

- FUREY, W. F., ROBBINS, A. H., CLANCY, L. L., WINGE, D. R., WANG, B. C. & STOUT, C. D. (1986). *Science*, **231**, 704-710.
- GLOVER, I., HANEEF, I., PITTS, J.-E., WOOD, S. P., MOSS, D., TICKLE, I. J. & BLUNDELL, T. L. (1983). *Biopolymers*, **22**, 293-304.
- HAUPTMAN, H. (1982). *Acta Cryst.* **A38**, 289-294.
- HULL, S. E. & IRWIN, M. J. (1978). *Acta Cryst.* **A34**, 863-870.
- KARLE, J. (1968). *Acta Cryst.* **B24**, 182-186.
- KARLE, J. (1984). *Methods and Applications in Crystallographic Computing*, edited by S. R. HALL & T. ASHIDA, pp. 120-140. Oxford: Clarendon Press.
- KARLE, J. (1986). *Acta Cryst.* **A42**, 246-253.
- MAIN, P. (1978). *Acta Cryst.* **A34**, 31-38.
- SAYRE, D. (1972). *Acta Cryst.* **A28**, 210-212.
- SAYRE, D. (1974). *Acta Cryst.* **A30**, 180-184.
- TICKLE, I. J. (1981). *Refinement of Protein Structures*. Proceedings of the Daresbury Study Weekend, 15-16 November 1980, edited by P. A. MACHIN, J. W. CAMPBELL & M. ELDER. Warrington: SERC Daresbury Laboratory.
- TSOUCARIS, G. (1970). *Acta Cryst.* **A26**, 492-499.
- WANG, B. C. (1984). *Acta Cryst.* **A40**, C12.
- WILSON, K. S. (1978). *Acta Cryst.* **B34**, 1599-1608.
- YAO JIA-XING (1983). *Acta Cryst.* **A39**, 35-37.

Acta Cryst. (1988). **A44**, 413-417

Exact Random-Walk Models in Crystallographic Statistics. IV. P.D.F.'s of $|E|$ Allowing for Atoms in Special Positions

BY URI SHMUELI

School of Chemistry, Tel Aviv University, Ramat Aviv, 69 978, Tel Aviv, Israel

AND GEORGE H. WEISS

Division of Computer Research and Technology, Laboratory of Physical Sciences, National Institutes of Health, Bethesda, Maryland 20892, USA

(Received 7 January 1987; accepted 17 September 1987)

Abstract

The effect of scatterers, located in variable special positions, on the probability density function of the magnitude of the normalized structure factor has been investigated. Exact characteristic functions have been obtained for all the statistically different variable special positions in triclinic, monoclinic and orthorhombic space groups except in *Fdd2* and in the space groups based on the point group 222, and the probability density functions have been evaluated from their Fourier or Fourier-Bessel series expansions. It is seen that the effect of heavy scatterers, located in the special positions investigated, is very marked and should be accounted for in cases of space-group ambiguities.

Introduction

The effects of the presence of atoms in special positions have been investigated in earlier work on intensity statistics (Karle & Hauptman, 1953; Hauptman & Karle, 1953; Collin, 1955; Hargreaves, 1956; Sim, 1958; Foster & Hargreaves, 1963; Ilyukhin & Nikitin, 1963), but no exact studies of these effects on the probability density function (p.d.f.) of the structure factor have so far been attempted. One of the reasons for avoiding the study of special positions is their very large number in all the 230 space groups,

and consequent apparent difficulties in arriving at tractable and reasonably concise formulae. Since, however, the qualitative effects of (heavy) scatterers in special positions on intensity statistics may well be of considerable significance in the determination of space-group symmetry in cases of ambiguities, a study of such effects was thought to be of interest.

The techniques used in this paper are similar to those we employed in our previous studies of intensity statistics, based on exact solutions of random-walk models (e.g. Shmueli, Weiss, Kiefer & Wilson, 1984; Shmueli & Weiss, 1987). Only low-symmetry space groups are treated, and it is seen that the number of different expressions that need to be developed is much smaller than the formal number of crystallographically different Wyckoff positions in the space groups investigated. The present treatment is confined to the variable special positions (*i.e.* lines and planes), since the contributions of scatterers located in fixed special positions can be calculated and subsequently subtracted from the (scaled) intensity; examples of the latter process can be found in the works of Collin (1955), Sim (1958), Srinivasan & Parthasarathy (1976) and Pradhan, Ghosh & Nigam (1985). The results presented in this paper encompass all the variable special positions in monoclinic and orthorhombic space groups, except those in space groups based on the point group 222 and in the space group *Fdd2*.

Derivation

In order to obtain the required dependence of the p.d.f. on the distribution of atoms among the symmetry sites, it is best to start from a general representation of the structure factor. All the special positions are adequately represented by the following formulation of the normalized structure factor:

$$E(\mathbf{h}) = \sum_{m=1}^k a_m \sum_{j=1}^{N_m} n_j (\xi_j^m + i\eta_j^m) = A(\mathbf{h}) + iB(\mathbf{h}), \quad (1)$$

where the index m ranges over the types of Wyckoff positions, their number being k , the index j ranges over the N_m atoms in the m th Wyckoff position, a_m is the occupancy factor of the m th Wyckoff position, and $\xi_j^m + i\eta_j^m$ is the conventional complex trigonometric structure factor of the j th atom in the m th Wyckoff position. The occupancy factor, *i.e.* the fraction of an atom occupying a Wyckoff position, enables one to use trigonometric structure factors with numerical coefficients as given in standard references (*e.g. International Tables for X-ray Crystallography*, 1965). For example, the occupancy factor for an atom located on an n -fold axis is $1/n$ *etc.*, and $a_m = 1$ characterizes general positions. Dispersion is neglected, *i.e.* the normalized scattering factor n_j in (1) is taken as a real quantity.

The required probability density function is univariate for centrosymmetric space groups (with the origin at a center of symmetry), and initially bivariate for the non-centrosymmetric ones. We represent these p.d.f.'s as (single or double) Fourier series, and make use of the fact that the Fourier coefficients are just the values of the corresponding characteristic function. This function has a property which is important in the present application: the characteristic function of a sum of random variables is a product of the characteristic functions of individual variables. Thus, if the atoms are distributed among k Wyckoff positions, the characteristic function of the p.d.f. of $A(\mathbf{h})$ [or the joint p.d.f. of $A(\mathbf{h})$ and $B(\mathbf{h})$] is a product of k characteristic functions, each corresponding to one type of Wyckoff position.

The characteristic function for $E(\mathbf{h})$ in the centrosymmetric case becomes

$$C(\omega_1) = \langle \exp(i\omega_1 A) \rangle \quad (2)$$

$$= \left\langle \exp \left(i\omega_1 \sum_{m=1}^k a_m \sum_{j=1}^{N_m} n_j \xi_j^m \right) \right\rangle \quad (3)$$

$$= \left\langle \prod_{m=1}^k \exp \left(i\omega_1 a_m \sum_{j=1}^{N_m} n_j \xi_j^m \right) \right\rangle \quad (4)$$

$$= \left\langle \prod_{m=1}^k \prod_{j=1}^{N_m} \exp(i\omega_1 a_m n_j \xi_j^m) \right\rangle \quad (5)$$

$$= \prod_{m=1}^k C_m(\omega_1), \quad (6)$$

where

$$C_m(\omega_1) = \prod_{j=1}^{N_m} \langle \exp(i\omega_1 a_m n_j \xi_j^m) \rangle \quad (7)$$

is the contribution of the m th Wyckoff position (the N_m atoms occupying it) to the characteristic function $C(\omega_1)$. The assumptions underlying the present derivations are that (i) the atomic phase factors, $2\pi\mathbf{h} \cdot \mathbf{r}_j$, are random variables which are uniformly distributed in the $(0, 2\pi)$ range, and (ii) the contributions of different atoms to the structure factor are independent (*cf. Shmueli, Weiss, Kiefer & Wilson, 1984; Shmueli & Weiss, 1985*).

The non-centrosymmetric case is treated similarly:

$$C(\omega_1, \omega_2) = \langle \exp[i(\omega_1 A + \omega_2 B)] \rangle \quad (8)$$

$$= \left\langle \prod_{m=1}^k \prod_{j=1}^{N_m} \exp[ia_m(\omega_1 \xi_j^m + \omega_2 \eta_j^m)] \right\rangle \quad (9)$$

$$= \prod_{m=1}^k C_m(\omega_1, \omega_2), \quad (10)$$

where

$$C_m(\omega_1, \omega_2) = \prod_{j=1}^{N_m} \langle \exp[ia_m n_j (\omega_1 \xi_j^m + \omega_2 \eta_j^m)] \rangle \quad (11)$$

is the contribution of the m th Wyckoff position to the (acentric) characteristic function $C(\omega_1, \omega_2)$ [note that ξ and η are denoted in *International Tables for X-ray Crystallography* (1965) as A and B respectively].

The integrals implied by the averages in (7) and (11) can now be evaluated in closed form for most space groups of low symmetry (triclinic, monoclinic and orthorhombic) (*cf. Shmueli & Weiss, 1987*), and their calculation is increasingly complex for higher symmetries. We shall therefore confine the exact characterization of the statistical properties of the special positions to the tractable systems only. The point groups of the special positions that are of interest are always subgroups of the point group isomorphic to the space group considered, and the functional forms of the corresponding characteristic functions are identical to those found for some lower-symmetry space groups. Furthermore, the characteristic functions for the low-symmetry non-centrosymmetric space groups can usually be expressed as a function of $\omega = (\omega_1^2 + \omega_2^2)^{1/2}$ and the required p.d.f.'s can then be represented as single Fourier-Bessel series rather than double Fourier series (*e.g. Shmueli & Weiss, 1987*). Table 1 summarizes the various characteristic functions which are needed for the computation of the p.d.f.'s $p(|E|)$ for the above mentioned space groups. It should be pointed out that these p.d.f.'s depend on point groups rather than on space groups (*cf. Shmueli, 1982*), and therefore a tabulation of the symmorphic groups with P -type lattices is sufficient for the present purpose.

Table 1. Atomic characteristic functions for some general and special positions

The first column lists the triclinic, monoclinic and orthorhombic symmorphic space groups with *P*-type lattices which are representative of all the space groups of the above systems, except for *Fdd2* and *Fddd*. The second column lists the types of Wyckoff positions – descriptively rather than by their conventional notations – which have been considered in enumerating all the variable special positions in the above space groups. The third column contains the atomic characteristic functions to be used with equations (7) and (11).

Space group	Wyckoff position	Atomic characteristic function
<i>P1</i>	General	$J_0(\omega n_j)$
$\overline{P1}$	General	$J_0(2\omega_1 n_j)$
<i>P2</i>	General	$J_0^2(\omega n_j)$
	Twofold axis	$J_0(\omega n_j)$
<i>P2/m</i>	General	$J_0^2(2\omega_1 n_j)$
	Twofold axis or mirror plane	$J_0(2\omega_1 n_j)$
<i>P222</i>	General	(cf. Shmueli & Weiss, 1987)
	Twofold axis	$J_0(2\omega_1 n_j)$
<i>Pmm2</i>	General	$(2/\pi) \int_0^{\pi/2} J_0^2(2\omega_1 n_j \cos \theta) d\theta$
	Mirror plane	$J_0^2(\omega n_j)$
	Twofold axis	$J_0(\omega n_j)$
<i>Pmmm</i>	General	$(2/\pi) \int_0^{\pi/2} J_0^2(4\omega_1 n_j \cos \theta) d\theta$
	Mirror plane	$J_0^2(2\omega_1 n_j)$
	Twofold axis	$J_0(2\omega_1 n_j)$

The space groups based on the point group 222 pose some difficulties in the computation of the p.d.f. of $|E|$ for scatterers distributed among general as well as special positions. Problems of a similar nature have been found in our study of intensity statistics for dispersive scatterers. These are being investigated, and the results will be reported at a later date. However, there is no difficulty in correctly formulating the atomic characteristic functions for the above situations.

The above mentioned property of the characteristic function as product of constituent characteristic functions can also be applied to situations involving hypersymmetry. We examined this in the case of a partially bivalent distribution (Shmueli & Weiss, 1985) in the space group $\overline{P1}$, where the functional form of the Fourier coefficient was found to be that of a product of the Fourier coefficient for the space group $\overline{P1}$ (with no hypersymmetry) (Shmueli, Weiss, Kiefer & Wilson, 1984) and the Fourier coefficient of the purely bivalent distribution for this space group (Shmueli, Weiss & Kiefer, 1985). Other related examples are being investigated.

Results

Theoretical p.d.f.'s based on the atomic characteristic functions shown in Table 1 were compared with simulated distributions of $|E|$ in order to assess the magnitude of the effect of the heavy atoms when these are located in special positions. The p.d.f.'s were computed for all except the space group *P222*, with heavy scatterers in general as well as special positions. The

simulated distributions were constructed as described in detail by Shmueli *et al.* (1984), by replacing the atomic phase factors with computer-generated random numbers. The agreement between the theoretical and simulated distributions was found to be very good (as is usual for Fourier p.d.f.'s with good convergence properties) and we shall content ourselves with some graphical representations of theoretical p.d.f.'s.

The importance of the effect studied can often be inferred from the analytical expressions for the atomic characteristic function. For example, this function for the space group *P2*, for an atom in general position, is $J_0^2(\omega n_j)$ while the atomic characteristic function for an atom located on the twofold axis is $J_0(\omega n_j)$. Since $|J_0(x)|$ is always smaller than unity for $x > 0$ (Abramowitz & Stegun, 1972), the effect of special positions is expected to be considerable for this space group.

We illustrate the effects of the distribution of heavy scatterers among the possible types of Wyckoff positions for the space groups *P2/m* and *Pmm2*. Fig. 1(a) shows the p.d.f. of $|E|$ for a C_9U asymmetric unit, with all the atoms occupying general positions of *P2/m*, while Fig. 1(b) displays a p.d.f. of $|E|$ for the same space group but with a unit cell containing 36 C and four U atoms in general positions, and two symmetry-related U atoms located on a twofold axis. Clearly, both p.d.f.'s deviate significantly from the ideal centric Gaussian, but the function shown in Fig. 1(b) is much closer to an ideal p.d.f., because it represents a less heterogeneous composition of the model structure. It should be pointed out that had we increased the number of C atoms in the asymmetric unit from nine to 49, while keeping the distribution of the U atoms as above, the p.d.f.'s corresponding to those in Fig. 1 would nearly coincide. This dependence on the number of light atoms may, however, differ widely from one space group to another. Far

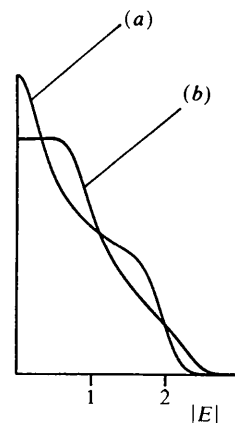


Fig. 1. Theoretical distributions for the space group *P2/m*. Both p.d.f.'s are on the same (arbitrary) scale. (a) Asymmetric unit C_9U , all the atoms in general positions. (b) As (a) with two more U atoms on a twofold axis.

more drastic effects are predicted if all the light atoms are located in general positions and the heavy atoms in special ones. Thus, the p.d.f. in Fig. 2(b) is based on 36 C atoms all in the general positions and two U atoms located on a twofold axis of $P2/m$. This p.d.f. is remarkably close to that predicted for the space group $P\bar{1}$, with all the atoms occupying general positions (*cf.* Shmueli *et al.*, 1984). Note that the same result would be obtained if the two heavy atoms were located on the mirror plane of $P2/m$. Figs. 3 and 4 illustrate similar phenomena for the space group $Pmm2$. Fig. 3(a) corresponds to a C_9U asymmetric unit with all the atoms in the general positions of $Pmm2$, Fig. 3(b) is based on the above composition plus two U atoms located on one of the mirrors of this space group, and Fig. 3(c) rests on the atomic content of Fig. 3(a) plus one U atom located on a twofold axis of the space group. The unit-cell contents are $C_{36}U_4$, $C_{36}U_5$ and $C_{36}U_6$, for the three cases, the last being least heterogeneous (Fig. 3b) and with a p.d.f. which is closest in its shape to the ideal acentric p.d.f. (Wilson, 1949). The effects of heavy atoms occupying only special positions, the light ones remaining in the general positions, are here also very pronounced. Thus Fig. 4(b) is based on two U atoms lying in a mirror plane of $Pmm2$, with the 36 C atoms left in general positions, and Fig. 4(c) is based on the same light-atom content and a single U atom in an $mm2$ position. As might have been expected, the p.d.f.'s in Figs. 4(b) and (c) are very close to those for the space groups Pm and $P1$ respectively.

The p.d.f.'s shown in the above figures are influenced by (i) atomic heterogeneity, (ii) paucity of the atoms in the asymmetric unit, and (iii) functional forms of the trigonometric structure factors (or symmetry of the atomic arrangement, which determines these factors). This applies to the extent of the departure of the p.d.f.'s from ideal behavior (Wilson, 1949),

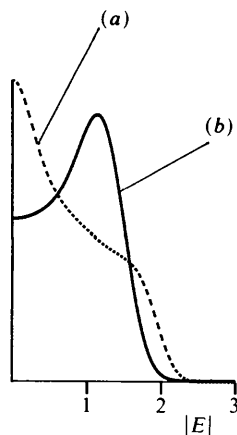


Fig. 2. Theoretical distributions for the space group $P2/m$. Both p.d.f.'s are on the same (arbitrary) scale. (a) As Fig. 1(a), for comparison. (b) Unit-cell contents: 36 C atoms in general positions, two U atoms on a twofold axis.

as well as to the effect of heavy scatterers in special positions. Thus for a fixed heterogeneity and paucity of atoms, the p.d.f.'s are different for different space groups and site symmetries. However, for a given symmetry of atomic arrangement the effects will generally decrease for decreasing heterogeneity; note that a decrease of heterogeneity may be caused by (i) an increased number of light atoms for a fixed number of heavy ones, (ii) a decrease of the atomic number of the heavy atom(s) for a fixed number of light ones, and (iii) an increased number of heavy atoms for a fixed number of light ones.

This study of effects of special positions on intensity statistics shows clearly that the exact Fourier

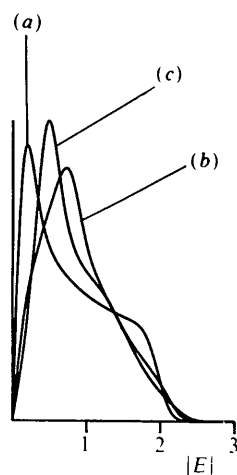


Fig. 3. Theoretical distributions for the space group $Pmm2$. The three p.d.f.'s are on the same (arbitrary) scale. (a) Asymmetric unit C_9U , all the atoms in general positions. (b) As (a) with two more U atoms on a mirror plane. (c) As (a) with one more U atom on the twofold axis ($mm2$).

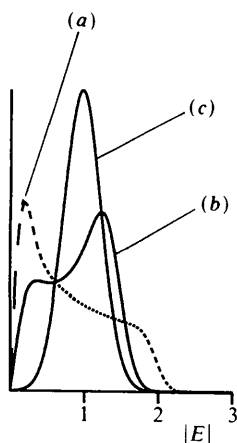


Fig. 4. Theoretical distributions for the space group $Pmm2$. The three p.d.f.'s are on the same (arbitrary) scale. (a) As Fig. 3(a), for comparison. (b) Unit-cell contents: 36 C atoms in general positions, two U atoms on a mirror plane. (c) Unit-cell contents: 36 C atoms in general positions, one U atom on twofold axis ($mm2$).

approach is capable of handling this predictable rational dependence. We intend to extend these investigations to multivariate p.d.f.'s of $|E|$, to be used in the field of direct methods.

This research has been supported in part by grant No. 84-00076 from the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel.

References

- ABRAMOWITZ, M. & STEGUN, I. A. (1972). *Handbook of Mathematical Functions*. New York: Dover.
 COLLIN, R. L. (1955). *Acta Cryst.* **8**, 499-502.
 FOSTER, F. & HARGREAVES, A. (1963). *Acta Cryst.* **16**, 1124-1133.
 HARGREAVES, A. (1956). *Acta Cryst.* **9**, 191-192.
 HAUPTMAN, H. & KARLE, J. (1953). *Acta Cryst.* **6**, 136-141.

- ILYUKHIN, V. V. & NIKITIN, A. V. (1963). *Sov. Phys. Crystallogr.* **8**, 3-6.
International Tables for X-ray Crystallography (1965). Vol. I, 2nd ed., edited by N. F. M. HENRY & K. LONSDALE. Birmingham: Kynoch Press.
 KARLE, J. & HAUPTMAN, H. (1953). *Acta Cryst.* **6**, 131-135.
 PRADHAN, D., GHOSH, S. & NIGAM, G. D. (1985). In *Structure and Statistics in Crystallography*, edited by A. J. C. WILSON, pp. 43-51. Guilderland: Adenine Press.
 SHMUELI, U. (1982). *Acta Cryst.* **A38**, 362-371.
 SHMUELI, U. & WEISS, G. H. (1985). In *Structure and Statistics in Crystallography*, edited by A. J. C. WILSON, pp. 53-66. Guilderland: Adenine Press.
 SHMUELI, U. & WEISS, G. H. (1987). *Acta Cryst.* **A43**, 93-98.
 SHMUELI, U., WEISS, G. H. & KIEFER, J. E. (1985). *Acta Cryst.* **A41**, 55-59.
 SHMUELI, U., WEISS, G. H., KIEFER, J. E. & WILSON, A. J. C. (1984). *Acta Cryst.* **A40**, 651-660.
 SIM, G. (1958). *Acta Cryst.* **11**, 123-124.
 SRINIVASAN, R. & PARTHASARATHY, S. (1976). *Some Statistical Applications in X-ray Crystallography*. Oxford: Pergamon Press.
 WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318-321.

Acta Cryst. (1988). **A44**, 417-421

POSIT – a Method for Structure Determination of Small Partially Known Molecules from Powder Diffraction Data. Structure of 6-Methyl-1,2,3,4-tetrahydropyrimidine-2,4-dione (6-Methyluracil)

BY G. RECK, R.-G. KRETSCHMER AND L. KUTSCHABSKY

*Zentralinstitut für Molekularbiologie der Akademie der Wissenschaften der DDR,
 Robert-Rössle-Strasse 10, Berlin DDR-1115, German Democratic Republic*

AND W. PRITZKOW

*Zentralinstitut für Physikalische Chemie der Akademie der Wissenschaften der DDR,
 Rudower Chaussee 5, Berlin DDR-1199, German Democratic Republic*

(Received 17 June 1987; accepted 6 January 1988)

Abstract

POSIT is a computer program package for the determination of the approximate structure of small organic molecules using known molecular fragments. The orientation and the translation vector of this fragment are determined and refined by a special *R*-value analysis. In contrast to other molecular replacement methods, only a few low-order strong reflections are needed to start the structure analysis. Therefore, this method is suitable for the determination of crystal structures of small rigid or semi-rigid organic molecules without heavy atoms from powder diffraction data. $C_5H_6N_2O_2$, $M_r = 126.11$, monoclinic, $P2_1/c$, $a = 4.5137(7)$, $b = 10.9888(12)$, $c = 11.7224(9)$ Å, $\beta = 97.554(8)^\circ$, $V = 576.39(7)$ Å³, $Z = 4$, $D_x = 1.453$ Mg m⁻³, λ (Cu $K\alpha$) = 1.54178 Å, $\mu = 0.987$ mm⁻¹, $F(000) = 264$, $T = 297$ K, final $R = 0.04$

for 165 approximate structure factors obtained from a powder diffraction measurement. The molecules were found in the diketo form. They are linked by a two-dimensional network of hydrogen bonds forming layers parallel to the *yz* plane.

Introduction

Normally, powder diffraction data are used for the identification of unknown crystalline samples, for symmetry and lattice-constant determination, for phase analysis and for refinement of approximately known structures. It is very difficult to determine an unknown structure because most reflections in the powder pattern overlap. Therefore, up to now only a few structures, most of them containing heavy atoms, have been solved from powder diffraction