Probability density functions of the average and difference intensities of Friedel opposites

U. Shmueli[a]* and H. D. Flack[b]

[a]School of Chemistry, Tel Aviv University, 69978 Tel Aviv, Israel, and [b]Département de Chimie Minérale, Analytique et Appliquée, Faculté des Sciences, Université de Genève, Switzerland. Correspondence e-mail: ushmueli@post.tau.ac.il

Trigonometric series for the average (A) and difference (D) intensities of Friedel opposites were carefully rederived and were normalized to minimize their dependence on $\sin(\vartheta)/\lambda$. Probability density functions (hereafter p.d.f.s) of these series were then derived by the Fourier method [Shmueli, Weiss, Kiefer & Wilson (1984). *Acta Cryst.* A**40**, 651–660] and their expressions, which admit any chemical composition of the unit-cell contents, were obtained for the space group $P1$. Histograms of A and D were then calculated for an assumed random-structure model and for 3135 Friedel pairs of a published solved crystal structure, and were compared with the p.d.f.s after the latter were scaled up to the histograms. Good agreement was obtained for the random-structure model and a qualitative one for the published solved structure. The results indicate that the residual discrepancy is mainly due to the presumed statistical independence of the p.d.f.'s characteristic function on the contributions of the interatomic vectors.

1. Introduction

Examination of the distribution of the Bragg reflection intensities can be useful at various stages of structure analysis. The most widely used application is the comparison of the distribution of normalized observed intensities with the theoretical ones for space groups $P1$ and $P\overline{1}$ (Wilson, 1949). The objective is to determine whether the space group of the unknown crystal is centrosymmetric or non-centrosymmetric. The theoretical distributions of Wilson (1949) have been derived on the basis of a large number of identical atoms in the triclinic system, ignoring the effect of resonant scattering and using the central limit theorem. In practice, erroneous indications of whether the structure is centrosymmetric or non-centrosymmetric often occur. Shmueli & Weiss (1995) have presented techniques to overcome the limitations of Wilson statistics, but practical applications have been limited. Moreover, the enigma that pseudo-centrosymmetric structures nevertheless give reliable indications of the Flack (1983) parameter led Flack & Shmueli (2007) to investigate the first and second moments of the intensity distributions of the average and difference of the Friedel opposites. This analysis resolved the enigma and resulted in Shmueli *et al.* (2008) investigating the effect of point-group symmetry and lattice centering on these moments. However, for the crucial step of determining whether an unknown crystal is centrosymmetric or non-centrosymmetric, the full probability distribution of intensity or, better, of the average and difference intensities of Friedel opposites free of the assumptions of Wilson (1949) are

required. The present paper makes an important contribution to the solution of this problem.

Parthasarathy & Srinivasan (1964), Parthasarathy (1967) and Srinivasan & Parthasarathy (1976) have undertaken to calculate the probability distribution of Bijvoet differences, the expectation values of Bijvoet ratios and various other functions associated with the effect of resonant scattering on Friedel differences. However, their starting models of the compounds always assume that there are P resonant atoms of the same chemical element along with Q non-resonant atoms. Q is taken to be large leading to ideal probability distributions for the non-resonant atoms. This starting assumption severely limits the applicability of their results. One never finds the important factor $(f_i f_j'' - f_j f_i'')$ appearing in their work, although it does occur in those of Okaya & Pepinsky (1955), Rossmann & Arnold (2001) and Shmueli *et al.* (2008).

The present probabilistic treatment of the problem was carried out along the lines of that described by Shmueli & Weiss (1995). The obvious advantage of the probability density functions (p.d.f.s) derived in this work over those given by the above authors are that:

(i) our p.d.f.s take into account explicitly *any* chemical composition of the compound investigated, and

(ii) our p.d.f.s are exact and are thus applicable to compositions for which the central limit theorem (as implied in the Wilson, 1949, statistics and used by the above authors) breaks down.

These p.d.f.s are compared with simulated distributions of the average and difference of the intensities of Friedel opposites,

as well as with such distributions based on explicitly calculated quantities from published parameters of a solved crystal structure.

2. Preliminaries

It appears that the main prerequisites for the derivation of the p.d.f. of the average and difference of the intensities of Friedel opposites are convenient functional forms of these quantities. The expressions given by Okaya & Pepinsky (1955) and re-derived later (*e.g.* Rossmann & Arnold, 2001), in their studies of the influence of resonant scattering on the Patterson function, seem to be eminently suitable. For the sake of completeness, they will be rederived here for the space group $P1$. We have

$$|F(\mathbf{h})|^2 = F(\mathbf{h})F^*(\mathbf{h})$$
$$= \sum_{i=1}^{N}(f_i + if_i'')\exp(2\pi i\mathbf{h}\cdot\mathbf{r}_i)\sum_{j=1}^{N}(f_j - if_j'')\exp(-2\pi i\mathbf{h}\cdot\mathbf{r}_j)$$
$$= \sum_{i=1}^{N}\sum_{j=1}^{N}[(f_if_j + f_i''f_j'') - i(f_if_j'' - f_i''f_j)]\exp[2\pi i\mathbf{h}\cdot(\mathbf{r}_i-\mathbf{r}_j)],$$
$$(1)$$

where f_i and f_j include the real parts of the resonant-scattering contribution. The components of the atomic scattering factor also include the displacement parameters.

Since $|F(\mathbf{h})|^2$ is of necessity real, we take the real parts of equation (1). This leads to

$$|F(\mathbf{h})|^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}[(f_if_j + f_i''f_j'')\cos[2\pi\mathbf{h}\cdot(\mathbf{r}_i-\mathbf{r}_j)]$$
$$+ \sum_{i=1}^{N}\sum_{j=1}^{N}(f_if_j'' - f_i''f_j)\sin[2\pi\mathbf{h}\cdot(\mathbf{r}_i-\mathbf{r}_j)] \quad (2)$$

and since only the sine term in equation (2) changes sign as $\mathbf{h}$ changes its sign

$$|F(-\mathbf{h})|^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}[(f_if_j + f_i''f_j'')\cos[2\pi\mathbf{h}\cdot(\mathbf{r}_i-\mathbf{r}_j)]$$
$$- \sum_{i=1}^{N}\sum_{j=1}^{N}(f_if_j'' - f_i''f_j)\sin[2\pi\mathbf{h}\cdot(\mathbf{r}_i-\mathbf{r}_j)]. \quad (3)$$

The average reduced intensity of Friedel opposites is now given by

$$A_F(\mathbf{h}) = \tfrac{1}{2}[|F(\mathbf{h})|^2 + |F(-\mathbf{h})|^2]$$
$$= \sum_{i=1}^{N}\sum_{j=1}^{N}(f_if_j + f_i''f_j'')\cos[2\pi\mathbf{h}\cdot(\mathbf{r}_i-\mathbf{r}_j)] \quad (4)$$

and the difference reduced intensity by

$$D_F(\mathbf{h}) \equiv |F(\mathbf{h})|^2 - |F(-\mathbf{h})|^2$$
$$= 2\sum_{i=1}^{N}\sum_{j=1}^{N}(f_if_j'' - f_i''f_j)\sin[2\pi\mathbf{h}\cdot(\mathbf{r}_i-\mathbf{r}_j)]. \quad (5)$$

Equation (5) is the expression given by Okaya & Pepinsky (1955). In the following we normalize the above expressions to remove as far as possible the variation of the free-electron atomic scattering factors and the Debye–Waller factors with $\sin(\vartheta)/\lambda$, as is common in all direct methods for the solution of crystal structures. To normalize A_F and D_F we may divide the real and imaginary parts of the scattering factor with $\sqrt{\Sigma}$, where Wilson's Σ is given by

$$\Sigma = \sum_{i}(f_i^2 + f_i''^2) \quad (6)$$

and equals the mean value of A_F (Flack & Shmueli, 2007). We hence come to

$$A_E = A_F/\Sigma \quad (7)$$

$$D_E = D_F/\Sigma \quad (8)$$

for the space group $P1$. It is necessary to make some comment on this choice of normalization. It is known to work well for the calculation of A_E. For the normalization of D_F it is not possible to use $\langle D_F\rangle$ as the latter is zero. The obvious second choice is to use $\langle D_F^2\rangle^{1/2}$ as the normalization factor in the denominator. For measured intensities this can lead to difficulties as often the values of $|D_{\text{obs}}|$ are subject to large uncertainty and, as shown by Flack *et al.* (2010), frequently $|D_{\text{obs}}|$ are considerably overestimated. Flack *et al.* (2010) convincingly demonstrated that the values of D have a different dependence on $\sin(\vartheta)/\lambda$ than the values of A. As a consequence, in one practical application Flack *et al.* (2010) preferred to first normalize the value of D, as given by equation (8), and then apply an additional correction which allowed for the different dependence on $\sin(\vartheta)/\lambda$ of the values of D from the values of A. In writing equation (8) this additional correction has been omitted and allows A_E and D_E to be written in terms of $|E(\mathbf{h})|$ and $|E(-\mathbf{h})|$, as given below.

In this paper A_E and D_E hence refer to the normalized average and difference reduced intensities of Friedel opposites, and since for the space group $P1$ the order of the isotropy subgroup (*e.g.* Shmueli *et al.*, 2008) equals unity, they can be defined as

$$A_E = \tfrac{1}{2}\big(|E(\mathbf{h})|^2 + |E(-\mathbf{h})|^2\big) \quad (9)$$
$$D_E = |E(\mathbf{h})|^2 - |E(-\mathbf{h})|^2, \quad (10)$$

where $E(\mathbf{h})$ is the normalized structure factor of reflection $\mathbf{h}$.

We wish to express equations (4)–(8) more concisely taking care of the diagonal terms ii and the related off-diagonal terms ij and ji. We write L_{ij} to mean a run of values over the lower off-diagonal triangle of a square $N \times N$ matrix with elements ij. L_{ij} will be used in sums and products such as

$$\sum_{L_{ij}} \equiv \sum_{i=2}^{N}\sum_{j=1}^{i-1} \quad \text{and} \quad \prod_{L_{ij}} \equiv \prod_{i=2}^{N}\prod_{j=1}^{i-1}.$$

Let also $\mathbf{R}_{ij} \equiv \mathbf{r}_i - \mathbf{r}_j$.

To rewrite equations (7) and (8), let

$$a_{ij} \equiv (f_if_j + f_i''f_j'')/\Sigma \quad (11)$$

and

$$d_{ij} \equiv (f_if_j'' - f_jf_i'')/\Sigma. \quad (12)$$

For $A_E(\mathbf{h})$, $a_{ij} = a_{ji}$ and $\cos(2\pi\mathbf{h} \cdot \mathbf{R}_{ij}) = \cos(2\pi\mathbf{h} \cdot \mathbf{R}_{ji})$. Therefore, the matrix $||a_{ij}\cos(2\pi\mathbf{h} \cdot \mathbf{R}_{ij})||$ is symmetric. However,

$$\sum_{i=1}^{N} a_{ii} = \sum_{i=1}^{N}(f_i^2 + f_i''^2)/\Sigma = 1$$

and therefore

$$A_E(\mathbf{h}) = 1 + 2\sum_{L_{ij}} a_{ij}\cos(2\pi\mathbf{h} \cdot \mathbf{R}_{ij}). \tag{13}$$

For $D_E(\mathbf{h})$, $d_{ji} = -d_{ij}$ and $\sin(2\pi\mathbf{h} \cdot \mathbf{R}_{ji}) = -\sin(2\pi\mathbf{h} \cdot \mathbf{R}_{ij})$. Therefore, $d_{ji}\sin(2\pi\mathbf{h} \cdot \mathbf{R}_{ji}) = d_{ij}\sin(2\pi\mathbf{h} \cdot \mathbf{R}_{ij})$ and the matrix $||d_{ij}\sin(2\pi\mathbf{h} \cdot \mathbf{R}_{ij})||$ is symmetric, while all its terms on the main diagonal have a value of zero. Hence, also in view of equation (5)

$$D_E(\mathbf{h}) = 4\sum_{L_{ij}} d_{ij}\sin(2\pi\mathbf{h} \cdot \mathbf{R}_{ij}). \tag{14}$$

3. The p.d.f. of A_E and its characteristic function

The expression for A_E, to be used in explicit computations, is given by equation (13). However, in the calculation of the p.d.f. of A_E we shall use the definition

$$A_E'(\mathbf{h}) \equiv A_E(\mathbf{h}) - 1 = 2\sum_{L_{ij}} a_{ij}\cos(2\pi\mathbf{h} \cdot \mathbf{R}_{ij}). \tag{15}$$

It is seen from equations (15) and (11), and the fact that $a_{ij} > 0$, that the maximum possible value of A_E' is

$$A_E'(\text{max}) = 2\sum_{L_{ij}} a_{ij}. \tag{16}$$

Hence, the probability of finding $A_E'(\mathbf{h})$ outside the range $[-A_E'(\text{max}), A_E'(\text{max})]$ is zero. The p.d.f. of A_E' ($\mathbf{h}$ omitted for simplicity) can therefore be expanded in a Fourier series within the range of its existence (Barakat, 1974; Shmueli *et al.*, 1984).

The general form of the p.d.f. expressed as a Fourier series of A_E' can be written as

$$p(A_E') = \frac{1}{2A_E'(\text{max})}\sum_{k=-\infty}^{\infty} Q_k \exp\left(-\frac{\pi i k A_E'}{A_E'(\text{max})}\right) \tag{17}$$

or

$$p(A_E') = \frac{\alpha}{2}\sum_{k=-\infty}^{\infty} Q_k \exp(-\pi i k \alpha A_E'), \tag{18}$$

where $\alpha = 1/A_E'(\text{max})$. From the basic Fourier theory the coefficients Q_k are given by

$$Q_k = \int_{-A_E'(\text{max})}^{A_E'(\text{max})} p(A_E') \exp(\pi i k \alpha A_E')\, dA_E'. \tag{19}$$

Since $p(A_E')$ vanishes outside the interval $[-A_E'(\text{max}), A_E'(\text{max})]$, the integration limits in equation (19) may be replaced by $\pm\infty$ without loss of generality (Shmueli & Weiss, 1995). We thus have

$$Q_k = \int_{-\infty}^{\infty} p(A_E') \exp(\pi i k \alpha A_E')\, dA_E'. \tag{20}$$

If we define, as is customary, the expected value of a function $f(x)$ as

$$E(f) = \int_{-\infty}^{\infty} f(x)p(x)\, dx \equiv \langle f \rangle,$$

the coefficient Q_k is the expected value of the exponential in equation (20) which can be written as

$$Q_k = \langle \exp(\pi i k \alpha A_E')\rangle \tag{21}$$

$$= \left\langle \exp\left[2\pi i k \alpha \sum_{L_{ij}} a_{ij}\cos(2\pi\mathbf{h} \cdot \mathbf{R}_{ij})\right]\right\rangle \tag{22}$$

$$= \left\langle \prod_{L_{ij}} \exp(2\pi i k \alpha a_{ij}\cos\varphi_{ij})\right\rangle \tag{23}$$

$$= \prod_{L_{ij}} \langle \exp(2\pi i k \alpha a_{ij}\cos\varphi_{ij})\rangle \tag{24}$$

$$= \prod_{L_{ij}} \langle \exp(2\pi i k \alpha a_{ij}\cos\varphi)\rangle \tag{25}$$

$$= \prod_{L_{ij}} (1/2\pi) \int_{-\pi}^{\pi} \exp(2\pi i k \alpha a_{ij}\cos\varphi)\, d\varphi \tag{26}$$

$$= \prod_{L_{ij}} J_0(2\pi k \alpha a_{ij}). \tag{27}$$

Equation (21) is the characteristic function of the p.d.f. of A_E' at the point $\pi k \alpha$ (*e.g.* Shmueli & Weiss, 1995), which was simplified to equation (27) thanks to two assumptions to be mentioned below. Equation (22) is obtained from equation (21) by inserting the expression for A_E' from equation (15). Since only the fractional part of $\mathbf{h} \cdot \mathbf{R}_{ij}$ is relevant, the argument of the cosine in equation (22) is replaced by the angle φ_{ij}, which exists somewhere in the $[0, 2\pi]$ interval. Since the complex exponentials in equation (23), also depending on the various interatomic vectors $\mathbf{R}_{ij}$ in the lower off-diagonal triangle, are assumed to be independent, the average of the product in equation (23) is replaced with a product of the averages in equation (24). Since the fractional part of $\mathbf{h} \cdot \mathbf{R}_{ij}$ is assumed to be uniform in the $[0, 1]$ interval, this implies the uniformity of φ_{ij} in the $[0, 2\pi]$ interval and the indices ij on φ are therefore omitted in equation (25). The integral in equation (26) is just a representation of the Bessel function of zero order and of the first kind (*e.g.* Abramowitz & Stegun, 1972) which appears in equation (27).

If we now insert equation (27) into equation (18), the p.d.f. of A_E' becomes

$$p(A_E') = \frac{\alpha}{2}\sum_{k=-\infty}^{\infty}\left[\prod_{L_{ij}} J_0(2\pi k \alpha a_{ij})\right]\exp(-\pi i k \alpha A_E'). \tag{28}$$

Since for $k = 0$ the argument of J_0 is zero, this term contributes $J_0(0) = 1$. Since $J_0(-x) = J_0(x)$, then $Q_{-k} = Q_k$. Further, $\exp(-\pi i k \alpha A_E') + \exp(\pi i k \alpha A_E') = 2\cos(\pi k \alpha A_E')$ and equation (28) therefore simplifies to

$$p(A'_E) = \frac{\alpha}{2} \left\{ 1 + 2 \sum_{k=1}^{\infty} \left[\prod_{L_{ij}} J_0(2\pi k \alpha a_{ij}) \right] \cos(\pi k \alpha A'_E) \right\}. \quad (29)$$

Finally, since $p(A'_E) = p(-A'_E)$, we find the required p.d.f. $p(|A'_E|)$ as

$$p(|A'_E|) = \alpha \left\{ 1 + 2 \sum_{k=1}^{\infty} \left[\prod_{L_{ij}} J_0(2\pi k \alpha a_{ij}) \right] \cos(\pi k \alpha |A'_E|) \right\}. \quad (30)$$

It is interesting to point out that the p.d.f. $p(|A'_E|)$ has the same functional form as the p.d.f. of $|E|$ for the centrosymmetric space group $P\bar{1}$ [*cf.* equation (5.1.9) in Shmueli & Weiss, 1995]. The reason for this is a similar functional form of A'_E in equation (15) to that of the expression for the normalized structure factor, E. The coefficients and their meaning are of course entirely different.

4. The p.d.f. of D_E and its characteristic function

It is seen from equation (14), upon taking the absolute magnitude of d_{ij}, that the maximum possible value of $D_E(\mathbf{h})$ is

$$D_E(\text{max}) = 4 \sum_{L_{ij}} |d_{ij}|. \quad (31)$$

Similar to the case of A'_E, the probability of finding $D_E(\mathbf{h})$ outside the range $[-D_E(\text{max}), D_E(\text{max})]$ is zero. The p.d.f. of D_E (**h** omitted for simplicity) can therefore be expanded in a Fourier series within the range of its existence (Barakat, 1974; Shmueli *et al.*, 1984).

The general form of the p.d.f. expressed as a Fourier series of D_E can be written as

$$p(D_E) = \frac{1}{2D_E(\text{max})} \sum_{k=-\infty}^{\infty} C_k \exp\left(-\frac{\pi i k D_E}{D_E(\text{max})}\right) \quad (32)$$

or

$$p(D_E) = \frac{\delta}{2} \sum_{k=-\infty}^{\infty} C_k \exp(-\pi i k \delta D_E), \quad (33)$$

where $\delta = 1/D_E(\text{max})$. In analogy with equation (21) the characteristic function of the p.d.f. of D_E is

$$C_k = \langle \exp(\pi i k \delta D_E) \rangle. \quad (34)$$

The subsequent derivation is exactly analogous to the derivation of the p.d.f. of A'_E. It differs in that d_{ij} replaces a_{ij}, δ replaces α and D_E replaces A'_E. The interpretation of the average [analogous to equation (26)] as a Bessel function $J_0(x)$ follows from the relation

$$J_0(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(ix \cos\theta)\, d\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(ix \sin\theta)\, d\theta,$$

which is readily derived from the generating function of Bessel functions of the first kind (*e.g.* Abramowitz & Stegun, 1972). The intermediate result for $p(D_E)$ is

$$p(D_E) = \frac{\delta}{2} \sum_{k=-\infty}^{\infty} \left[\prod_{L_{ij}} J_0(4\pi k \delta d_{ij}) \right] \exp(-\pi i k \delta D_E). \quad (35)$$

Since for $k = 0$ the argument of J_0 is zero, this term contributes $J_0(0) = 1$. Since $J_0(-x) = J_0(x)$, we have $C_{-k} = C_k$. Further, $\exp(-\pi i k \delta D_E) + \exp(\pi i k \delta D_E) = 2\cos(\pi k \delta D_E)$ and equation (35) therefore simplifies to

$$p(D_E) = \frac{\delta}{2} \left\{ 1 + 2 \sum_{k=1}^{\infty} \left[\prod_{L_{ij}} J_0(4\pi k \delta d_{ij}) \right] \cos(\pi k \delta D_E) \right\}. \quad (36)$$

5. Methodology for comparison of the p.d.f.s with various histograms of A_E and D_E

The above theoretical derivation is of interest for its own sake, however, it appears to be very important to find out to what extent these p.d.f.s of A'_E and D_E are meaningful when tested against data from a crystal belonging to the space group $P1$. The essence of such a study is the construction of various histograms of A_E and D_E and their comparison with the p.d.f.s of A'_E and D_E, respectively, when placed on the same scale. There are three distinct ways of creating the histograms of A_E and D_E described in the following subsections. Each provides a different way of comparing theory with experiment.

All the computations were programmed with *GNU FORTRAN* software, and the drawings were produced with the aid of the *PGPLOT FORTRAN* subroutine package. The discrepancy or agreement between the p.d.f.s and the histograms was computed as the conventional R factor defined as

$$R = \frac{\sum_i |h_i - p_i|}{\sum_i h_i},$$

where h_i is the height of the ith histogram bin and p_i is the value of the p.d.f. at the midpoint of the basis of the ith histogram bin. The R factors are given at the top of the figure boxes.

5.1. Random-structure model

The histograms of A_E and D_E are produced by a simulation creating a set of structures each with a set of interatomic vectors generated randomly. This is achieved by assuming that the fractional part of the scalar product $\mathbf{h} \cdot \mathbf{R}_{ij}$ in equations (13) and (14) is uniformly distributed on the [0, 1] range. In practice, each fractional part of $\mathbf{h} \cdot \mathbf{R}_{ij}$ is produced by drawing separately a pseudo-random number based on a uniform distribution on the [0, 1] range. Each structure in the set produces one value each of A_E and D_E, and the multitude of structures results in the histogram. Normalized scattering factors were assumed to be independent of $\sin(\vartheta)/\lambda$ and were computed on the basis of the composition of the simulated structure. The p.d.f.s were only computed at the midpoints of the histogram bins. It is a multiple-structure single-reflection calculation. This is analogous to the fixed-index and random-

position assumption underlying most direct methods (*e.g.* Giacovazzo, 1998).

For our tests of the random-structure model, the hypothetical $P1$ structure of composition U_2C_{19} measured with Cu $K\alpha$ radiation was used. The computation of the scattering factors was carried out as indicated in *Appendix B* of Flack & Shmueli (2007). The computation of the $|A_E|$ and D_E sums was carried out for 50 000 random structures from which the histograms were constructed and to which the p.d.f.s were scaled up.

The underlying assumptions for this simulation are the same as those used in the analytical derivation of the p.d.f.s in §§3 and 4. A reasonable agreement between the histograms and the p.d.f.s would confirm the analysis.

5.2. Published crystal-structure model

The histograms of A_E and D_E are constructed from a published crystal structure. The atomic coordinates are taken from the literature and an overall isotropic displacement parameter was employed. It was taken as an average of all isotropic or equivalent isotropic displacement parameters of the atoms. A_E and D_E are calculated from the atomic coordinates using equations (13) and (14) for all Friedel pairs reported as having been observed, although the observed intensities themselves are not used. In spite of the normalization of A_F and D_F by equations (7) and (8), respectively, there remained some dependence on $\sin(\vartheta)/\lambda$, especially at low values of the diffraction angle. We therefore averaged the coefficients a_{ij} and d_{ij} as well as the maximum values $A'_E(\max)$ and $D_E(\max)$ over all the Friedel pairs and used these averages in the computation of the p.d.f.s $p(A'_E)$ and $p(D_E)$. The computation comprised:

(i) a retrieval of the positional and displacement parameters of the structure examined from its CIF file,

(ii) an analysis of the deposited hkl file in terms of Friedel pairs and unpaired reflections,

(iii) computation of all the A_E's and D_E's from the equations (13) and (14) and construction of their histograms, and

(iv) computation of the p.d.f.s at four points at the basis of each histogram bin, including the midpoint of the bin. (In one instance only the scalar products were replaced by pseudo-random numbers.)

It is a single-structure multiple-reflection calculation. This is analogous to the variable-index and fixed structure assumption underlying intensity statistics (*e.g.* Shmueli & Weiss, 1995; Section 1.3).

For our tests of the published crystal-structure method, the reflection data and the structural parameters for the example considered were taken from the CIF and hkl files accompanying the work of Chartrand *et al.* (2007) (CSD code YIDYIF). The structure belongs to the triclinic system, space group $P1$, the formula of the unit cell content being $C_{46}H_{32}Br_4F_{12}N_8P_2Ru$. High-quality intensity data, according to their analysis by Flack *et al.* (2010), were collected with Cu $K\alpha$ radiation by Chartrand *et al.* (2007). The structure thus displays considerable atomic heterogeneity and significant resonant scattering, which justify its choice in the comparison

with the p.d.f.s derived in this work. All the real and imaginary parts of the resonant scattering contributions and all the positional and isotropic/equivalent displacement parameters were included in the computations. The free-electron atom scattering factors were computed from the analytical approximation of Brown *et al.* (1999) and the resonant scattering contributions were taken from Creagh (1999). The total number of reflections is 7003, containing 3135 Friedel pairs and 733 unpaired reflections. The latter were not included in the computation. Measured intensities were not used.

The discrepancies between the histograms and the p.d.f.s would attest to the validity of the uniform-distribution approximation for the structure examined, and the assumption of statistically independent contributions of interatomic vectors. These assumptions underlie the derivations of the p.d.f.s.

[Note: Whether we have the multiple-structure and single-reflection or the single-structure and multiple-reflection situation we presume that the components of the interatomic vectors are rationally independent (*e.g.* Shmueli & Weiss, 1995; Section 1.3). In simpler terms, we require that the interatomic vectors be in general positions of the space group of the Patterson function.]

5.3. Observed intensities

The histograms of A_E and D_E are constructed directly from the reduced and normalized intensities in the literature without any reference to the crystal-structure model.

Since the purpose of our study was to establish the limits of validity of the analytical p.d.f.s derived in §§3 and 4, we have not produced histograms from the intensity data in the literature. Of course, if it turns out that our p.d.f.s have a wide range of validity, comparison of the histograms from intensity data with the p.d.f.s would help in the choice of space group. However, Flack *et al.* (2010) observed that values of D for published crystal structures are often dominated by random and systematic errors in measurement and data reduction.

6. Results

Figs. 1 and 2 show the results for the U_2C_{19} compound using the random-structure model described in full in §5.1. The histogram and p.d.f. of $|A_E|$ for U_2C_{19} are shown in Fig. 1. Agreement between the two is very good, $R = 0.042$, thus supporting the correctness of the above analysis. The histogram and p.d.f. of D_E for U_2C_{19} can be seen in Fig. 2. Agreement between the two is also good, $R = 0.070$, thus again confirming the above analysis.

Figs. 3 and 4 show the results for YIDYIF using the published crystal-structure model described in full in §5.2. The histogram and p.d.f. of $|A_E|$ for YIDYIF are shown in Fig. 3. There is a qualitative agreement but the discrepancy is significant, $R = 0.244$. This may be due to the fact that in the derivation of the p.d.f. $p(|A'_E|)$ we assumed that:

(i) the functions depending on the interatomic vectors are statistically independent, and

(ii) the fractional parts of the scalar products $\mathbf{h} \cdot \mathbf{R}_{ij}$ are uniformly distributed on the [0, 1] interval.

Either or both of these assumptions may not be exactly fulfilled in the case of a real crystal. It would, in principle, be possible to test the fulfillment of the second assumption by replacing the scalar products $\mathbf{h} \cdot \mathbf{R}_{ij}$ with pseudo-random numbers, uniform on the [0, 1] interval. However, the cosine series in equation (13) would then be as likely positive as negative, while A_E is positive by definition. The physical reason for such an effect is that in general a random set of interatomic vectors is not the vector set of a random crystal structure. Such a simulation would make no physical sense.

Note, however, that all the A_E's on which Fig. 3 is based are positive as they are derived from a real crystal structure.

The situation with the D_E sums is similar. Fig. 4 shows the comparison of the histogram and p.d.f. of D_E. The qualitative agreement seems to be better than that in Fig. 3 and the discrepancy between the histogram and p.d.f. is smaller, $R = 0.146$, than that for the A_E sums. In a separate test, since D_E may be positive as well as negative, we replaced the explicit expressions for the scalar products $\mathbf{h} \cdot \mathbf{R}_{ij}$ with pseudo-random numbers uniform on the [0, 1] interval. The result is shown in Fig. 5. A significantly better agreement was obtained in this partial random-structure model approach, $R = 0.076$, and it is
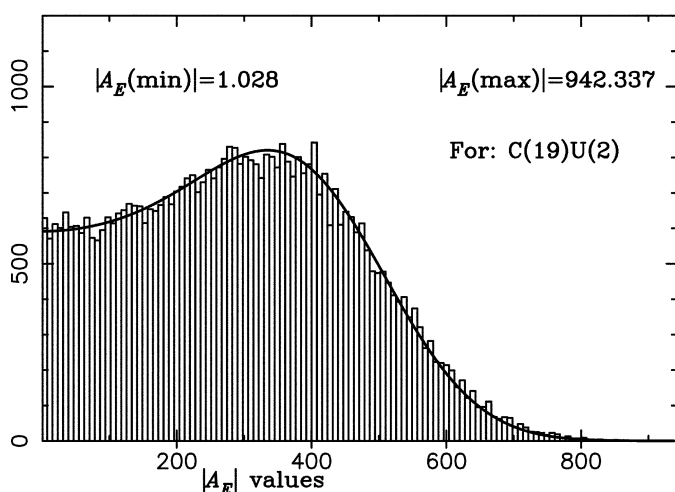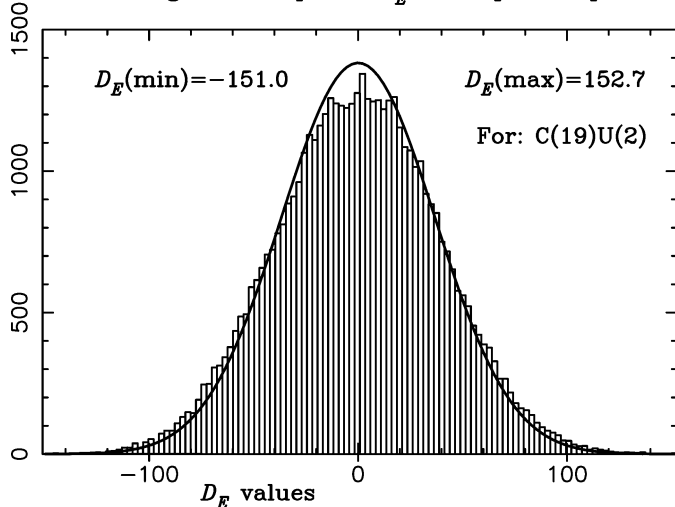

Figure 1
Histogram and p.d.f. of $|A_E|$ for compound U_2C_{19} calculated by the random-structure model. The p.d.f. is indicated by a solid line. The discrepancy factor is $R = 0.042$. Full details are given in §§5.1 and 6.
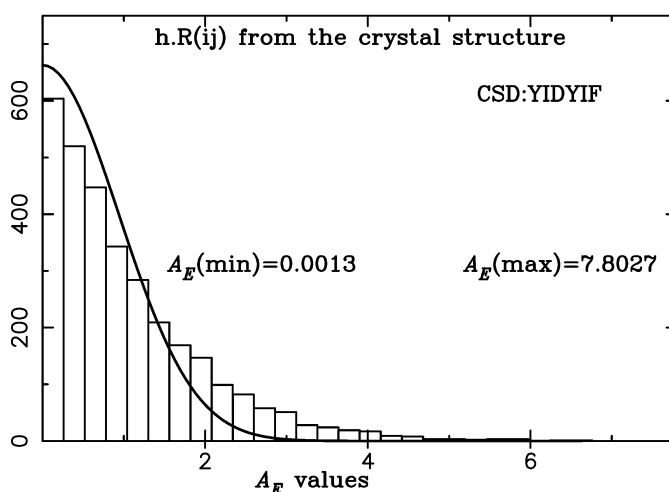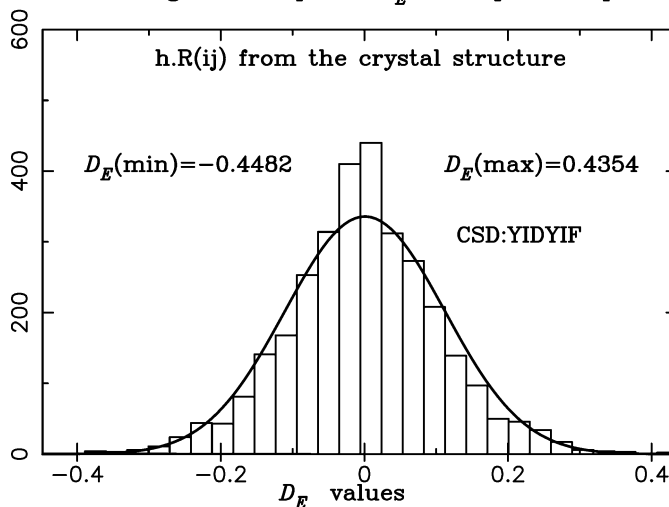

Figure 3
Histogram and p.d.f. of $|A_E|$ for compound YIDYIF calculated by the published crystal-structure model. The p.d.f. is indicated by a solid line. The discrepancy factor is $R = 0.244$. Full details are given in §§5.2 and 6.
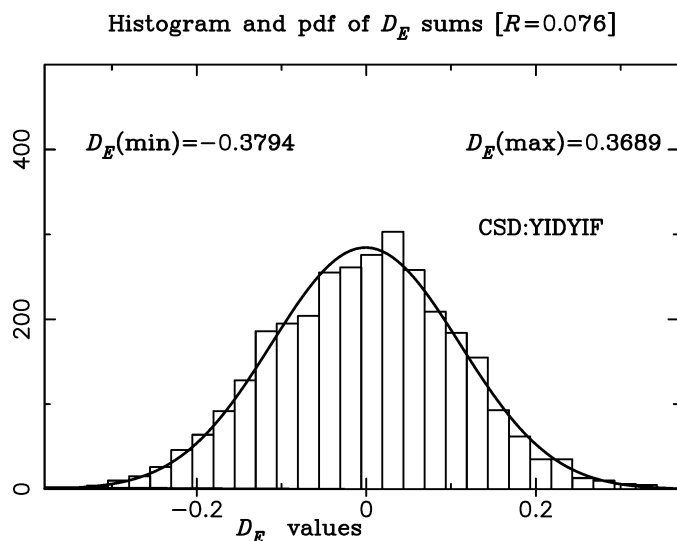

Figure 2
Histogram and p.d.f. of D_E for compound U_2C_{19} calculated by the random-structure model. The p.d.f. is indicated by a solid line. The discrepancy factor is $R = 0.070$. Full details are given in §§5.1 and 6.

Figure 4
Histogram and p.d.f. of D_E for compound YIDYIF calculated by the published crystal-structure model. The p.d.f. is indicated by a solid line. The discrepancy factor is $R = 0.146$. Full details are given in §§5.2 and 6.

Figure 5
Histogram and p.d.f. of D_E for compound YIDYIF calculated by the random-structure model. The p.d.f. is indicated by a solid line. The discrepancy factor is $R = 0.076$. Full details are given in §§5.2 and 6.

in software is presented in the current work. Moreover, for use as an aid in space-group determination, the simulations do not need to be of high accuracy, thus limiting the calculation time necessary.

We also wish to point out that the analytical expressions for $A(\mathbf{h})$ and $D(\mathbf{h})$ in terms of interatomic vectors in equations (4) and (5) led Flack *et al.* (2010) to investigate the practical applications of their Fourier transforms. These are the A and D Patterson functions, and are found to have real use in the validation of a crystal-structure determination.

tempting to conclude that the residual discrepancy is mainly due to the assumed statistical independence of the complex exponentials in equation (23), which also depend on the interatomic vectors.

7. Concluding remarks

The current work is limited to the triclinic space group $P1$. A natural extension will thus be to extend the work to the other 229 space groups, taking care of the general/special centrosymmetric/non-centrosymmetric status of the reflections in these space groups (see Shmueli & Flack, 2009). An important question of strategy is the choice of the form in which the p.d.f.s should be prepared so as to be of maximum use to the structure analyst. Historically, p.d.f.s have been preferentially prepared in analytical form, which are then evaluated numerically using parameters appropriate to the crystal structure under consideration. However, Shmueli & Weiss (1995) and the current work have used numerical simulation to check the analytical solution. Considering the immense calculation power and low price of modern computers, it may not in fact be necessary to seek analytical solutions and one may rely on suitably optimized simulations. Most, but not all, of the analysis necessary for implementing the latter approach

References

Abramowitz, M. & Stegun, I. A. (1972). *Handbook of Mathematical Functions*. New York: Dover.
Barakat, R. (1974). *Opt. Acta*, **21**, 903–921.
Brown, P. J., Fox, A. G., Maslen, E. N., O'Keefe, M. A. & Willis, B. T. M. (1999). *International Tables for Crystallography*, Vol. C, edited by A. J. C. Wilson & E. Prince, Table 6.1.1.4. Dordrecht: Kluwer Academic Publishers.
Chartrand, D., Theobald, I. & Hanan, G. S. (2007). *Acta Cryst.* E**63**, m1561.
Creagh, D. C. (1999). *International Tables for Crystallography*, Vol. C, edited by A. J. C. Wilson & E. Prince, Table 4.2.6.8. Dordrecht: Kluwer Academic Publishers.
Flack, H. D. (1983). *Acta Cryst.* A**39**, 876–881.
Flack, H. D., Sadki, M., Thompson, A. L. & Watkin, D. J. (2010). *Acta Cryst.* A. Submitted.
Flack, H. D. & Shmueli, U. (2007). *Acta Cryst.* A**63**, 257–265.
Giacovazzo, C. (1998). *Direct Phasing in Crystallography*, Section 1.11. Oxford University Press.
Okaya, Y. & Pepinsky, R. (1955). *Computing Methods and the Phase Problem in X-ray Crystal Analysis*, p. 276. Oxford: Pergamon Press.
Parthasarathy, S. (1967). *Acta Cryst.* **22**, 98–103.
Parthasarathy, S. & Srinivasan, R. (1964). *Acta Cryst.* **17**, 1400–1407.
Rossmann, M. G. & Arnold, E. (2001). *International Tables for Crystallography*, Vol. B, edited by U. Shmueli, ch. 2.3, pp. 235–263. Dordrecht: Kluwer Academic Publishers.
Shmueli, U. & Flack, H. D. (2009). *Acta Cryst.* A**65**, 322–325.
Shmueli, U., Schiltz, M. & Flack, H. D. (2008). *Acta Cryst.* A**64**, 476–483.
Shmueli, U. & Weiss, G. H. (1995). *Introduction to Crystallographic Statistics*, IUCr Monograph No. 6. Oxford University Press.
Shmueli, U., Weiss, G. H., Kiefer, J. E. & Wilson, A. J. C. (1984). *Acta Cryst.* A**40**, 651–660.
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.