=1}^N |f_{j\mathbf{H}}|^2, \quad (5)$$

where N is the number of atoms in the unit cell and the atomic scattering factor, $f_{j\mathbf{H}} = f_{j\mathbf{H}}^0 + f_j' + if_j''$, is expressed as

$$f_{j\mathbf{H}} = |f_{j\mathbf{H}}| \exp(i\delta_{j\mathbf{H}}). \quad (6)$$

The estimate (1) is reliable in the case that the variance of the distribution is small. Now, assume that a crystal belongs to a given noncentrosymmetric space group with m equivalent positions. The relationship between the phases of equivalent reflections is known (Waser, 1955):

$$\varphi_{\mathbf{H}} = \varphi_{\mathbf{H}_i} + 2\pi\mathbf{H} \cdot \mathbf{T}_i \quad (i = 1, 2, \dots, m), \quad (7)$$

where \mathbf{H}_i is the i th equivalent vector of \mathbf{H} in reciprocal space and \mathbf{T}_i is the translation vector of the i th equivalent position in direct space. With (7), (1) can be written as

$$\varphi_{\mathbf{H}} + \varphi_{\bar{\mathbf{H}}} = n\pi - \xi, \quad (8)$$

where

$$n = 2\mathbf{H} \cdot \mathbf{T}_i \quad (i = 1, 2, \dots, m). \quad (9)$$

Since $\varphi_{\mathbf{H}} + \varphi_{\bar{\mathbf{H}}}$ in (1), which corresponds to $i = 1$ in (8), is a TPSI, $\varphi_{\mathbf{H}} + \varphi_{\bar{\mathbf{H}}_i}$ ($i = 2, \dots, m$) in (8), derived from the equivalent transformation of $\varphi_{\bar{\mathbf{H}}}$, must be a TPSS for the given space group. When i changes from 2 to m , we can obtain different types of TPSS of this space group. Clearly, (8) and (9), the estimate formulas for the TPSSs are suitable for all noncentrosymmetric space groups.

It is well known that the space groups with the same structure seminvariants belong to the same H-K (Hauptman-Karle) group. For each H-K group, the \bar{H}_i 's can be divided into 1-4 sets (only one set for the monoclinic and orthorhombic systems) according to the equivalent positions. There are $m-1$ types of \bar{H}_i in each set. Table 1 lists the types of \bar{H}_i and n values for the monoclinic and orthorhombic systems.

2.2. The probabilistic background of the TPSS estimate

The formulas (8) and (9) are derived from the joint probability distribution followed by the conditional probability distribution of the TPSS according to Hauptman's (1975) neighborhood theory.

Assume again that the number of equivalent positions is m for a given space group. Then, as described above, $\psi = \varphi_{\mathbf{H}} + \varphi_{\overline{\mathbf{H}}_k}$ is a TPSS involving the k th equivalent vector of $\overline{\mathbf{H}}$ and its first neighborhood is defined to consist of the two magnitudes $|E_{\mathbf{H}}|$ and $|E_{\overline{\mathbf{H}}}|$ because $|E_{\overline{\mathbf{H}}}| = |E_{\overline{\mathbf{H}}_k}|$. In the presence of anomalous scatterers, the normalized structure factor is defined by

$$E_{\mathbf{H}} = (1/\alpha_{\mathbf{H}}^{1/2}) \sum_{j=1}^N |f_{j\mathbf{H}}| \exp [i(\delta_{j\mathbf{H}} + 2\pi\mathbf{H} \cdot \mathbf{r}_j)]. \quad (10)$$

Replacing \mathbf{H} by $\overline{\mathbf{H}}_k$ in (10) and noticing that $f_{j\mathbf{H}} = f_{j\overline{\mathbf{H}}_k}$, we have

$$E_{\overline{\mathbf{H}}_k} = (1/\alpha_{\overline{\mathbf{H}}_k}^{1/2}) \sum_{j=1}^N |f_{j\mathbf{H}}| \exp [i(\delta_{j\mathbf{H}} - 2\pi\mathbf{H}_k \cdot \mathbf{r}_j)]. \quad (11)$$

Suppose that the reciprocal-lattice vector \mathbf{H} is fixed; the atomic position vectors \mathbf{r}_j are assumed to be the primitive random variables, which are uniformly distributed in the unit cell. Then, $E_{\mathbf{H}}$ and $E_{\overline{\mathbf{H}}_k}$, as functions of the primitive random variables, are themselves random variables. Let

$$R_1 = |E_{\mathbf{H}}|, \quad \phi_1 = \varphi_{\mathbf{H}}, \quad R_2 = |E_{\overline{\mathbf{H}}_k}|, \quad \phi_2 = \varphi_{\overline{\mathbf{H}}_k};$$

the joint probability distribution of the magnitudes $|E_{\mathbf{H}}|$, $|E_{\overline{\mathbf{H}}_k}|$ and the phases $\varphi_{\mathbf{H}}$, $\varphi_{\overline{\mathbf{H}}_k}$ of the structure factors $E_{\mathbf{H}}$, $E_{\overline{\mathbf{H}}_k}$ is given by the fourfold integral (Karle & Hauptman, 1958)

$$P(R_1, R_2; \phi_1, \phi_2) = [R_1 R_2 / (2\pi)^4] \int_{\rho_1, \rho_2=0}^{\infty} \int_{\theta_1, \theta_2=0}^{2\pi} \rho_1 \rho_2 \times \exp \{ -i[R_1 \rho_1 \cos(\theta_1 - \phi_1) + R_2 \rho_2 \cos(\theta_2 - \phi_2)] \} \times \prod_{j=1}^{N/m} q_j(\rho_1, \rho_2; \theta_1, \theta_2) d\rho_1 d\rho_2 d\theta_1 d\theta_2, \quad (12)$$

where

$$q_j(\rho_1, \rho_2; \theta_1, \theta_2) = \langle \exp \{ (i|f_j|/\alpha^{1/2}) [\rho_1 \cos(\delta_j + 2\pi\mathbf{H} \cdot \mathbf{r}_j - \theta_1) + \rho_1 \cos(\delta_j + 2\pi\mathbf{H} \cdot \mathbf{R}_2 \cdot \mathbf{r}_j + 2\pi\mathbf{H} \cdot \mathbf{T}_2 - \theta_1) + \dots + \rho_1 \cos(\delta_j + 2\pi\mathbf{H} \cdot \mathbf{R}_m \cdot \mathbf{r}_j + 2\pi\mathbf{H} \cdot \mathbf{T}_m - \theta_1) + \rho_2 \cos(\delta_j - 2\pi\mathbf{H}_k \cdot \mathbf{r}_j - \theta_2) + \rho_2 \cos(\delta_j - 2\pi\mathbf{H}_k \cdot \mathbf{R}_2 \cdot \mathbf{r}_j - 2\pi\mathbf{H}_k \cdot \mathbf{T}_2 - \theta_2) + \dots + \rho_2 \cos(\delta_j - 2\pi\mathbf{H}_k \cdot \mathbf{R}_m \cdot \mathbf{r}_j - 2\pi\mathbf{H}_k \cdot \mathbf{T}_m - \theta_2)] \} \rangle_{r_j}, \quad (13)$$

From the work of Hauptman (1982a),

$$q_j(\rho_1, \rho_2; \theta_1, \theta_2) = 1 - \{ (|f_j|^2/\alpha) [(m/4)(\rho_1^2 + \rho_2^2) + (m/2)\rho_1\rho_2 \cos(2\delta_j + 2\pi\mathbf{H} \cdot \mathbf{T}_k - \theta_1 - \theta_2)] \}$$

and

$$\prod_{j=1}^{N/m} q_j(\rho_1, \rho_2; \theta_1, \theta_2) = \exp [-\frac{1}{4}(\rho_1^2 + \rho_2^2)] \exp [-\frac{1}{2}\rho_1\rho_2 X \times \cos(\xi + 2\pi\mathbf{H} \cdot \mathbf{T}_k - \theta_1 - \theta_2)], \quad (14)$$

where

$$X = (C_{\mathbf{H}}^2 + S_{\mathbf{H}}^2)^{1/2} \quad (15)$$

and ξ , $C_{\mathbf{H}}$ and $S_{\mathbf{H}}$ are defined by (2), (3) and (4).

Substitution of (14) into (12) and completion of the fourfold integral give

$$P(R_1, R_2; \phi_1, \phi_2) = [R_1 R_2 / \pi^2 (1 - X^2)] \times \exp \{ -[(R_1^2 + R_2^2)/(1 - X^2)] + [2XR_1 R_2 / (1 - X^2)] \times \cos(\phi_1 + \phi_2 - 2\pi\mathbf{H} \cdot \mathbf{T}_k + \xi) \}. \quad (16)$$

From (16), when R_1 and R_2 are given, the conditional probability distribution of $\psi = \varphi_{\mathbf{H}} + \varphi_{\overline{\mathbf{H}}_k}$ is

$$P(\psi/R_1, R_2) = [2\pi I_0(A)]^{-1} \exp [A \cos(\psi - n\pi + \xi)], \quad (17)$$

where

$$n = 2\mathbf{H} \cdot \mathbf{T}_k, \quad (18)$$

$$A = 2XR_1 R_2 / (1 - X^2) \quad (19)$$

and I_0 is the modified Bessel function. Since (17) has a unique maximum at $\psi = n\pi - \xi$, it follows that

$$\varphi_{\mathbf{H}} + \varphi_{\overline{\mathbf{H}}_k} \approx n\pi - \xi \quad (20)$$

provided that the variance of the distribution (17) is small, *i.e.* A is large. Equations (20) and (18) just correspond to (8) and (9) when $i = k$, suitable for all noncentrosymmetric space groups.

2.3. The estimates of the OPSSs

According to Cochran (1955), the conditional probability distribution of $\varphi_{\mathbf{H}}$ for a set of triplet structure invariants of the form $\varphi_{\mathbf{H}} - \varphi_{\mathbf{H}'} - \varphi_{\mathbf{H}-\mathbf{H}'}$ given $|E_{\mathbf{H}}|$, $E_{\mathbf{H}'}$ and $E_{\mathbf{H}-\mathbf{H}'}$ is

$$P(\varphi_{\mathbf{H}}) = [2\pi I_0(\kappa)]^{-1} \times \exp \left[\sum_{\mathbf{H}'} \kappa \cos(\varphi_{\mathbf{H}} - \varphi_{\mathbf{H}'} - \varphi_{\mathbf{H}-\mathbf{H}'}) \right], \quad (21)$$

where

$$\kappa = 2\sigma_3 \sigma_2^{-3/2} |E_{\mathbf{H}} E_{\mathbf{H}'} E_{\mathbf{H}-\mathbf{H}'}|,$$

$$\sigma_n = \sum_{j=1}^N Z_j^n$$

and Z_j is the atomic number of the j th atom in the unit cell.

Suppose now that $\varphi_{\mathbf{H}_i}$ is an OPSS and $\varphi_{\mathbf{H}} + \varphi_{\overline{\mathbf{H}}_i}$ a TPSS, which are related to each other by

$$\mathbf{H}_i = \mathbf{H} + \overline{\mathbf{H}}_i \quad (i = 2, \dots, m). \quad (22)$$

In (22), \mathbf{H} corresponds to the index hkl and thus the index type of \mathbf{H}_i depends upon that of $\overline{\mathbf{H}}_i$, which is given in Table 1. If $|E_{\mathbf{H}_i}|$, $|E_{\mathbf{H}}|$, $|E_{\overline{\mathbf{H}}_i}|$ and $\varphi_{\mathbf{H}} + \varphi_{\overline{\mathbf{H}}_i}$ are given, using (21) we have

$$P(\varphi_{\mathbf{H}_i}) = [2\pi I_o(K)]^{-1} \times \exp \left[\sum_p K \cos(\varphi_{\mathbf{H}_i} - \varphi_{\mathbf{H}} - \varphi_{\overline{\mathbf{H}}_i}) \right], \quad (23)$$

where

$$K = 2\sigma_3\sigma_2^{-3/2}|E_{\mathbf{H}_i}E_{\mathbf{H}}E_{\overline{\mathbf{H}}_i}|. \quad (24)$$

The summation over p involves those pairs of reflections (\mathbf{H} , $\overline{\mathbf{H}}_i$) that satisfy (22) and have larger $|E|$ values.

In addition, following Hauptman (1982b), we obtained the probability distribution of one structure factor $E_{\mathbf{H}_i}$, whose phase $\varphi_{\mathbf{H}_i}$ is a structure seminvariant with a value near 0 or π , when anomalous scatterers are present (see Appendix). The conditional probability distribution of $\varphi_{\mathbf{H}_i}$ assuming as known the magnitude $|E_{\mathbf{H}_i}|$ has a unique maximum at $2\varphi_{\mathbf{H}_i} = -\xi$. This result implies that for the single phase the increment of the phase arising from anomalous scattering is $-\xi/2$. Hence, when one deals with the OPSSs by making use of (8), a modified form should be employed:

$$\varphi_{\mathbf{H}} + \varphi_{\overline{\mathbf{H}}_i} \approx n\pi - \xi/2. \quad (25)$$

With $\alpha \cos \beta = \sum_p K \cos(\varphi_{\mathbf{H}} + \varphi_{\overline{\mathbf{H}}_i})$, $\alpha \sin \beta = \sum_p K \sin(\varphi_{\mathbf{H}} + \varphi_{\overline{\mathbf{H}}_i})$ and combination with (25), (23) becomes

$$P(\varphi_{\mathbf{H}_i}) = [2\pi I_o(\alpha)]^{-1} \exp[\alpha \cos(\varphi_{\mathbf{H}_i} - \beta)], \quad (26)$$

where

$$\alpha = \left\{ \left[\sum_p K \cos(n\pi - \xi/2) \right]^2 + \left[\sum_p K \sin(n\pi - \xi/2) \right]^2 \right\}^{1/2}, \quad (27)$$

$$\tan \beta = \left[\sum_p |E_{\mathbf{H}}E_{\overline{\mathbf{H}}_i}| \sin(n\pi - \xi/2) \right] \times \left[\sum_p |E_{\mathbf{H}}E_{\overline{\mathbf{H}}_i}| \cos(n\pi - \xi/2) \right]^{-1}. \quad (28)$$

Provided that the variance of the distribution (26) is small, *i.e.* α is large, (26) has a unique maximum at $\varphi_{\mathbf{H}_i} = \beta$. The larger α , the more reliable is the $\varphi_{\mathbf{H}_i}$ estimated by (28). Since the sine part of (28) is generally small, the estimated value of $\varphi_{\mathbf{H}_i}$ is determined by the sign of the cosine part to be near 0 or π .

Table 2. Average magnitudes of the errors in the estimated TPSS values

The 2000 for cocaine methiodide (1) and 1931 for rubredoxin (2) are divided into groups, each contains 200 data but the last for rubredoxin containing 331 data.

Group no.	$\langle E \rangle$		$\langle A \rangle$		$\langle \varphi_i - \varphi_e \rangle$ (°)	
	(1)	(2)	(1)	(2)	(1)	(2)
1	1.79	1.81	5789	224	0.5	1.1
2	1.44	1.36	3591	86	0.9	1.7
3	1.24	1.15	2581	58	1.2	1.9
4	1.07	1.01	1876	39	1.3	2.2
5	0.91	0.87	1397	29	1.4	2.7
6	0.80	0.79	1041	21	1.5	3.2
7	0.68	0.65	745	15	1.7	3.9
8	0.55	0.53	506	9	2.3	5.0
9	0.40	0.32	269	4	2.9	9.6
10	0.26		115		5.5	

If the positions of heavy atoms are available, the known structure information can be utilized to improve the estimates (Fan & Gu, 1985). From the product of (26) and Sim's (1959) distribution, (27) and (28) become

$$\alpha = \left\{ \left[\sum_p K \cos(n\pi - \xi/2) + x \cos \varphi_{\mathbf{H}_i, \varrho} \right]^2 + \left[\sum_p K \sin(n\pi - \xi/2) + x \sin \varphi_{\mathbf{H}_i, \varrho} \right]^2 \right\}^{1/2}, \quad (29)$$

$$\tan \beta = \left[\sum_p K \sin(n\pi - \xi/2) + x \sin \varphi_{\mathbf{H}_i, \varrho} \right] \times \left[\sum_p K \cos(n\pi - \xi/2) + x \cos \varphi_{\mathbf{H}_i, \varrho} \right]^{-1}, \quad (30)$$

where

$$x = 2\sigma_2 \left(\sum_u Z_u^2 \right)^{-1} |E_{\mathbf{H}_i} E_{\mathbf{H}_i, \varrho}|,$$

$E_{\mathbf{H}_i, \varrho}$ is the heavy-atom contribution to $E_{\mathbf{H}_i}$ and the corresponding phase is $\varphi_{\mathbf{H}_i, \varrho}$, and Z_u is the atomic number of the u th unknown atom in the unit cell.

3. Applications

The formulas for the TPSS and OPSS estimates were tested with error-free diffraction data. The normalized structure-factor magnitudes $|E_{\mathbf{H}}|$ and the true phases $\varphi_{\mathbf{H}}$ were calculated using the known atomic coordinates from a small molecule, cocaine methiodide (Shen, Ruble & Hite, 1975), $\text{C}_{18}\text{H}_{24}\text{NO}_4\text{I}$, space group $P2_12_12_1$, $Z = 4$ with Mo $K\alpha$ wavelength up to $2\theta = 50^\circ$, and a protein, rubredoxin (Adman, Sieker, Jensen, Bruschi & LeGall, 1977), $\text{C}_{245}\text{FeN}_{58}\text{O}_{80}\text{S}_5$ (H atoms not included), space group $P2_1$, $Z = 2$ with Cu $K\alpha$ wavelength to a resolution of 1.5 Å.

3.1. Test of the TPSS estimates

For cocaine methiodide, the calculations were done for 2000 TPSSs of three types of \bar{H}_i , namely $\bar{H}_i = \bar{h}, k, l$, h, \bar{k}, l and h, k, \bar{l} with $|E| > 0.1$ [$|E| = (|E_H| |E_{\bar{H}_i}|)^{1/2}$]. The results were arranged in descending order of A values. Table 2 gives the averages $\langle |E| \rangle$, $\langle A \rangle$ and the error $\langle |\varphi_t - \varphi_e| \rangle$ (φ_t and φ_e denote true and estimated values of the TPSS, respectively) in the ten groups, each containing 200

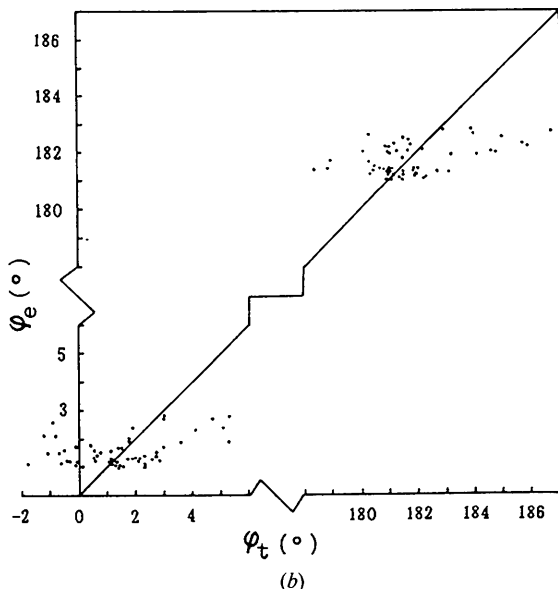
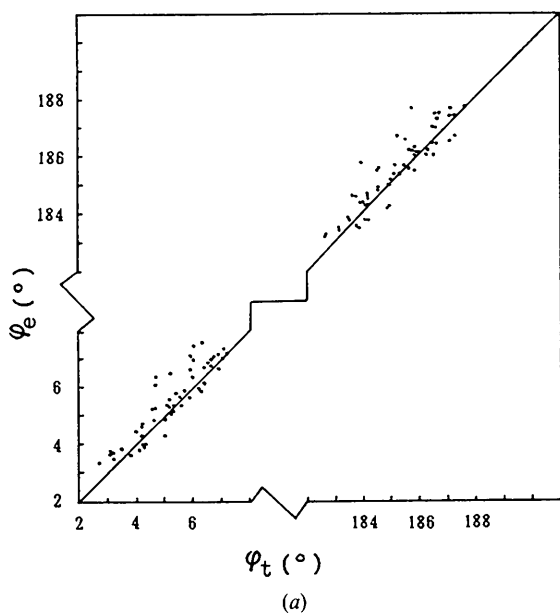


Fig. 1. Scatter diagrams of φ_e versus φ_t for the TPSS estimates, as well as the lines $\varphi_e = \varphi_t$, (a) using 116 data with $|E| > 1.7$ for cocaine methiodide, (b) using 125 data with $|E| > 1.7$ for rubredoxin.

data. A scatter diagram of φ_e versus φ_t is shown in Fig. 1(a).

1931 TPSSs of $\bar{H}_i = h, \bar{k}, l$ type with $|E| > 0.1$ were computed for rubredoxin. Table 2 also lists the averages for the nine groups in descending order of A values. Fig. 1(b) shows a scatter diagram of φ_e versus φ_t .

The computational results show that, firstly, a large number of estimates of the TPSSs can be obtained by (8) and (9) with quite a good accuracy in a statistical sense. Secondly, the larger the A values, i.e. the smaller the variance of the distribution (17), the closer the agreements between φ_e and φ_t . The errors $\langle |\varphi_t - \varphi_e| \rangle$ decrease also with increasing $|E|$, as is shown in Fig. 2. Fig. 1 shows that for the large $|E|$ values the diagonal distribution holds, even in the protein case. Finally, owing to the relative weakness of the anomalous-scattering intensities, the errors are larger for the protein structure than for the small-molecule structure. From Fig. 2, we see that the average errors begin to increase substantially for $|E| < 0.5$ for cocaine methiodide but for $|E| < 0.8$ for rubredoxin.

3.2. Test of the OPSS estimates

In the space group $P2_12_12_1$, the φ_{H_i} is of three forms: $\varphi_{0,2k,2l}$, $\varphi_{2h,0,2l}$, $\varphi_{2h,2k,0}$. The results of the three kinds of phase were arranged in descending order of α values for cocaine methiodide. Table 3 was constructed using the top 121 with $\alpha \geq 0.5$ by accumulation into the four groups shown. The seventh column (Nwr) shows the number of signs of $\cos \varphi_H$, incorrectly estimated.

In the space group $P2_1$, there is only one kind of φ_{H_i} , $\varphi_{2h,0,2l}$. In order to examine the influence of the strength of the anomalous signal on the estimates, the three derivative structures were made from rubredoxin by replacing the Fe atom by Cd, Pt and U

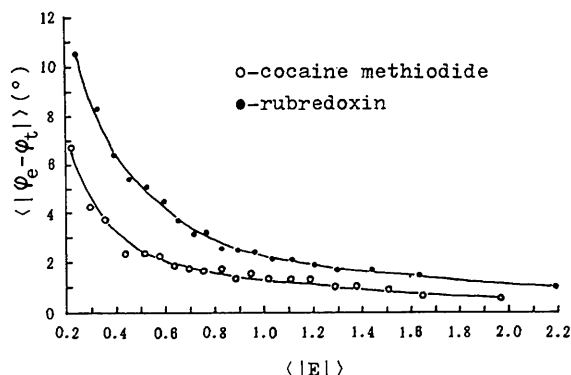


Fig. 2. The error $\langle |\varphi_t - \varphi_e| \rangle$ distribution of the TPSS estimates as a function of $\langle |E| \rangle$. The data were ranked on $|E|$ and the averages $\langle |E| \rangle$ and $\langle |\varphi_t - \varphi_e| \rangle$ were computed for the ranked groups of 100 data.

Table 3. Average magnitudes of the errors in the 121 seminvariants accumulated in groups according to α_{min} for cocaine methiodide

Group no.	Number in group	α_{min}	$\langle \alpha \rangle$	$\langle E_{H_s} \rangle$	$\langle \varphi_{H_s} - \beta \rangle$ (°)	Nwr
1	21	10.0	18.6	1.57	0.5	0
2	46	5.0	12.2	1.37	0.7	0
3	91	1.5	7.7	1.22	0.7	0
4	121	0.5	6.0	1.08	8.3	5

atoms, respectively. The results were arranged in descending order of α values and those for the top 30 are listed in Table 4. The last column (Per) gives the percentage of the seminvariants with the sign of $\cos \varphi_{H_s}$ correctly determined.

The results show that, as expected, the larger the α values, the more reliable are the phase estimates. Comparison of Table 3 with Table 4 indicates that the results are much better in the small-molecule case than in the macromolecular case. For cocaine methiodide, all the signs of the cosines are correctly determined when $\alpha \geq 1.5$. It can be seen from Table 4 that the results are relatively poor for rubredoxin containing one Fe atom but the estimates are gradually improved when the heavy atom changes from iron to uranium. The average value of α also increases with the replacement of the heavier atom. For a given α value, there are far more seminvariants reliably estimated in the case of the U atom as major anomalous scatterer than there are in the case of the Fe atom. The results calculated from (29) and (30) are also given in Table 4 by the entries for method II, which show that the incorporation of the heavy-atom structure information leads to an effective increase in the number of seminvariants whose values are correctly estimated.

4. Concluding remarks

The extension of the probabilistic estimates of the TPSIs from Hauptman's (1982b) formula to the TPSS estimates has been proved to be valid by applications to the known structures using error-free data. In the case that the variance of the distribution is small, which corresponds to the large $|E|$ value involved, the estimate is reliable. The analytical features are consistent with those of the TPSI estimates (Guo, Blessing & Hauptman, 1991).

The estimate formulas for the OPSSs (near 0 or π) when anomalous scatterers are present, assuming as known the magnitudes $|E_{H_s}|$, $|E_{H_t}|$, $|E_{H_r}|$ and the TPSS $\varphi_H + \varphi_{\bar{H}}$, have also been tested with the error-free data. The accuracy of the estimates depends on the complexity of the structure and the strength of the anomalous scattering. Quite accurate estimates can be obtained from smaller structures. The fraction of

Table 4. Average magnitudes of the errors in the top 30 seminvariants arranged in descending order of α for rubredoxin and its Cd-, Pt- and U-derivative structures

Method I gives the results calculated from (27) and (28) and method II those from (29) and (30).

Method	Heavy atom	$\langle \alpha \rangle$	$\langle E_{H_s} \rangle$	$\langle \varphi_{H_s} - \beta \rangle$ (°)	Nwr	Per (%)
I	Fe	0.9	1.27	60.3	10	66.7
	Cd	1.4	1.40	43.1	7	76.7
	Pt	3.4	1.37	31.4	5	83.3
	U	4.9	1.36	15.5	2	93.3
II	Fe	3.3	1.24	49.9	8	73.3
	Cd	4.3	1.33	31.3	5	83.3
	Pt	7.2	1.33	7.7	1	96.7
	U	9.3	1.36	8.5	1	96.7

seminvariants correctly estimated may reach more than 90% for macromolecular structures containing as many as 400 non-H atoms with such a major anomalous scatterer as a U atom in the asymmetric unit. The improvement of the accuracy can be achieved by the incorporation of the heavy-atom-structure information. These results suggest that the method for the TPSS estimates described here as a supplement to existing techniques may find application in the determination of real crystal structures. An analogous study on the three-phase structure seminvariants is in progress.

APPENDIX

The reciprocal-lattice vector H_s is fixed and the atomic position vectors r_j are assumed to be the primitive random variables. Following the work of Hauptman (1982b), the joint probability distribution of the magnitude $|E_{H_s}|$ and the phase φ_{H_s} of the structure factor E_{H_s} , where φ_{H_s} is a structure seminvariant having a value near 0 or π , when anomalous scatterers are present, is given by the twofold integral

$$P(R, \phi) = [R/(2\pi)^2] \int_{\rho=0}^{\infty} \int_{\theta=0}^{2\pi} \rho \exp[-iR\rho \cos(\theta - \phi)] \times \prod_{j=1}^{N/m} q_j(\rho, \theta) d\rho d\theta, \quad (A1)$$

where

$$q_j(\rho, \theta) = \left\langle (i|f_j|/\alpha^{1/2}) \rho \sum_{i=1}^m \cos(\delta_j + 2\pi H_s \cdot C_i \cdot r_j - \theta) \right\rangle_r \\ \approx -m\rho^2(|f_j|^2/4\alpha)[1 + \cos(2\delta_j - 2\theta)]. \quad (A2)$$

Substituting (A2) into (A1) and completing the twofold integral, we have

$$P(R, \phi) = (R/\pi) \exp[-R^2 + XR^2 \cos(2\phi + \xi)]. \quad (A3)$$

When $R = |E_{H_s}|$ is given, the conditional probability

distribution of φ_H is

$$P(\phi/R) = [2\pi I_0(XR^2)]^{-1} \exp[XR^2 \cos(2\phi + \xi)], \quad (A4)$$

where ξ is defined in (2)–(5) and X in (15). Clearly, (A4) has a unique maximum at $2\varphi_H = -\xi$ when XR^2 is large.

References

- ADMAN, E. T., SIEKER, L. C., JENSEN, L. H., BRUSCHI, M. & LEGALL, J. (1977). *J. Mol. Biol.* **112**, 113–120.
- CASCARANO, G. & GIACOVAZZO, C. (1984). *Acta Cryst.* **A40**, 305–306.
- COCHRAN, W. (1955). *Acta Cryst.* **8**, 473–478.
- FAN, H.-F. & GU, Y.-X. (1985). *Acta Cryst.* **A41**, 280–284.
- FORTIER, S., FRASER, M. E. & MOORE, N. J. (1986). *Acta Cryst.* **A42**, 149–156.
- GIACOVAZZO, C. (1983). *Acta Cryst.* **A39**, 585–592.
- GUO, D. Y. (1990). *Acta Cryst.* **A46**, 942–944.
- GUO, D. Y., BLESSING, R. & HAUPTMAN, H. (1991). *Acta Cryst.* **A47**, 340–345.
- GUO, D. Y. & HAUPTMAN, H. (1989). *Chin. Sci. Bull.* **34**, 137–141.
- HAO, Q. & FAN, H.-F. (1988). *Acta Cryst.* **A44**, 379–382.
- HAUPTMAN, H. (1975). *Acta Cryst.* **A31**, 680–687.
- HAUPTMAN, H. (1982a). *Acta Cryst.* **A38**, 289–294.
- HAUPTMAN, H. (1982b). *Acta Cryst.* **A38**, 632–641.
- KARLE, J. & HAUPTMAN, H. (1958). *Acta Cryst.* **11**, 264–269.
- SHEN, M., RUBLE, J. R. & HITE, G. (1975). *Acta Cryst.* **B31**, 2706–2709.
- SIM, G. A. (1959). *Acta Cryst.* **12**, 813–815.
- VELMURUGAN, D. & HAUPTMAN, H. (1989). *Acta Cryst.* **A45**, 158–163.
- VELMURUGAN, D., HAUPTMAN, H. & POTTER, S. A. (1989). *Acta Cryst.* **A45**, 163–165.
- WASER, J. (1955). *Acta Cryst.* **8**, 595.

Acta Cryst. (1994). **A50**, 601–605

The Use of Continuous Atomic Distributions in Structural Investigations

BY VLADIMIR V. CHERNYSHEV, SERGEI G. ZHUKOV, ALEXANDER V. YATSENKO AND LEONID A. ASLANOV

Department of Chemistry, Moscow State University, 119899 Moscow, Russia

AND HENK SCHENK

Laboratory of Crystallography, University of Amsterdam, Nieuwe Achtergracht 166, Amsterdam 1018 WV, The Netherlands

(Received 1 November 1993; accepted 17 March 1994)

Abstract

A simple method of describing the scattering from an atom continuously distributed in the unit cell is proposed. Some experimental applications in structural investigations are presented.

1. Introduction

It is not unusual in structural analysis that an atom does not occupy a definite point position in the unit cell but is distributed among a number of positions, discretely or continuously. The problems connected with atomic disorder in crystals have been discussed extensively for a long time (Krivoglaz, 1969; Dunitz, Schomaker & Trueblood, 1988; Kuhs, 1992). If the number of occupied positions is finite, it is possible to refine these positions with partial populations. For more complicated distributions, caused by statistical displacements or thermal vibrations, a general approach has been introduced by Johnson (1969). It is based on a differential expansion of the atomic Gaussian probability density function (p.d.f.), which after Fourier transformation leads to a series expansion

of tensorial coefficients. The limitations of this approach were discussed by Kuhs (1992). However, there are many cases when one wishes to express the p.d.f. for disordered atoms in real space. This p.d.f. will not only be similar to the Gaussian or Johnson expansions but, also, its type may be established on the basis of physical considerations. Sometimes, it is possible to parametrize such a p.d.f. with a small number of parameters and use these as variables in the refinement. This direct approach to define the p.d.f. of disordered atoms was applied in the structural investigations of two compounds. Preliminary results were published in brief (Zhukov, 1991; Chernyshev, 1992). In this paper, a more detailed description is presented.

2. An atom uniformly distributed on a sphere

Let us define the average unit cell (a.u.c.) as that obtained by averaging the crystal unit cells. Such a.u.c.'s are used in ordinary structure-factor calculations. Let atom A in the a.u.c. be continuously distributed with equal probability on some set of