

# Comment on *Probability density functions of the average and difference intensities of Friedel opposites* by Shmueli & Flack (2010)

Andrzej Olczak

Received 4 November 2010

Accepted 2 March 2011

 © 2011 International Union of Crystallography  
Printed in Singapore – all rights reserved

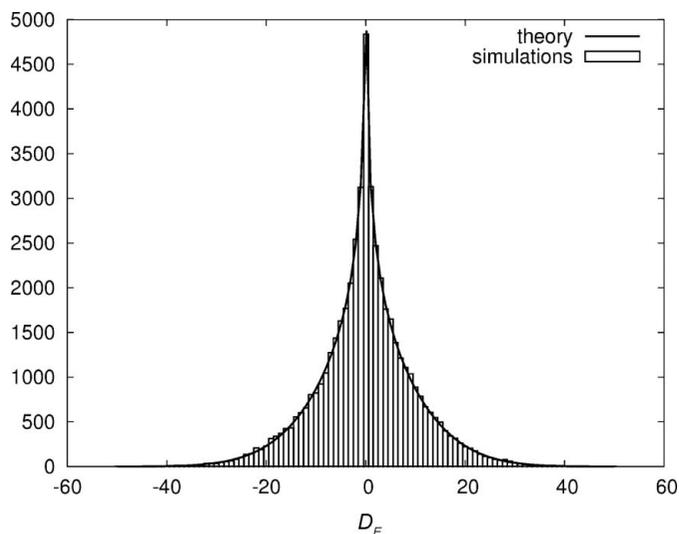
Institute of General and Ecological Chemistry, Technical University of Łódź, Żeromskiego 116, 90-924 Łódź, Poland. Correspondence e-mail: olczakan@p.lodz.pl

An interesting article by Shmueli & Flack (2010) (referred to as S&F in the following) was published recently [*Acta Cryst.* **A66**, 669–675]. Part of this article is devoted to probability density functions (p.d.f.s) and simulations of Friedel intensity differences, upon which I would like to share some remarks:

(1) Simulations presented in S&F are based on the assumption that  $\mathbf{h} \cdot \mathbf{R}_{jk}$  ( $\mathbf{R}_{jk} = \mathbf{r}_j - \mathbf{r}_k$ , where  $\mathbf{r}_j$  denote atom positions,  $j, k = 1, \dots, N$  and  $N$  is the number of atoms in the unit cell) are random, statistically independent variables whose fractional parts are uniformly distributed in the  $[0, 1]$  range. This assumption made by the authors (hereafter referred to assumption A) is inconsistent with the commonly held assumption that atoms are independently and uniformly distributed in the unit cell. This second assumption (hereafter referred to as assumption B) differs from assumption A in that fractional parts of  $\mathbf{h} \cdot \mathbf{r}_j$  (not  $\mathbf{h} \cdot \mathbf{R}_{jk}$ ) are random, independent variables uniformly distributed in the  $[0, 1]$  range.

Under assumption B one can define  $\mathbf{R}_{jk}$  vectors as the  $\mathbf{r}_j - \mathbf{r}_k$  differences ( $\mathbf{R}_{jk} \equiv \mathbf{r}_j - \mathbf{r}_k$ ). These symbols ( $\mathbf{R}_{jk}$  and  $\mathbf{r}_j - \mathbf{r}_k$ ) are different ways of denoting the same thing.

Under assumption A, however,  $\mathbf{R}_{jk}$  vectors can be generated independently. This means that what was the identity  $\mathbf{R}_{jk} \equiv \mathbf{r}_j - \mathbf{r}_k$  now becomes the system of equations  $\mathbf{R}_{jk} = \mathbf{r}_j - \mathbf{r}_k$  (with  $\mathbf{r}_j$  as unknowns), which can be fulfilled or not. It is worth noting that there are a lot more ‘degrees of freedom’ on the left-hand side of this system  $[N(N-1)/2]$  than on the right-hand side ( $N$ ), which suggests that the solution to this system need not exist.



**Figure 1**  
Histogram calculated by the random-structure model ( $5 \times 10^4$  simulations) and p.d.f. (under assumption B) of  $D_E$  for the compound  $U_2C_{19}$ .  $R = 0.03$ . The abscissa is on an arbitrary scale.

Let us consider a simple example showing the difference between these two approaches. The structure factor  $F_{\mathbf{h}} = \sum_{j=1}^N f_j \exp(2\pi\mathbf{h} \cdot \mathbf{r}_j)$  under assumption B is a function of  $\mathbf{r}_j$  vectors,  $F_{\mathbf{h}}(\mathbf{r}_1, \dots, \mathbf{r}_N)$ . The square of its modulus (and the modulus itself) is always a non-negative real number,  $|F_{\mathbf{h}}|^2 \geq 0$ .  $|F_{\mathbf{h}}|^2$  can be expressed in the form

$$|F_{\mathbf{h}}|^2 = \sum_{j=1}^N f_j^2 + 2 \sum_{j>k} f_j f_k \cos(2\pi\mathbf{h} \cdot (\mathbf{r}_j - \mathbf{r}_k)).$$

If we replace  $\mathbf{r}_j - \mathbf{r}_k$  by  $\mathbf{R}_{jk}$  (which are independent under assumption A) in this expression, we get a completely new function [let us call it  $M(\mathbf{R}_{12}, \mathbf{R}_{13}, \dots, \mathbf{R}_{N-1,N})$ ]:

$$M = \sum_{j=1}^N f_j^2 + 2 \sum_{j>k} f_j f_k \cos(2\pi\mathbf{h} \cdot \mathbf{R}_{jk}),$$

which can take negative values, if for example all  $\cos(2\pi\mathbf{h} \cdot \mathbf{R}_{jk}) = -1$ . Therefore, this function can no longer be interpreted as the square of the modulus of the familiar structure factor  $F_{\mathbf{h}}$ .

(2) Simulations based on assumptions A and B lead to completely different results, which can be illustrated by using simulation results for the hypothetical structure  $U_2C_{19}$ . These simulations are based on the following formula for the anomalous differences [equation (14) in S&F]:

$$D_E(\mathbf{h}) = 4 \sum_{L_{jk}} d_{jk} \sin(2\pi\mathbf{h} \cdot \mathbf{R}_{jk}). \quad (1)$$

Assumption A produces the histogram presented in Fig. 2 in S&F, while assumption B leads to the histogram presented in Fig. 1 of this letter.

Which of these two assumptions is more appropriate for describing the structure with atoms randomly distributed in the unit cell?

Assumption B demonstrates that  $\mathbf{h} \cdot \mathbf{r}_j$  are uniformly distributed. From a statistical point of view there is no difference whether  $\mathbf{h}$  or  $\mathbf{r}_j$  is varying in the  $\mathbf{h} \cdot \mathbf{r}_j$  expression while the other parameter is fixed. Consequently, a very similar histogram to that presented in Fig. 1 of this letter should be obtained by the following procedure: using a random-number generator we generate independently atom positions for two U atoms and 19 C atoms, then we calculate structure factors, normalize them, calculate all anomalous differences and finally build the respective histogram out of these differences.

Let us try to apply this procedure in the case of assumption A. First, using the random-number generator we generate independently a set of  $\mathbf{R}_{jk}$  interatomic vectors, then in order to calculate structure factors we need to determine the atom positions,  $\mathbf{r}_j$ , for all atoms, or at least the fractional parts of the  $\mathbf{h} \cdot \mathbf{r}_j$  scalar products. Unfortunately, in most cases this is not possible, because of the reason given below.

(3) Let us consider  $N(N-1)/2$  independent random variables  $\mathbf{h} \cdot \mathbf{R}_{jk}$ , each of which is uniformly distributed over the  $[0, 1]$  range.

Using a random-number generator one can generate a set of  $N(N-1)/2$  values (let us call them  $T_{jk}$ ) of these variables.

To determine  $\mathbf{h} \cdot \mathbf{r}_i$  corresponding to the generated set of values  $T_{jk}$ , we need to solve the system of linear equations

$$T_{jk} = \mathbf{h} \cdot \mathbf{r}_j - \mathbf{h} \cdot \mathbf{r}_k.$$

This system is, however, overdetermined (for  $N > 3$ ) because it contains  $N(N-1)/2$  equations and only  $N$  unknowns  $\mathbf{h} \cdot \mathbf{r}_k$ . Usually, there is no solution to such a system and consequently the values  $T_{jk}$  are purely abstract numbers, which do not correspond to any configuration of atoms in the unit cell (see Fig. 2 for more geometrical explanation). In such a case  $D_E$  calculated on the basis of the formula

$$D_E(\mathbf{h}) = 4 \sum_{L=jk} d_{jk} \sin(2\pi T_{jk})$$

loses its physical meaning. Hence the histograms are built of these 'unphysical states'.

In summary, the results concerning the simulations of Friedel intensity differences and the respective p.d.f.s presented in the S&F paper are interesting from a mathematical point of view, but do not appear to relate to any crystal structure.

(4) One exception exists when independence of the  $\mathbf{h} \cdot \mathbf{R}_{jk}$  variables is consistent with assumption B: all but one of the atoms in the structure have the same atomic factor phase  $\delta_Q$  [ $f_Q = f_{oQ} \exp(i\delta_Q)$ , where  $f_{oQ} = [(f_Q + f'_Q)^2 + (f''_Q)^2]^{1/2}$ ] (Olczak *et al.*, 2003). In such a case we can choose the origin of the coordinate system at this particular atom, which allows for the transformation of all  $\sin(2\pi \mathbf{h} \cdot \mathbf{R}_{jk})$  components in equation (1) into the  $\sin(2\pi \mathbf{h} \cdot \mathbf{r}_k)$  form. It can be shown using the central limit theorem that in this case the

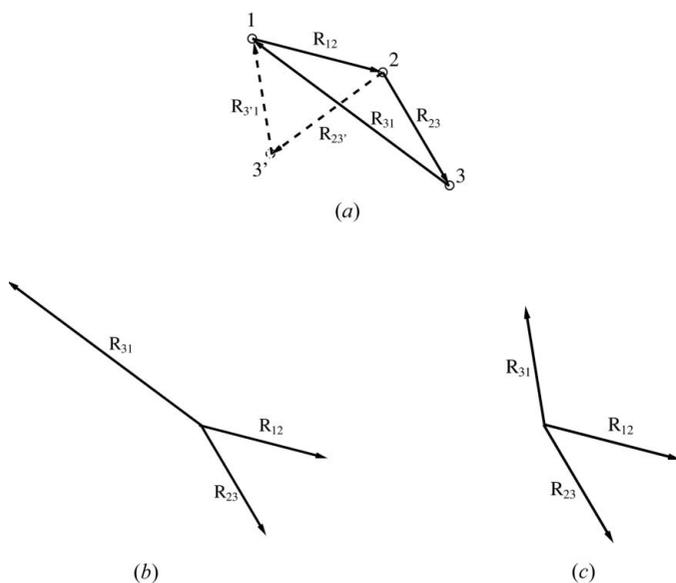


Figure 2

(a) Three atoms represented by small circles and their interatomic vectors,  $\mathbf{R}_{jk}$ . Movement of atom 3 to position 3' causes a change of two vectors:  $\mathbf{R}_{13}$  and  $\mathbf{R}_{23}$ , which means that interatomic vectors  $\mathbf{R}_{13}$  and  $\mathbf{R}_{23}$  are not independent. They always have to fulfil the relation  $\mathbf{R}_{12} + \mathbf{R}_{23} + \mathbf{R}_{31} = 0$ . (b) Possible collection of three randomly generated vectors. These vectors fulfil the condition  $\mathbf{R}_{12} + \mathbf{R}_{23} + \mathbf{R}_{31} = 0$  and, consequently, there exists an arrangement of three atoms (the structure) for which these vectors can be interpreted as interatomic vectors [see part (a)]. (c) Another possible collection of three vectors generated randomly. These vectors do not fulfil the relation  $\mathbf{R}_{12} + \mathbf{R}_{23} + \mathbf{R}_{31} = 0$  and consequently there is no configuration of three atoms for which these vectors could be interpreted as interatomic vectors. This means that they cannot be expressed as  $\mathbf{R}_{12} = \mathbf{r}_1 - \mathbf{r}_2$ ,  $\mathbf{R}_{23} = \mathbf{r}_2 - \mathbf{r}_3$ ,  $\mathbf{R}_{31} = \mathbf{r}_3 - \mathbf{r}_1$ , where  $\mathbf{r}_j$  represent atom positions. These vectors should not contribute to  $D_E$  and  $A_E$  histograms, because they are not associated with any arrangement of atoms in the unit cell.

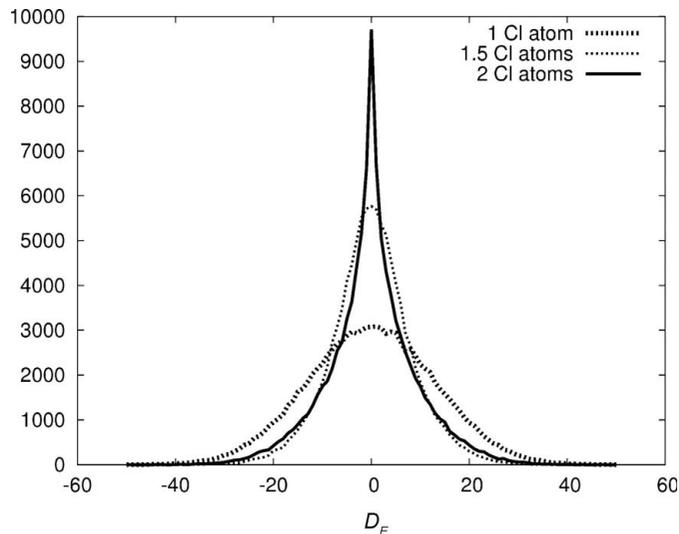


Figure 3

Histograms of  $D_E$  for the compound  $C_{15}H_{23}NO_2$  (Cu  $K\alpha$  radiation) by the random-structure model ( $10^5$  simulations under assumption B) containing in addition one Cl atom (thick dotted line), two Cl atoms (solid line), or one Cl atom with an occupancy factor 0.5 and one with an occupancy factor 1 (thin dotted line). The abscissa is on an arbitrary scale and is not common to all plots. All plots are normalized to  $10^5$ .

p.d.f. for  $D_E$  is a Gaussian function if the number of atoms with phase  $\delta_Q$  is large enough (in practice greater than six or seven) (Olczak *et al.*, 2003; Olczak, 2004).

(5) Under assumption B, the p.d.f. of the Friedel differences for the  $U_2C_{19}$  structure (in which two atoms are distinguished) takes a quite different form from the Gaussian shape. In this case  $D_E$  can be represented in the form

$$D_E = -8d_{UC} \cos(2\pi \mathbf{h} \cdot \mathbf{r}_{U_1}) \sum_{j=1}^{N-2} \sin(2\pi \mathbf{h} \cdot \mathbf{r}_{C_j})$$

if one puts the origin of the coordinate system exactly in the middle between the two U atoms ( $\mathbf{r}_{U_1} = -\mathbf{r}_{U_2}$ ).  $D_E$  is a product of two independent random variables:  $\cos(2\pi \mathbf{h} \cdot \mathbf{r}_{U_1})$  and  $\sum_{j=1}^{N-2} \sin(2\pi \mathbf{h} \cdot \mathbf{r}_{C_j})$ . Now we can follow the reasoning of Parthasarathy & Srinivasan (1964) to get a p.d.f. in the form

$$2^{1/2} / (\pi^{3/2} a) K_0(D_E^2/a^2) \exp(-D_E^2/a^2),$$

where  $a$  is a constant and  $K_0$  is a modified Bessel function of the second kind. This function fits perfectly the simulations of  $D_E$  based on assumption B (Fig. 1) and is a generalization of results obtained by Parthasarathy and Srinivasan, who considered structures containing  $P$  resonant atoms (in this case  $P = 2$ ) of the same chemical element along with  $Q$  non-resonant atoms where  $Q$  has to be large. The generalization is that  $Q$  atoms can also be resonant ones.

Some other cases considered by Parthasarathy and Srinivasan can be generalized as well, but it is much more difficult (if at all possible) to construct a universal p.d.f. embracing all possibilities (any chemical composition of the structure) at once, especially when partially occupied sites are considered (Olczak, 2008) (Fig. 3).

(6) There is also a good side to this diversity of p.d.f. shapes, which opens some new opportunities. In principle, it should be possible to determine from diffraction data whether the structure contains one or more strong anomalous scatterers. It is necessary, however, to remember that experimental differences, along with anomalous differences, also contain measurement uncertainties (see §3.6 in Olczak *et al.*, 2010), which should be deconvoluted before the experimental differences are compared to the theoretical ones.

### References

- Olczak, A. (2004). *46th Polish Crystallographic Meeting*, p. 16. Wrocław, Poland.
- Olczak, A. (2008). *50th Polish Crystallographic Meeting*, p. 191. Wrocław, Poland.
- Olczak, A., Cianci, M., Hao, Q., Rizkallah, P. J., Raftery, J. & Helliwell, J. R. (2003). *Acta Cryst.* **A59**, 327–334.
- Olczak, A., Głowska, M. L., Szczesio, M., Bojarska, J., Wawrzak, Z. & Duax, W. L. (2010). *Acta Cryst.* **D66**, 874–880.
- Parthasarathy, S. & Srinivasan, R. (1964). *Acta Cryst.* **17**, 1400–1407.
- Shmueli, U. & Flack, H. D. (2010). *Acta Cryst.* **A66**, 669–675.

## Response to Olczak's comment on *Probability density functions of the average and difference intensities of Friedel opposites*

U. Shmueli<sup>a\*</sup> and H. D. Flack<sup>b</sup>

Received 27 January 2011  
 Accepted 2 March 2011

© 2011 International Union of Crystallography  
 Printed in Singapore – all rights reserved

<sup>a</sup>School of Chemistry, Tel Aviv University, 69978 Tel Aviv, Israel, and <sup>b</sup>Département de chimie minérale, analytique et appliquée, Faculté des sciences, Université de Genève, Switzerland. Correspondence e-mail: ushmueli@post.tau.ac.il

The recently published article by Shmueli & Flack (2010) deals with average and difference reduced intensities of  $|F(\mathbf{h})|^2$  and  $|F(-\mathbf{h})|^2$ , known as Friedel opposites. First, well established trigonometric series of these quantities are re-derived and normalized and their probability density functions (hereafter: p.d.f.s), admitting any chemical composition and not relying in any way on the central limit theorem, are derived for the space group *P1*. This derivation rests on an assumption which is the basis of major grievances of Olczak (2011), to which we shall respond.

Consider the expressions for the structure factor and for its squared magnitude:

$$F(\mathbf{h}) = \sum_{j=1}^N f_j \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j) \quad (1)$$

and

$$|F(\mathbf{h})|^2 = \sum_{j=1}^N \sum_{k=1}^N f_j f_k^* \exp(2\pi i \mathbf{h} \cdot \mathbf{R}_{jk}), \quad (2)$$

where  $\mathbf{R}_{jk} = \mathbf{r}_j - \mathbf{r}_k$  and the (complex) atomic scattering factor contains the displacement parameters and those of resonant scattering. Obviously, the argument of the exponential in equation (2) will also appear in the expressions for the average and difference reduced intensities, normalized or not.

The fractional part of the scalar product  $\mathbf{h} \cdot \mathbf{r}_j$  in equation (1) is assumed in various applications to direct methods and intensity statistics to be uniformly distributed in the [0, 1] interval. This assumption has a number-theoretical justification (e.g. Weyl, 1916) and can be used if all the atoms (and especially the heavy atoms) are in general positions. Olczak (2011) calls it assumption B.

In our derivations of the p.d.f.s of the normalized average and difference reduced intensities we naturally came across the scalar product  $\mathbf{h} \cdot \mathbf{R}_{jk}$  shown in equation (2). We assumed that the interatomic vectors  $\mathbf{R}_{jk} = \mathbf{r}_j - \mathbf{r}_k$  are in the general positions of the Patterson unit cell and hence that the fractional part of the scalar product  $\mathbf{h} \cdot \mathbf{R}_{jk} = \mathbf{h} \cdot (\mathbf{r}_j - \mathbf{r}_k)$  is uniformly distributed in the [0, 1] interval. The overriding advantage of this approach is that it allows p.d.f.s to be obtained in analytical form, but of course they needed to be subjected to several tests to establish their usefulness (see Shmueli

& Flack, 2010). The text of Shmueli & Flack (2010) makes it clear that it is being used as an assumption (approximation) in need of further verification.

According to Olczak (2011), the use of this assumption (i) is inconsistent with the 'commonly used' assumption B, (ii) the result obtained from our calculation 'loses its physical sense' and (iii) our results 'are interesting from the mathematical point of view but not related to any crystal structure'.

Having derived the p.d.f.s, we performed two runs of tests: First, we compared the p.d.f.s with simulated histograms recalculated from a hypothetical structure. The results, shown in Figs. 1 and 2 of Shmueli & Flack (2010), display very good to good agreement of the histograms with the p.d.f.s. This mainly confirmed the correctness of the expressions.

However, a p.d.f. is physically meaningful if it agrees with a distribution related to experimental data, and a crucial test was indicated. We therefore compared the p.d.f.s with histograms recalculated from the parameters of a solved *P1* structure (CSD: YIDYIF). The results, shown in Figs. 3 and 4 of Shmueli & Flack (2010), display qualitative agreement and it is concluded that the p.d.f.s are useful.

Objection (i) of Olczak (2011) would preclude assuming anything during one's scientific work, and is therefore not reasonable. Objection (ii) referring to results that lose physical sense is taken care of by Figs. 3 and 4 of Shmueli & Flack (2010) and his comment (iii) is flattering but incorrect, as shown above. We have not been able to derive analytical forms of the p.d.f.s without assuming a uniform distribution of the fractional parts of the scalar products  $\mathbf{h} \cdot \mathbf{R}_{jk}$ . On the other hand, it is possible to undertake simulations similar to those described in Shmueli & Flack (2010) in which, instead of  $\frac{1}{2}N(N-1)$  independent vectors, one takes  $N-1$  independent vectors and evaluates the others from these. This concludes our response.

### References

- Olczak, A. (2011). *Acta Cryst.* **A67**, 315–317.  
 Shmueli, U. & Flack, H. D. (2010). *Acta Cryst.* **A66**, 669–675.  
 Weyl, H. (1916). *Math. Ann.* **77**, 313–352.