A New Probability Distribution of the Triplet from Patterson Function Arguments. V

JORDI RIUS

Institut de Cikncia de Materials de Barcelona (CSIC) Campus de la UAB, 08193-Cerdanyola, Catalunya, Spain. E-mail: jordi, rius @ icmab, es

(Received 2 December 1996; accepted 21 March 1997)

Abstract

The physical basis of the conditional probability distributions of the quartet and of the triplet are investigated by means of the 'Patterson' and 'modulus' sum functions, respectively. From this study, a new conditional probability distribution of the triplet follows which has been tested on real data from three structures of different size. The empirically found distribution is compared with Cochran's closely related distribution.

1. Introduction

Historically, the practical application of structure invariants is related to the availability of the respective conditional probability distributions. Cochran (1955) showed that the probability distribution of the phase of a triplet,

$$
\Phi_3(\mathbf{H}, \mathbf{K}) = \varphi(-\mathbf{H}) + \varphi(\mathbf{K}) + \varphi(\mathbf{H} - \mathbf{K}), \qquad (1)
$$

given the product $|E(H)E(K)E(H - K)|$, corresponds to a 'von Mises' distribution centered at 0,

$$
\mathcal{P}(\Phi_3) = \frac{1}{2\pi I_o[K(\mathbf{H}, \mathbf{K})]} \exp[K(\mathbf{H}, \mathbf{K}) \cos \Phi_3], \quad (2)
$$

with I_0 being a modified Bessel function of the second kind (Watson, 1922, p. 181). Assuming throughout that the crystal structure consists of N equal atoms in the unit cell and that the space group is $P1$, then the concentration parameter determining the variance of this distribution reduces to

$$
K(H, K) = \frac{2}{N^{1/2}} |E(H)E(K)E(H - K)|.
$$
 (3)

Since the distribution is always centred at 0, no negative triplets can be predicted with (2). In the case of a quartet,

$$
\Phi_4(H, K, L) = \varphi(-H) + \varphi(K) + \varphi(L) + \varphi(M), \quad (4)
$$

with $M = H - K - L$, the situation is different. The use of quartets in direct procedures for phase solution was first introduced by Schenk (1973a,b, 1974). Later, Hauptman $(1975a,b)$ and Giacovazzo $(1976a,b)$ gave the probabilistic theories from quartets for $P1$. In practice, the following approximation (Giacovazzo, 1976b) to the conditional probability distribution of the quartet is normally used (DeTitta, Weeks, Thuman, Miller & Hauptman, 1994), namely

$$
\mathcal{P}(\Phi_4) = \frac{1}{2\pi I_o[K(\mathbf{H}, \mathbf{K}, \mathbf{L})]} \exp\{K(\mathbf{H}, \mathbf{K}, \mathbf{L})
$$

× cos[$\varphi_4 - \psi(\mathbf{H}, \mathbf{K}, \mathbf{L})$]}
(5)

where the concentration parameter is,

$$
K(\mathbf{H}, \mathbf{K}, \mathbf{L}) = \frac{2}{N} |E(\mathbf{H})E(\mathbf{K})E(\mathbf{L})E(\mathbf{M})| |X(\mathbf{H}, \mathbf{K}, \mathbf{L})|
$$
\n(6)

$$
X(H, K, L) = |E(-H + K)|^2 + |E(-H + L)|^2
$$

+ |E(K + L)|^2 - 2. (7)

This approximation to the probability distribution of the quartet is a 'von Mises' distribution conditioned by the seven magnitudes $|E(H)|$, $|E(K)|$, $|E(L)|$, $|E(M)|$, $|E(-H + \mathbf{K})|$, $|E(-H + L)|$ and $|E(\mathbf{K} + L)|$. Unlike (2), it is centered either at $\psi = 0$ or π depending on the sign of $X(H, K, L)$. The most reliable estimates of ψ are obtained when the four main terms $|E(H)|$, $|E(K)|$, $|E(L)|$ and $|E(M)|$ are all large, and simultaneously, the three 'cross terms' $|E(-H + K)|$, $|E(-H + L)|$, $|E(K + L)|$ are either all large or all small. In the first case positive quartets are predicted while the second case corresponds to the mostly negative quartets.

Most conventional direct-methods procedures for solving small crystal structures are based on phaserefinement functions derived from the probability distributions of the structure invariants. Consequently, one possible way to gain information about the physical basis of the conditional probability distributions, is to convert these phase-refinement functions to real space. The simplest phase-refinement function is the Z function,

$$
Z(\Phi) = \sum_{\mathbf{H}}^A \sum_{\mathbf{K}}^{\infty} \frac{2}{N^{1/2}} |E(\mathbf{H})E(\mathbf{K})E(\mathbf{H} - \mathbf{K})|
$$

× cos[Φ_3 (**H**, **K**)] = maximum, (8)

which follows from the product of the probabilities of the triplets, and where Φ denotes the collectivity of refined phases and A the asymmetrical unit of reciprocal space. Unfortunately, for crystal structures with polar or symmorphic space groups, refinement of phases with the Z function often ends in the unconstrained global maximum of Z. Similarly, the phaserefinement function,

$$
Q(\Phi) = \sum_{\mathbf{H}}^{A} \sum_{\mathbf{K}}^{\infty} \sum_{\mathbf{L}}^{\infty} \frac{2}{N} |E(\mathbf{H})E(\mathbf{K})E(\mathbf{L})E(\mathbf{M})| |X(\mathbf{H}, \mathbf{K}, \mathbf{L})|
$$

× cos[Φ_4 (**H**, **K**, **L**) – ψ (**H**, **K**, **L**)] = maximum, (9)

can be derived from the probability distributions of the quartets. The principal difference between (8) and (9) is that ψ can be either 0 or π in the O function. In this way, the incidence of the incorrect solutions (especially the U-atom solution) is greatly reduced. Both the Z and Q functions are employed in their reciprocal-space form. So far, little attention has been paid to the realspace meaning of Z and Q . One early interpretation of the Z function was given by Cochran (1952) who said that the integral,

$$
\int\limits_{V} \rho^3 \, dV, \tag{10}
$$

which is essentially equivalent to Z, must be a large and positive quantity. More recently, Rius (1993) showed that the Z function can be reinterpreted in the form of a 'direct methods' sum function involving the modulus function. According to this new view, the close relationship between the Q function and the Patterson sum function is investigated in the present contribution. As will be seen, the capability of predicting negative quartets is related to the removal of the origin peak of the observed Patterson function. By applying the same procedure to the triplet case, a new conditional probability distribution of the triplet follows.

2. The Patterson sum function

The Patterson sum function is a direct methods sum function in which the Patterson-type synthesis is the 'true' Patterson function. It is given by the integral,

$$
Q(\Phi) = V^3 \int_{V} P'(u)P(u, \Phi) du,
$$
 (11)

where $P'(u)$ is the observed Patterson function of the squared point atom structure with partially substracted origin peak, $P(u, \Phi)$ is the Patterson function expressed in terms of the phases Φ , and finally V is the volume of the unit cell. The observed Patterson function of the squared point atom structure is evaluated as,

$$
\mathbf{P}'(\mathbf{u}) = \frac{1}{V} \sum_{\mathbf{H}}^{\infty} \left[|G(\mathbf{H})|^2 - p \langle |G|^2 \rangle \right] \exp(-i2\pi \mathbf{H} \mathbf{u}), \quad (12)
$$

where the $|G(H)|$ are the amplitudes of the structure factors of the squared point atom structure, $\langle |G|^2 \rangle$ is the value of $|G(H)|^2$ averaged over all H, and p is a factor which determines the fraction of origin peak removed.

According to their definition and since the space group is assumed to be $P1$, the structure factors of the point atom structure are the normalized structure factors. Thus, for an equal atom structure, the identities,

$$
|G(\mathbf{H})|^2 = \frac{1}{N} |E(\mathbf{H})|^2
$$
 (13)

and

$$
\langle |G|^2 \rangle = \frac{1}{N} \langle |E|^2 \rangle, \tag{14}
$$

hold. Using Parseval's theorem and taking into account (12), (13) and (14), the reciprocal-space form of $\mathcal{Q}(\phi)$ follows

$$
Q(\Phi) = \frac{V^2}{N} \sum_{\mathbf{h}}^{\infty} \left[|E(\mathbf{h})|^2 - p \langle |E|^2 \rangle \right] |G(\mathbf{h}, \Phi)|^2, \quad (15)
$$

where, for convenience, the usual symbols H , K , L of the reciprocal lattice vectors have been replaced by h, k, l. $|G(h, \phi)|^2$ results from multiplying $G(h, \phi)$ (see the *Appendix)* by its complex conjugate, so that

$$
|G(\mathbf{h}, \Phi)|^2 = \frac{1}{V^2} \sum_{\mathbf{k}}^{\infty} \sum_{\mathbf{l}}^{\infty} \mathbf{E}(\mathbf{k}) \mathbf{E}(\mathbf{h} - \mathbf{k}) \mathbf{E}(\mathbf{l}) \mathbf{E}(-\mathbf{h} - \mathbf{l}).
$$
\n(16)

Introducing (16) in (15) and since $\langle |E|^2 \rangle$ is equal to unity, $Q(\phi)$ turns out to be,

$$
\mathcal{Q}(\Phi) = \frac{1}{N} \sum_{\mathbf{h}}^{\infty} \sum_{\mathbf{k}}^{\infty} \sum_{\mathbf{l}}^{\infty} \mathbf{E}(\mathbf{k}) \mathbf{E}(\mathbf{h} - \mathbf{k}) \mathbf{E}(\mathbf{l}) \mathbf{E}(-\mathbf{h} - \mathbf{l})
$$

$$
\times [|E(\mathbf{h})|^2 - p].
$$
 (17)

Now, if the reciprocal vectors are transformed according to $k=-H$, $h-k=K$, $l=L$ then $h=-H+K$, $-h-l=H-K-L=M$ and (17) takes the more familiar form,

$$
\mathcal{Q}(\Phi) = \frac{2}{N} \sum_{\mathbf{H}}^{\mathcal{A}} \sum_{\mathbf{K}}^{\infty} \sum_{\mathbf{L}}^{\infty} |E(\mathbf{H})E(\mathbf{K})E(\mathbf{L})E(\mathbf{M})|
$$

$$
\times [|E(-\mathbf{H} + \mathbf{K})|^2 - p] \cos[\Phi_4(\mathbf{H}, \mathbf{K}, \mathbf{L})]. \quad (18)
$$

Since Q is a large quantity, then the most probable values of Φ_4 are those which maximize the corresponding summation terms. However, there are three different summation terms for a given $\cos[\Phi_{A}(\mathbf{H}, \mathbf{K}, \mathbf{L})]$. Effectively, from the 24 summation terms generated by permutation of the $-H$, K, L and M indices having identical product *IE(-H)E(K)E(L)E(M)I* and $cos[\Phi_{4}(H, K, L)]$, there are three types differing in the factors $[|E(-\mathbf{H} + \mathbf{K})|^2 - p]$, $[|E(-\mathbf{H} + \mathbf{L})|^2 - p]$ and $[|E(\mathbf{K} + \mathbf{L})|^2 - p]$, *i.e.*

$$
(2/N)|E(\mathbf{H})E(\mathbf{K})E(\mathbf{L})E(\mathbf{M})|\left[|E(-\mathbf{H} + \mathbf{K})|^2 - p\right] \times \cos[\Phi_4(\mathbf{H}, \mathbf{K}, \mathbf{L})]
$$
(19)

$$
(2/N) |E(\mathbf{H})E(\mathbf{K})E(\mathbf{L})E(\mathbf{M})| [|E(-\mathbf{H} + \mathbf{L})|^2 - p]
$$

$$
\times \cos[\Phi_4(\mathbf{H}, \mathbf{K}, \mathbf{L})]
$$
 (20)

$$
(2/N) |E(\mathbf{H})E(\mathbf{K})E(\mathbf{L})E(\mathbf{M})| [|E(\mathbf{K} + \mathbf{L})|^2 - p]
$$

× cos[Φ_4 (**H**, **K**, **L**)]. (21)

The information contained in these three terms can be easily combined by assuming that the value of $\cos \Phi_4$ which maximizes their sum,

$$
(2/N) |E(\mathbf{H})E(\mathbf{K})E(\mathbf{L})E(\mathbf{M})| |[E(-\mathbf{H} + \mathbf{K})]^2
$$

+ |E(-\mathbf{H} + \mathbf{L})|^2 + |E(\mathbf{K} + \mathbf{L})|^2 - 3p]
× cos[Φ_4 (**H**, **K**, **L**)], (22)

will the most probable. Notice that, for $p = 2/3$, (22) coincides with the argument of the exponential in the quartet distribution (5). This result is thus the demonstration that the $\ddot{\theta}$ function (9) derived from the probability distribution of the quartet is essentially equivalent to the Patterson sum function (15) with $p=2/3$, and illustrates the physical basis of the conditional probability distribution of the quartet.

3. The modulus sum function

On the analogy of the Patterson sum function, the modulus sum function is defined by the expression,

$$
\mathcal{Z}(\boldsymbol{\Phi}) = V^2 \int\limits_{\mathbf{v}} P(\mathbf{u}) P(\mathbf{u}, \boldsymbol{\Phi}) \, \mathrm{d}\mathbf{u}, \tag{23}
$$

where the Patterson function has been replaced by the modulus function $P(u)$. According to Ramachandran & Raman (1959), the principal difference between the modulus and the Patterson function is the change of the relative heights of origin to non-origin peaks. The modulus function P' of the squared structure with partially removed peaks is,

$$
P'(\mathbf{u}) = \frac{1}{V} \sum_{\mathbf{H}}^{\infty} \left[|G(\mathbf{H})| - p \langle |G| \rangle \right] \exp(-i2\pi \mathbf{H} \mathbf{u}), \quad (24)
$$

where the coefficients $|G(H)|$ are obtained by taking the square root of (13), so that,

$$
|G(\mathbf{H})| = \frac{1}{N^{1/2}} |E(\mathbf{H})|
$$
 (25)

$$
\langle |G| \rangle = \frac{1}{N^{1/2}} \langle |E| \rangle, \tag{26}
$$

and the Fourier coefficients in (24) become

$$
|G(\mathbf{H})| - p\langle |G| \rangle = \frac{1}{N^{1/2}}[|E(\mathbf{H})| - p\langle |E| \rangle]. \tag{27}
$$

Transformation of $\mathcal{Z}(\Phi)$ to reciprocal space is performed by applying Parseval's theorem to (23), so that, in view of (27),

$$
\mathcal{Z}(\Phi) = \frac{V}{N^{1/2}} \sum_{\mathbf{H}}^{\infty} \left[|E(\mathbf{H})| - p \langle |E| \rangle \right] |G(\mathbf{H}, \Phi)|. \tag{28}
$$

Since for an equal atom structure the phases of $G(H)$ and $E(H)$ are alike, then $\phi(H) = \varphi(H)$, and from (38),

$$
|G(\mathbf{H}, \Phi)| = \frac{1}{V} \sum_{\mathbf{K}}^{\infty} |E(\mathbf{K})E(\mathbf{H} - \mathbf{K})| \exp[i\Phi_3(\mathbf{H}, \mathbf{K})].
$$
\n(29)

Substitution of (29) into (28) leads to,

$$
\mathcal{Z}(\Phi) = \frac{2}{N^{1/2}} \sum_{\mathbf{H}}^{A} \sum_{\mathbf{K}}^{\infty} [|E(\mathbf{H})| - p\langle |E| \rangle] |E(\mathbf{K})E(\mathbf{H} - \mathbf{K})|
$$

× cos[$\Phi_3(\mathbf{H}, \mathbf{K})$], (30)

where **H** ranges uniformly over those vectors in the asymmetrical unit A of reciprocal space. There are three different terms in the double summation for each $\cos[\Phi_3(H, K)]$. These terms are

$$
(2/N^{1/2})[|E(\mathbf{H})| - p\langle |E| \rangle] |E(\mathbf{K})E(\mathbf{H} - \mathbf{K})|
$$

$$
\times \cos[\Phi_3(\mathbf{H}, \mathbf{K})] \tag{31}
$$

$$
(2/N^{1/2})[|E(\mathbf{K})| - p\langle |E| \rangle] |E(-\mathbf{H})E(\mathbf{H} - \mathbf{K})|
$$

$$
\times \cos[\Phi_3(\mathbf{H}, \mathbf{K})] \tag{32}
$$

$$
(2/N^{1/2})[|E(\mathbf{H} - \mathbf{K})| - p\langle |E|\rangle] |E(-\mathbf{H})E(\mathbf{K})|
$$

$$
\times \cos[\Phi_3(\mathbf{H}, \mathbf{K})].
$$
 (33)

As for the quartet case, it will be assumed that the most probable value of $cos[\varphi_3(H, K)]$ is that which maximizes the sum of the three terms namely

$$
\frac{2}{N^{1/2}}|E(\mathbf{H})E(\mathbf{K})E(\mathbf{H}-\mathbf{K})|X_p(\mathbf{H},\mathbf{K})\cos[\Phi_3(\mathbf{H},\mathbf{K})],
$$
\n(34)

with

$$
X_{p}(\mathbf{H}, \mathbf{K}) = 3 - p\langle |E| \rangle \left[\frac{1}{|E(\mathbf{H})|} + \frac{1}{|E(\mathbf{K})|} + \frac{1}{|E(\mathbf{H} - \mathbf{K})|} \right].
$$
\n(35)

Due to the close resemblance between expressions (22) for the quartets and (34) for the triplets, it is to be expected that (34) corresponds to the argument of the exponential in the probability distribution of the triplet. Since the value of p can vary between 0 and 1, the best p value will be determined from a series of test calculations. Notice that it is not possible to derive Cochran's distribution from (34). Effectively, for $p = 0$, the value of X_p (**H**, **K**) is three and not unity as Cochran's distribution demands.

4. Test calculations

The finality of the test calculation is to study the effect of introducing different p values in the concentration parameter,

$$
K_p = \frac{1}{N^{1/2}} |E(\mathbf{H})E(\mathbf{K})E(\mathbf{H} - \mathbf{K})| X_p(\mathbf{H}, \mathbf{K}).
$$
 (36)

The real data employed to perform this study correspond to the structures given in Table 1. To measure the agreement between real and calculated data, the mean phase error (MPE) defined by

$$
MPE(p) = \langle \left| \arccos\{\cos[\Phi_3(H, K)] \} \right|
$$

- arccos[$I_1(K_p(H, K)/I_o(K_p(H, K))]$ (37)

has been used wherein the quotient $I_1(K_p(H, K)/I_p(K_p(H, K))$ is the expected value of the cosine of the triplet $-H + K + (H - K)$ (Germain, Main & Woolfson, 1970). For the sake of completeness, the MPE's for Cochran's distribution have also been computed. The values of Φ_3 in (37) have been derived from the published refined atomic positions. All calculations have been performed with a modified version of the program *XLENS* (Rius, 1993).

Inspection of Table 2, indicates that, in general, the best results are obtained for p in the range $0-2/3$, since, for $p = 1$, the MPE's of the triplets involving one weak E value are very large (specially for PGE2).

Inspection of Table 2, also shows a discrepancy between the MPE's of Cochran's distribution and the rest. This discrepancy which is less obvious for MBH2, becomes more evident for the largest structure TVAL. This result suggests that there is a factor or three missing in the concentration parameter of Cochran's distribution.

This work was partially supported by the Dirección Gral de Ensefianza Superior (Project PB95-0115) and the Direcci6 Gral de Recerca de la Generalitat de Catahnya (Grant No. SGR 00460, 1995).

APPENDIX The structure factor $G(H,\phi)$

The structure factor $G(H, \phi)$ is the Fourier transform of the squared point atom structure expressed in terms of the phases φ (H) associated with the normalized structure factors $E(H)$. According to the convolution theorem which states that the Fourier transform of the product of two functions is the convolution of the Fourier transforms of the individual functions, $G(H, \phi)$ can be expressed as the sum

$$
\mathbf{G}(\mathbf{H}, \boldsymbol{\Phi}) = |G(\mathbf{H}, \boldsymbol{\Phi})| \exp i\boldsymbol{\phi}(\mathbf{H})
$$

$$
= \frac{1}{V} \sum_{\mathbf{K}}^{\infty} \mathbf{E}(\mathbf{K}) \mathbf{E}(\mathbf{H} - \mathbf{K}). \tag{38}
$$

Table 1. *Relevant crystal data of the examples used for the test calculations*

Structure code	Formula	Space group	N	Reference
PGE ₂	$C_{20}H_{32}O_5$	P1.	25	(a)
MBH ₂	C_1 ₅ H_2 ₄ O_3	P1.	54	(b)
TVAL	$C_{54}H_{90}N_6O_{18}$	P1	156	(c)

References: (a) DeTitta, Langs, Edmonds & Duax (1980); (b) Poyser, Edwards, Anderson, Hursthouse, Walker, Sheldrick & Walley (1986); (c) Karle (1975); Smith, Duax, Langs, DeTitta, Edmonds, Rohrer & Weeks (1975).

Table 2. MPE's (°) for the distributions with concen*tration parameters Kp* $(p = 0, 1/3, 2/3, 2/3)$ *and 1) and for Cochran's distribution*

 E_1, E_2, E_3 are the E amplitudes involved in the triplet and N_T the number of triplets. The results indicated that, in general, the smallest MPE's are obtained for p between 0 and $2/3$. Note the discrepancies between the MPE's found for Cochran's distribution and for the rest, specially in the case of the largest structure TVAL.

References

- Cochran, W. (1952). *Acta Cryst.* 5, 65-67.
- Cochran, W. (1955). *Acta Cryst.* 8, 473-478.
- DeTitta, G., Langs, D. A., Edmonds, J. W. & Duax, W. L. (1980). *Acta Cryst.* B36, 638-645.
- DeTitta, G., Weeks, C., Thuman, P., Miller, R. & Hauptman, H. (1994) *Acta Cryst.* A50, 203-210.
- Germain, G., Main, P. & Woolfson, M. M. (1970). *Acta Cryst.* B26, 274-285.
- Giacovazzo, C. (1976a). *Acta Cryst.* A32, 74-82.
- Giacovazzo, C. (1976b). *Acta Cryst.* A32, 91-99.
- Hauptman, H. (1975a). *Acta Cryst.* A31,671-679.
- Hauptman, H. (1975b). *Acta Cryst.* A31,680-687.
- Karle, I. L. (1975). *J. Am. Chem. Soc.* 97, 4379-4386.
- Poyser, J., Edwards, R., Anderson, J., Hursthouse, M., Walker, N., Sheldrick, G. & Walley, A. (1986). J. *Antibiot.* 39, 167-169.
- Ramachandran, G. N. & Raman, S. (1959). *Acta Cryst.* 12, 957-964.
- Rius, J. (1993). *Acta Cryst.* A49, 406-409.
- Schenk, H. (1973a). *Acta Cryst.* A29, 77-82.
- Schenk, H. (1973b). *Acta Cryst.* A29, 480-481.
- Schenk, H. (1974). *Acta Cryst.* A30, 477-481.
- Smith, G. D., Duax, W. L., Langs, D. A., DeTitta, G. T., Edmonds, J. W., Rohrer, D. C. & Weeks, C. M. (1975). J. *Am. Chem. Soc.* 97, 7242-7247.
- Watson, G. N. (1922). *A Treatise on the Theory of Bessel Functions.* Cambridge University Press.